

**Settimo Convegno Annuale
AIUCD 2018
Bari, 31 gennaio – 2 febbraio 2018**

Book of Abstracts

**Patrimoni culturali nell'era digitale.
Memorie, culture umanistiche e tecnologia**

***Cultural Heritage in the Digital Age.
Memory, Humanities and Technologies***

edited by Daria Spampinato

ASSOCIAZIONE PER
L'INFORMATICA UMANISTICA
E LA CULTURA DIGITALE



2018

ISBN: 9788894253528

Copyright © 2018

ASSOCIAZIONE PER
L'INFORMATICA UMANISTICA
E LA CULTURA DIGITALE



Associazione per l'Informatica Umanistica e la Cultura Digitale



Copyright of each individual chapter is maintained by the authors.

This work is licensed under a Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text and to make commercial use of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Daria Spampinato (ed.), AIUCD 2018 - *Book of Abstracts*, Bologna 2018.

Available online as a supplement of *Umanistica Digitale*:

<https://umanisticadigitale.unibo.it>

Cover image has been created by Paolo Azzella - Quorum Italia Srl.

If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

All links have been visited on 29th December 2017, unless otherwise indicated.

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notified to the editor: daria.spampinato@cnr.it

Organizzazione

Gli abstract pubblicati in questo volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante double-blind peer review sotto la responsabilità del Comitato Scientifico di AIUCD 2016. Il programma della conferenza AIUCD 2018 è disponibile online all'indirizzo <http://www.aiucd2018.uniba.it/>.

All abstracts published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double peer review process under the responsibility of the AIUCD 2016 Scientific Committee. The AIUCD 2018 conference program is available online at <http://www.aiucd2018.uniba.it/>.

Comitato di Programma Internazionale

Daria Spampinato (Chair) - Istituto di Scienze e Tecnologie della Cognizione - CNR
Nicola Barbuti (Co-chair and local organizer) - Università di Bari "Aldo Moro"
Federico Boschetti - Istituto di Linguistica Computazionale "Antonio Zampolli" - CNR
Fabio Ciotti - Università di Roma "Tor Vergata"
Giorgio Maria Di Nunzio - Università di Padova
Stefano Ferilli - Università di Bari "Aldo Moro"
Greta Franzini - "Georg-August" - Universität Göttingen
Emiliano Giovannetti - Istituto di Linguistica Computazionale Antonio Zampolli - CNR
Alessandro Lenci - Università di Pisa
Paola Moscati - Istituto di Studi sul Mediterraneo Antico - CNR
Serge Noiret - European University Institute
Silvia Orlandi - Università di Roma "La Sapienza"
Claudia Villani - Università di Bari "Aldo Moro"

Comitato di Organizzazione Locale

Andrea Paziienza - DIB Università di Bari "Aldo Moro"
Michele Laricchia - DISUM Università di Bari "Aldo Moro"
Ottavio Ardillo - DISUM Università di Bari "Aldo Moro"
Luca Avellis - DISUM Università di Bari "Aldo Moro"
Ludovica Marinucci - Scuola a Rete Nazionale DiCultHer

Indice

Prefazione <i>Daria Spampinato</i>	11
---------------------------------------	----

Keynote lectures

Oltre i tradizionali confini disciplinari. Archeologia, filologia, codicologia e informatica umanistica: quattro discipline in dialogo per la realizzazione di un atlante digitale della letteratura copta. L'esperienza della gestione del progetto ERC 'PATHs' <i>Paola Buzi</i>	16
Saltare la carta, consentire l'accesso, disegnare gli scenari. Digital humanities: la pianura digitale verso il patrimonio comune europeo <i>Michele Rak</i>	17
What is cultural innovation? How do we measure it? <i>Riccardo Pozzo</i>	18

Long papers

A sustainable work flow for a multi-layer edition of the <i>Chronicon</i> by Romualdus Salernitanus <i>Paolo Monella</i>	21
AMiGre: A unified framework for archiving and processing oral and written dialectal data <i>Eleni Galiotou, Angela Ralli</i>	25
Approaches to the German reception of Verdi's »Messa da Requiem« through metadata analysis and »horizontal reading« <i>Torsten Roeder</i>	28
Building A Digital Ottoman/Turkish Serial Novel Archive <i>Ali Serdar, Reyhan Tutumlu</i>	33
Creating a Community on Film-induced Tourism through a Crowdsourcing Approach <i>Andrea Micheletti, Daniel Zilio, Nicola Orio</i>	38
Dai giacimenti culturali ai repository digitali: il Museo virtuale dell'informatica archeologica <i>Paola Moscati</i>	41
Exploiting ICT to Create Smart Visit Experiences at Cultural Heritage Sites <i>Carmelo Ardito, Paolo Buono, Maria Costabile, Danilo Caivano, Giuseppe Desolda, Rosa Lanzilotti, Maristella Matera, Antonio Piccinno</i>	45
Get Ready, Get Set, Curate: Understanding the 'Everyday Curator' <i>Michele Reilly, Santi Thompson</i>	49
Invading the Italian Literary Polysystem. A Distant Reading Case Study of Sienkiewicz's "Quo vadis" <i>Jan Rybicki, Katarzyna Biernacka-Licznar, Monika Woźniak</i>	52

L'analisi delle reti come strumento a supporto della ricerca storica. Da Ceprano a Benevento (1266) <i>Alessandro De Troia, Vito Ricci</i>	55
Le lettere di Bellini: dalla carta al Web <i>Angelo Mario Del Grosso, Salvatore Cristofaro, Maria Rosa De Luca, Emiliano Giovannetti, Simone Marchi, Graziella Seminara, Daria Spampinato</i>	60
Linked data ed edizioni scientifiche digitali. Esperimenti di trasformazione di un <i>Quaderno di appunti</i> <i>Marilena Daquino, Francesca Giovannetti, Francesca Tomasi</i>	65
L'habitat come patrimonio e le sue rappresentazioni. Per un atlante digitale dell'insediamento meridionale (XIII-XXI secolo) <i>Rosanna Rizzi, Biagio Salvemini, Lorena Maria Calculli, Giovanna Patruno</i>	70
Presentazione di Czech-IT!: Un corpus linguistico di parlanti cechi apprendenti la lingua italiana <i>Marco Petolicchio, Marcello Bolpagni</i>	74
Temporal Dimension in Alcide De Gasperi: Past, Present and Future in Historical Political Discourse <i>Rachele Sprugnoli, Giovanni Moretti, Sara Tonelli</i>	77
Thinking to Preservation. A Proposal for Innovation in Digital Library <i>Nicola Barbuti, Stefano Ferilli, Domenico Redavid, Tommaso Caldarola</i>	81
<i>Short papers</i>	
An experiment on the development of a digital edition for ancient Greek fragmentary poetry: A case study on Archilochus of Paros <i>Alberto Stefanini, Anika Nicolosi, Monica Monachini</i>	85
A.R.C.A. Project: uno strumento per la creazione di web applications per la pubblicazione degli scavi archeologici <i>Irene Carpanese, Guido Lucci Baldassari</i>	89
Bibliotheca Perspectivae <i>Stefano Casati, Andrea De Calisti</i>	91
CEED: a Cooperative Web-Based Editor for Critical Editions <i>Simone Zenzaro, Daniele Marotta, Amos Bertolacci</i>	93
Close and Computational Reading of Galileo Galilei and Alessandro Piccolomini <i>Crystal Hall</i>	97
Dalle Stampe Antiche al Museo Virtuale. Best practice nell'acquisizione di competenze digitali per l'archiviazione, conservazione e visualizzazione di materiale antico a stampa: il caso di un percorso di alternanza scuola-lavoro per gli studenti di un Liceo classico <i>Barbara Balbi, Francesca De Ruvo, Alessandra Monica Mazzaro, Alfredo Apicella, Roberto Montanari</i>	99
Designing a multi-layered User Interface for EVT 2: a development report <i>Roberto Rosselli Del Turco, Chiara Di Pietro, Chiara Martignano</i>	105

Digitalizzare la documentazione archeologica: metodologie e “buone pratiche” per la lavorazione, gestione e il riuso dei dati. L’archivio della Missione Archeologica Italiana in Anatolia Orientale (Arslantepe-Malatya, Turchia) <i>Margherita Bartoli, Heidi Paffrath</i>	109
Digitalizzazione delle risorse storico-artistiche e sustainability management a supporto dello sviluppo turistico-culturale in Puglia <i>Floriana Conte, Fabio De Matteis</i>	114
Do alchemists dream of electronic sheets? <i>Ilaria Cicola</i>	118
Edizioni digitali di opere a tradizione complessa: il caso del Conte di Carmagnola <i>Beatrice Nava</i>	123
From space to space: Image-based digital 3d-reconstruction of Indian Nagara architecture <i>Gerald Kozicz, Ludwig Grimm</i>	127
From the Clavius Correspondence to Linked Data: the CoW-LD Project <i>Angelica Lo Duca, Andrea Marchetti</i>	130
Gli Acta Eruditorum come esempio di valorizzazione di una collezione digitale <i>Stefano Casati, Federica Viazzi</i>	133
Historical Images and the Recovery of the Past. <i>The Medieval Kingdom of Sicily Image Database</i> <i>Caroline Bruzelius, Paola Vitolo, Joseph Williams</i>	135
IDEA e la conservazione dei dati epigrafici di EAGLE <i>Pietro Liuzzo, Franco Zoppi, Antonio E. Felle, Giuseppe Amato</i>	139
Identification of Documents in Written Cultural Heritage Digitalization – Description of Implemented Solution <i>Tomáš Klimek, Tomáš Psohlavec, Olga Čiperová</i>	141
Il Progetto Traduzione del Talmud Babilonese: il Ruolo della Tecnologia e della Linguistica Computazionale <i>Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, David Dattilo, Michael Dollinar, Alessandra Pecchioli, Clelia Piperno</i>	144
Integrazione di QR-code e localizzazione di prossimità in visite educative museali basate su pervasive gaming <i>Giovanni Luca Dierna, Alberto Machì</i>	147
<i>Labeculae Vivae</i> . Building a reference library of stains found on books <i>Alberto Campagnolo, Erin Connelly, Heather Wacha</i>	154
Mapping Bucharest from a Geo-Ontological Point of View <i>Timothy Tambassi</i>	156
Mapping Santiago de Compostela from Gonzalo Torrente Ballester's <i>Fragmentos de Apocalipsis</i> : uncertainty and spacial transformation <i>Alba Rozas Arceo</i>	159
Measuring the “Critical Distance”. A Corpus-Based Analysis of Italian Book Reviews <i>Massimo Salgaro, Simone Reborà</i>	161

‘Migrazioni e colonizzazione interna nel Mediterraneo d’età moderna’: un progetto di Umanistica digitale <i>Giampaolo Salice</i>	164
Patrimoni transdisciplinari. Memoria e cultura del <i>sensus loci</i> in ambiente digitale <i>Alessia Scacchi</i>	168
Pattern della globalizzazione – Trieste e il suo porto, 1859-1910 <i>Gaetano Dato</i>	170
Per un approccio sistemico alla rilevazione, misurazione e valutazione dell'applicazione delle tecnologie digitali alla tutela, gestione e valorizzazione del patrimonio culturale <i>Paolo Clini, Pierluigi Feliciati, Ramona Quattrini</i>	173
Semiosi e archeologia del Web. Recuperare eccellenze di fine millennio <i>Alessia Scacchi</i>	176
The electronic linguistic atlas of the Aegean island of Lesbos (EDAL) <i>Angela Ralli, Vaso Alexelli, Charalambos Tsimpouris</i>	179
The “Shape of Monuments” project: traditional methods and new technologies in the Basilica of San Lorenzo Fuori le Mura, Rome <i>Corrado Alvaro, Simone Amici, Jade Bajet, Valeria Danesi, Gian Michele Gerogiannis, Chiara La Marca, Giovanna Liberotti, Daniele Moscone, Antonella Pansini, Enrico Pizzoli, Martina Zinni</i>	183
Tracing Showrunners’ Impact <i>Joanna Byszuk</i>	187
Una proposta di annotazione semantica di concetti musicali nel <i>de institutione musica</i> di Boezio <i>Francesca Michelone</i>	190
<i>Voci della Grande Guerra. Preserving the Digital Memory of World War I</i> <i>Alessandro Lenci, Nicola Labanca, Claudio Marazzini, Simonetta Montemagni, Federico Boschetti, Irene De Felice, Stefano Dei Rossi, Felice Dell’Orletta, Michele Di Giorgio, Lucia Passaro, Giulia Venturi</i>	193
Poster papers	
<i>Aristotele sull’holodeck</i> <i>Luca Bandirali</i>	196
<i>Automatic for the people. Modelli di analisi automatica di testi ed immagini applicati alla ricerca sociale. Un progetto sull’Albania</i> <i>Simone Fagioli</i>	197
Colette, Willy and the <i>Claudine</i> book series: exploitation or collaboration? Authorship attribution in the novel <i>Claudine à l’école (Claudine at school, 1900)</i> <i>Marie Puren</i>	203
Gustave Roud, « Œuvres complètes ». Edizione digitale e RDF <i>Daniel Maggetti, Elena Spadini, Marion Rivoal, Loïc Jaouen, Julien Burri, Alessio Christen, Bruno Pellegrino</i>	205
LETTERE: LETters Transcription Environment for REsearch <i>Giovanni Moretti, Rachele Sprugnoli, Sara Tonelli</i>	207

Modelling between digital and humanities. A project, its research questions and its outcomes	210
<i>Cristina Marras, Arianna Ciula, Øyvind Eide, Patrick Sahle</i>	
On Developing a Virtual Museum of Greek Immigration and Language in Canada	211
<i>Angela Ralli, Charalampos Tsimpouris, Christos Papanagiotou, Tonia Tzanavara</i>	
Physical methods in cuneiform tablet classification - a possibility for a new narrative	214
<i>Jaroslav Valach, Petra Štefcová</i>	
Pinte di storia: un progetto di Public History su Youtube	216
<i>Domenico Matteo Frisone, Antonio Iodice, Michele Lacriola, Edoardo Nicoletti, Pietro Rubini</i>	
Power of Algorithms for Cultural Heritage Classification: The Case of Slovenian Hayracks	218
<i>Ajda Pretnar, Lan Žagar, Blaž Zupan, Dan Podjed</i>	
Scholarly Digital Edition: Manuscripts, Texts and TEI Encoding un corso online per l'Informatica Umanistica	221
<i>Marjorie Burghart</i>	
Stemmatology An R package for the computer-assisted analysis of textual traditions	223
<i>Jean-Baptiste Camps, Florian Cafiero</i>	
Stories in Situ Humanities–informed media design for location–based storytelling	225
<i>Gunnar Liestøl</i>	
The ERC Project “Localizing 4000 Years of Cultural History. Texts and Scripts from Elephantine Island in Egypt”	227
<i>Andrea Hasznos, Ahmed Kamal Mamdoh</i>	
ToposText: Broadening the Digital Humanities Public	230
<i>Brady Kiesling</i>	
TRI: a tool for the diachronic analysis of large corpora and social media	232
<i>Pierpaolo Basile, Annalina Caputo, Giovanni Semeraro</i>	
TriMED: banca dati terminologica multilingue	237
<i>Federica Vezzani, Giorgio Maria Di Nunzio, Geneviève Henrot</i>	
“Two days we have passed with the ancients...”: a Digital Resource of Historical Travel Writings on Italy	242
<i>Rachele Sprugnoli</i>	
Using Formal Ontologies for the Annotation and Study of Literary Texts	246
<i>Fahad Khan, Gloria Mugelli, Federico Boschetti, Francesca Frontini, Andrea Bellandi</i>	
Workshops	
Data Mining through Image Analytics	250
<i>Ajda Pretnar, Lan Žagar</i>	
How to Create a Memory Box with Raspberry PI and RFID Without Being a Programmer	252
<i>Liviu Pop</i>	

Panels

Competenze culturali per un Umanesimo Digitale. La via italiana verso Europa 2018 <i>Nicola Barbuti, Carmine Marinucci, Simonetta Buttò, Fabio Ciotti, Emiliano Degli Innocenti, Luigi Spadari, Giovanna Barni, Fabio Viola, Serge Noiret</i>	254
Common Strategies and Aims of Italian Research Centers for Digital Humanities <i>Fabio Ciraci, Enrica Salvatori, Cristina Marras, Franco Nicolucci, Gianfranco Crupi, Francesca Tomasi, Luigi Catalani, Nicola Barbuti, Fabio Ciotti</i>	255
Archivi digitali d'autore: ipotesi di lavoro <i>Pierluigi Feliciati, Stefano Allegrezza, Ernesto Belisario, Salvatore Vassallo, Emmanuela Carbé, Primo Baldini, Paul Gabriele Weston</i>	256
Digital practices and teaching Humanities <i>Antonio Brusa, Annalisa Caputo, Giuliano De Felice, Isidoro Davide Mortellaro, Claudia Villani</i>	260
Digitalizzare la memoria. Studi, ricerche e sperimentazioni in Terra d'Otranto <i>Giuliana Iurlano, Anna Maria Colaci, Demetrio Ria, Deborah De Blasi, Giovanna Bino, Francesca Salvatore</i>	264
La storia alla prova dei like. Il caso del neoborbonismo mediatico e digitale <i>Annastella Carrino, Gian Luca Fruci, Christopher Calefati, Antonella Fiorio, Maria Teresa Milicia, Federico Palmieri, Carmine Pinto, Silvia Sonetti, Claudia Villani</i>	268
Nuove Competenze e Nuovi Servizi per la Ricerca nella European Open Science Cloud: Il Futuro degli Open Research Data nelle Digital Humanities <i>Giorgio Maria Di Nunzio, Paolo Manghi, Marisol Occioni, Francesca Tomasi, Marialaura Vignocchi</i>	275
<i>Indice degli autori</i>	279

Prefazione

Daria Spampinato

Istituto di Scienze e Tecnologie della Cognizione - CNR
Chair del Comitato di Programma

Il tema principale della Settima Conferenza annuale AIUCD 2018 è Patrimoni culturali nell'era digitale. Memorie, culture umanistiche e tecnologie. Un ambito vasto che accomuna pratiche e settori di studio interdisciplinari e multidisciplinari e che riguarda il ruolo nell'era digitale del patrimonio culturale in tutti i suoi aspetti, il modo nel quale possono essere declinati lo studio e la ricerca scientifica del sapere umanistico (letterario, storico, artistico, archeologico, filosofico) con metodologie, logiche e tecnologie informatiche oltre che con l'uso di diverse tipologie di media e risorse digitali. Riguarda anche lo spinoso problema della memoria dell'evo digitale, il ripensamento della cultura digitale contemporanea come patrimonio culturale del futuro; il modo in cui il digitale sta cambiando la definizione, la valorizzazione, la ricerca, la conservazione dei patrimoni culturali; e la possibilità che la cultura umanistica possa trovare nel digitale un deciso rilancio.

Le conferenze annuali dell'Associazione per l'Informatica Umanistica e la Cultura Digitale si pongono l'obiettivo di far incontrare gli studiosi delle diverse discipline che fanno capo alle cosiddette Digital Humanities: linguisti, storici, archeologi, umanisti digitali, informatici e ingegneri informatici e non solo. Ogni anno la conferenza si propone inoltre di dialogare con altre associazioni del settore per diventare un momento di confronto sempre più ricco e ampio tra gli studiosi. In particolare quest'anno la conferenza si è inserita tra le iniziative della Scuola a rete DiCultHer per il 2018, anno europeo del Patrimonio Culturale, ospitando un panel organizzato in forma congiunta da l'AIUCD e DiCultHer.

La settima edizione della Conferenza nazionale dell'Associazione si è svolta per la prima volta in una città del meridione d'Italia, a Bari, e questo è sintomo della crescente attenzione che l'Associazione riscuote, oltre che di una sua presenza più omogeneamente diffusa sul territorio nazionale. Chair del Comitato di Programma è stata una studiosa di Informatica a conforto del fatto che la comunità dell'informatica umanistica e della cultura digitale non rappresenta soltanto una comunità di umanisti che utilizzano strumenti informatici, ma è diventata realmente interdisciplinare e multidisciplinare.

La Conferenza annuale dell'Associazione sta diventando negli anni sempre più un punto di riferimento, come dimostrano le 86 proposte che sono state presentate: 30 long paper, 39 short paper, 7 long panel, 8 poster e 2 pre-conference workshop. Gli autori sono stati 238, il 74% proveniente dall'Italia e il restante da 15 stati, per lo più europei, anche se sono giunte proposte di relatori dal Canada e dagli Stati Uniti. Le relatrici sono state il 42% del totale. Ci permettiamo inoltre di sottolineare che un

The main topic of the 7th annual AIUCD 2018 Conference is Cultural Heritage in the Digital Age. Memory, Humanities and Technologies. This vast field brings together interdisciplinary and multidisciplinary areas and examines the role of cultural heritage in all its aspects in the digital age. It includes the way in which study and scientific research into the Humanities (literary, historical, artistic, archaeological and philosophical) combines with different methodologies, logic and computer technologies and various types of media and digital resources. It also concerns the critical problem of digital memory, the re-thinking of contemporary digital culture as the cultural heritage of future generations. It also looks at how the digital is changing the definition, enhancement, research and conservation of cultural heritage, as well as the way in which the Humanities can be given a decisive relaunch by the digital world.

The annual conference of the Associazione per l'Informatica Umanistica e la Cultura Digitale, as its core objective, aims to bring together scholars from the diverse fields of the so called Digital Humanities; linguists, historians archaeologists, digital humanists, computer engineers and scientists, and others. Moreover, each year the conference provides a platform for encounters with other associations in the field to offer all those involved an ever more enriching and varied meeting of minds. In particular this year, the conference was included as one of the initiatives of the Scuola a rete DiCultHer 2018, European Year of Cultural Heritage, hosting a joint panel of l'AIUCD e DiCultHer.

The 7th edition of the National conference of the Association, took place for the first time in a city in the South of Italy, Bari, sign of the growing attention that the Association is drawing, as well as a more widespread presence throughout the whole of the country. The Chair of the programme committee was a computer scientist demonstrating the fact that the community of digital humanities and digital culture represents not only researchers and academics in the humanities who use computational tools but has truly become an interdisciplinary, multidisciplinary group.

The annual Conference of the Association is becoming ever more a point of reference, demonstrated by the 86 submissions: 30 long papers, 39 short papers, 7 long panels, 8 posters and 2 pre-conference workshops. The authors of these submissions were 238 altogether, of which 74% were Italian while the others came from 15 other states, mainly European, though two proposals arrived from Canada and the US. The female speakers were 42% of the total. We would also like to stress that

workshop, *How to Create a Memory Box with Raspberry PI and RFID Without Being a Programmer*, è stato organizzato da un autore rumeno, mentre l'altro, *Data Mining through Image Analytics*, da autori slovacchi.

La valutazione delle proposte è stata condotta mediante double-blind peer review da studiosi competenti in discipline umanistiche, informatiche e di Digital Humanities.

Per svolgere le 258 review (tre per ogni proposta) sono stati coinvolti 88 valutatori provenienti da 11 stati (il 75% dall'Italia).

La maggior parte dei valutatori appartiene alle aree CUN (Consiglio Nazionale Nazionale) 10 - Scienze dell'antichità, filologico-letterarie e storico-artistiche e 11 - Scienze storiche, filosofiche, pedagogiche e psicologiche, ambiti in cui l'umanistica digitale è storicamente più forte e che esprimono ancora oggi una forte prevalenza nell'Associazione. Si registra tuttavia anche la presenza anche di valutatori che rientrano nelle aree 8 - Ingegneria civile ed architettura e 12 - Scienze giuridiche e nell'area delle Digital Humanities (attribuita a studiosi stranieri in quanto non ancora riconosciuta dal sistema italiano) e sempre più crescente risulta la presenza (20%) di valutatori delle aree 1 e 9 delle Scienze informatiche e ingegneristiche. Si può notare, inoltre, che gli insiemi degli autori e dei valutatori sono differenti anche se molti studiosi sono sia autori che valutatori.

Dopo il processo di review delle 86 proposte presentate, ne sono state accettate 55 nella stessa tipologia (long, short, ...) in cui erano state presentate, 20 sono state accettate secondo una tipologia differente, mentre 11 non hanno superato la valutazione (pari al 13% del totale). Alla fine il processo di revisione ha portato ad avere: 16 long paper, 33 short paper, 5 long panel e 19 poster e 2 workshop.

La Conferenza ha voluto essere un momento di discussione della comunità italiana dell'Informatica Umanistica; quindi il Comitato di Programma ha ritenuto utile lasciare uno spazio a due panel di confronto, che pertanto non sono stati sottoposti a valutazione. Un panel, dal titolo *Competenze culturali per un Umanesimo Digitale*. La via italiana verso Europa 2018, è stato organizzato in collaborazione con DiCultHer (di cui ho detto sopra). Un altro panel, *Common Strategies and Aims of Italian Research Centers for Digital Humanities*, ha visto il coinvolgimento dei responsabili dei Centri di ricerca di Digital Humanities italiani che hanno fornito spunti e riflessioni sul ruolo che questi centri possono assumere nel quadro dell'organizzazione squisitamente monodisciplinare dei settori accademici in Italia.

Le proposte presentate hanno coperto moltissimi ambiti delle ricerche di Informatica Umanistica, facendone emergere al meglio l'aspetto caratteristico e ineliminabile dell'interdisciplinarietà.

I settori tradizionali in Italia delle Digital Humanities italiane hanno riscontrato una presenza sempre numerosa, con 7 proposte nel settore della Digital Libraries and Cultural Heritage, 6 in Digital editions, 6 in Computational Linguistics, 3 in Stylometry e 4 in Digital Archaeology.

one of the workshops, *How to Create a Memory Box with Raspberry PI and RFID Without Being a Programmer*, was organised by a Romanian author, while another one, *Data Mining through Image Analytics*, was organised by Slovakian authors.

The papers submitted were reviewed through the double-blind peer review method, by scholars competent in disciplines such as humanities, computer science and Digital Humanities.

88 reviewers from 11 different countries (75% Italian) were involved in the task of working through the 258 reviews (three for each proposal).

Most of the reviewers were from the National University Council (CUN) area 10 – which includes studies of antiquity, philology and literature, history or art history, and CUN area 11 – which covers historical studies, philosophy, education and psychology, the areas in which digital humanities is historically stronger and which even today are still prevalent within the Association. However, the presence of reviewers from areas 8 – civil engineering and architecture and 12 – legal sciences and from the area of Digital Humanities (all from outside Italy as the Italian system is yet to recognise this discipline) has also been recorded, and the presence of reviewers from areas 1 and 9 of computer sciences and engineering has also increased (20%). Furthermore, it can be observed that even though many of the scholars are both authors and reviewers, these two sets are still very different.

After the review process, of the 86 proposals, 55 were accepted in their original proposal form (long, short, ...), 20 were accepted in a different form, and 11 did not pass the review process (equal to 13% of the total). The final result was: 16 long papers, 33 short papers, 5 long panels and 19 posters, and 2 workshops.

The Conference wished to give the Italian Digital Humanities Community space for discussion; therefore, the Programme Committee decided to leave room for two free panels, which for this reason did not undergo any assessment. One of the panels, called *Competenze culturali per un Umanesimo Digitale*. La via italiana verso Europa 2018 was organised in collaboration with DiCultHer (as mentioned above). The second, *Common Strategies and Aims of Italian Research Centers for Digital Humanities*, involved the heads of the Italian Research centres for Digital Humanities which provided thoughts and reflections on the role that these centres can play in the organisational framework of the very monodiscipline academic sectors in Italy.

The presented proposals covered many different research areas of the Digital Humanities, helping to highlight the characteristic interdisciplinary approach.

The traditional sectors in Italy of Digital humanities were represented with a high number of papers, with 7 papers about Digital Libraries and Cultural Heritage, 6 on Digital editions, 6 on Computational Linguistics, 3 on Stylometry and 4 on Digital Archaeology.

Both because an interdisciplinary approach has long been the norm and because of the nature of the domain

Sia perché il dialogo interdisciplinare è attivo da più tempo, sia per la natura del dominio e delle metodologie di studio, nella maggior parte degli studi in questi settori si registra una trasformazione più profonda che coinvolge metodi, strumenti, paradigmi ed epistemologie: ormai è consolidato che il digitale non appare solo come strumento di mera rappresentazione, ma anche come strumento di computazione, utile ad acquisire conoscenza attraverso la manipolazione computazionale dei dati e l'utilizzo di nuove metodologie di lavoro. Si evince bene, inoltre, una reale intersezione tra le discipline umanistiche, le scienze dell'informazione e l'utilizzo delle tecnologie digitali. Quindi un ruolo attivo della scienza dell'informazione e non solo del digitale nelle scienze umane.

La Conferenza ha visto la partecipazione anche di interventi in due ambiti di studio meno tradizionali nelle DH italiane, 6 in Linked Data – Semantics and Text Analysis e 4 in Digital History. Entrambi, pur più recenti, hanno già maturato una reale consapevolezza e maturità nell'adozione di metodologie nuove e tipiche del settore interdisciplinare.

Inoltre sono state presentate 4 proposte nell'area degli Archives and social media. La presenza di una grande quantità di dati digitali, risorse e infrastrutture digitali, che forniscono una diversa rappresentazione del sapere, ha permesso infatti l'avvicinamento di altre scienze umanistiche (archivistica, biblioteconomia, information science, scienze sociali) a studi di Digital Humanities. In particolare le scienze sociali, che hanno sempre fatto uso di modelli matematici e statistici, con l'avvento di internet e soprattutto dei social media e la gestione di big data hanno modificato le metodologie di ricerca, utilizzando da un lato tecnologie computazionali in modo sempre più massivo e sfruttando dall'altro i dati digitali come oggetto dei loro studi.

Altro discorso si deve fare per alcune linee di ricerca in cui l'utilizzo delle tecniche multimediali e delle immagini bi e tridimensionali sta assumendo uno spazio assai più ampio. Anche se forse la consapevolezza del ruolo dei metodi analitici computazionali in tali settori non è del tutto compiuta, la presenza, in questa Conferenza, di 5 proposte nel settore Architecture and Geography e di 4 proposte sui Virtual Museums fa ben sperare per il futuro e incoraggia l'avvio di interessanti progetti interdisciplinari.

Esprimo le mie congratulazioni alle giovani studiose che si sono aggiudicate il Premio Giuseppe Gigliozzi, intitolato alla memoria di una delle figure più prominenti nella storia dell'Informatica Umanistica, che consisteva in una borsa di studio per giovani studiosi e studiose che avessero presentato un intervento (esteso o breve). La selezione, avvenuta secondo le modalità previste dalla call for paper, ha visto come vincitrici Francesca Michelone con il lavoro Una proposta di annotazione semantica di concetti musicali nel "De institutione musica" di Boezio e Beatrice Nava con Edizioni digitali di opere a tradizione complessa: il caso del Conte di Carmagnola.

La Conferenza ha avuto una nutrita presenza di partecipanti. Nei 3 giorni a Bari le diverse sessioni sono state seguite da 144 studiosi, di questi, 45 precari o in

and the method of study, much of the research in this sector shows an underlying change involving methods, tools, paradigms and epistemology. It is already clear that digital is not just used as a representative tool but as a computational tool useful for acquiring knowledge through the computational manipulation of data, and the use of new working methods. A real intersection emerges between the humanities, computer science and the use of digital technologies. The result is an active role for information technology, not simply digital elements within the humanities.

Two areas of study which are less traditional in the Italian Digital Humanities were also present at the conference, with 6 presentations in Linked Data - Semantics and text analysis and 4 in Digital history. Both of these are relative newcomers but, have already acquired a real awareness and maturity in their adoption of new methods, typical of interdisciplinary sectors.

There were also 4 proposals from the field of Archives and Social media. The presence of a large quantity of digital data, resources and digital infrastructure which provide a different representation of knowledge, has permitted other sciences of the humanities (archiving, librarianship, information science, social sciences) to approach studies in Digital Humanities. In particular, social sciences have always made use of mathematical models and statistics, with the arrival of the internet and above all with social media, the management of big data has changed methods of research, using on the one hand, increasingly extensive computational techniques and on the other, taking digital data as object of their research.

It is important to mention other lines of research where the use of multimedia techniques and two or three dimensional imaging is taking on a more important role. Even though we should be aware that the role of analytical, computational methods in these sectors are not yet totally developed, the presence of 5 proposals in the sector of Architecture and Geography and of 4 from Virtual museums allows us to hope for the future and encourages the launch of interesting interdisciplinary projects.

I would like to express my congratulations to the young researchers who won the Giuseppe Gigliozzi Prize, named in memory of one of the most prominent figures in the history of Digital Humanities, which gives a grant to young researchers who have presented a paper (long or short). The selection made following the criteria laid down in the Call for Papers, saw Francesca Michelone emerge as one of the winners with her work, Una proposta di annotazione semantica di concetti musicali nel "De institutione musica" di Boezio and Beatrice Nava as the other with Edizioni digitali di opere a tradizione complessa: il caso del Conte di Carmagnola.

The Conference was well attended. In the 3 days in Bari, the various sessions were followed by 144 researchers, of these, 45 without a fulltime contact or still studying and 28 students from the university of Bari, the conference host. Although most of the participants were

formazione e 28 studenti dell'Università di Bari, sede ospitante della Conferenza. Anche se la maggior parte degli studiosi proveniva dall'Italia (il 79%), hanno partecipato anche studiosi provenienti da altri 11 paesi europei e da 2 extraeuropei.

Vorrei ringraziare sentitamente i tre relatori invitati, i professori Paola Buzi, Riccardo Pozzo e Michele Rak, che hanno tenuto le keynote lecture, tutte di notevole rilievo e su aspetti e progetti di interesse europeo, ognuna con un taglio diverso, ma sempre idonee a dare ampio respiro al dibattito scientifico in seno alla Conferenza.

Un caloroso ringraziamento va anche a tutti i membri del Comitato di Programma con cui ho condiviso le diverse scelte e ai valutatori, che con il loro puntuale lavoro hanno svolto la non facile fase della valutazione delle proposte.

Un ultimo sentito ringraziamento va all'Università di Bari e al Comitato di organizzazione locale e in particolare al carissimo prof. Nicola Barbuti che ha saputo offrire un'accoglienza davvero speciale.

from Italy (79%), there were participants from 11 European countries and 2 from outside Europe.

Special thanks go to the invited speakers, Professors Paola Buzi, Riccardo Pozzo and Michele Rak who gave the keynote lectures, all three of the highest quality and dealing with themes and projects of European interest, each one with a different aspect but all 3 giving a wide opening to the scientific debate at the heart of the Conference.

Warm thanks are also extended to all members of the Programme Committee with whom I shared the various choices and also to the reviewers who undertook the difficult task of reviewing the proposed papers.

A final warm vote of thanks to the University of Bari and to the local organising Committee and in particular to Professor Nicola Barbuti who welcomed us with a truly special hospitality.

AIUCD 2018
Keynote lectures

Oltre i tradizionali confini disciplinari. Archeologia, filologia, codicologia e informatica umanistica: quattro discipline in dialogo per la realizzazione di un atlante digitale della letteratura copta. L'esperienza della gestione del progetto ERC 'PATHs'

Prof. Paola Buzi
Università di Roma La Sapienza

ABSTRACT

Il progetto 'PATHs' (dal copto $\rho\alpha\tau$ = "tracce", "impronte") mira a produrre un'efficace rappresentazione diacronica della geografia della produzione letteraria in lingua copta, ovvero del corpus di opere, quasi esclusivamente di contenuto religioso, prodotto in Egitto tra il III e l'XI secolo. Il progetto si basa su un approccio originale e multidisciplinare, combinando, per la prima volta in questo campo, filologia, codicologia, archeologia e digital humanities, allo scopo di esplorare i processi di produzione, copia, uso, diffusione e conservazione delle opere copte, in relazione ai contesti geografico-archeologici d'origine sia dei testi che dei relativi supporti scrittori.

Analizzando testi e contesti, paratesti (titoli e colofoni) e livelli linguistici (stile, dialetti), i prodotti letterari verranno connessi non solo ai luoghi in cui sono stati copiati ma anche ai singoli ambienti intellettuali responsabili della loro creazione.

Nell'ambito di una tradizione manoscritta che offre i più antichi testimoni dell'uso del codice verranno isolati orientamenti culturali e gusti letterari di specifiche aree dell'Egitto ed emergeranno cambiamenti nella manifattura del codice.

Al termine del progetto verrà prodotto un esaustivo atlante digitale dell'Egitto tardoantico e altomedievale, basato su strumenti interattivi, flessibili e versatili, che permetteranno una dettagliata e mirata ricerca, nonché la correlazione di dati cronologici, regionali e tematici.

Tutto ciò illustrerà il rapporto tra insediamenti (sulla base della documentazione archeologica) e produzione intellettuale (sulla base dei manoscritti), e fornirà una nuova prospettiva della diffusione della letteratura e della tradizione manoscritta copta.

Un portale integrerà l'atlante con vari sotto-prodotti, consistenti in banche dati che raccolgono informazioni su autori, opere, manoscritti e siti archeologici.

L'intervento ha lo scopo di illustrare le riflessioni teoriche e le scelte metodologiche che sono alla base di tale progetto.

BIOGRAPHY

Paola Buzi è Professore Ordinario di Egittologia e Civiltà Copta presso la Sapienza Università di Roma. Laureatasi presso lo stesso ateneo nel 1995 - dove ha studiato Egittologia con Alessandro Roccati e Lingua e letteratura copta con Tito Orlandi - ha conseguito nel 2002 il dottorato di Ricerca in Egittologia. Per cinque anni ha poi lavorato presso il Dipartimento di Archeologia dell'Alma Mater Studiorum Università di Bologna in qualità di assegnista di ricerca. Da gennaio a ottobre 2010 ha lavorato presso la Universität Hamburg, su incarico dell'Akademie der Wissenschaften zu Göttingen, allo scopo di catalogare i manoscritti copti conservati in Germania. Dal 1 novembre 2010 al 31 gennaio 2017 è stata Ricercatore a tempo indeterminato. Agli interessi di tipo storico-letterario e codicologico unisce da sempre quelli per la ricerca archeologica, egittologica e coptologica. Ha partecipato a numerosi scavi in Italia e in Egitto. Dal 2008 è condirettore della Missione Archeologica dell'ateneo bolognese e della Sapienza a Bakchias (Fayyum, Egitto) e dal 2014 è anche vice-direttore della missione archeologica a Manqabad (Assyut, Egitto) diretta da Rosanna Pirelli (Università di Napoli "L'Orientale").

È membro dell'Accademia Ambrosiana (Classe di Scienze Africane), dell'Accademia Europaea e del board dell'International Association of Coptic Studies (IACS).

È Senior Research Fellow della Scuola Superiore di Studi Avanzati della Sapienza Università di Roma. È collaboratore scientifico del Corpus dei Manoscritti Copti Letterari (CMCL), impresa diretta da Tito Orlandi. Attualmente dirige il progetto ERC Advanced Grant "PATHs - Tracking Papyrus and Parchment Paths: An Archaeological Atlas of Coptic Literature. Literary Texts in their Geographical Context: Production, Copying, Usage, Dissemination and Storage" con sede presso il Dip. di Storia Culture Religioni della Sapienza.

Saltare la carta, consentire l'accesso, disegnare gli scenari.

Digital humanities: la pianura digitale verso il patrimonio comune europeo

Michele Rak

ABSTRACT

I paesi membri tendono a conferire al loro patrimonio un valore identitario. L'immagine del patrimonio è diversa tra gli stati membri europei ed è diversa tra le comunità che li abitano. Parte delle Università, accorciando i tempi e la qualità delle prove, ha di fatto eliminato la saggistica. I linguaggi digitali sono in grado di ripristinare la circolazione della saggistica e facilitare l'accesso a un patrimonio comune. I processi di digitazione interessano tutti gli aspetti del patrimonio con la formula del facile accesso, dell'incremento della circolazione e dell'acquisizione dei testi e delle immagini degli oggetti/merci.

BIOGRAFIA

Già Professore Ordinario dell'Università di Siena, ha insegnato Sociologia della letteratura presso l'Università di Napoli (1965-1985), Letteratura italiana presso l'Università di Palermo (1985-1989). Ha tenuto la cattedra di Storia della critica e della storiografia letteraria dal 1989 presso l'Università di Siena e di Sociologia dell'arte e della letteratura (Facoltà di Lettere di Arezzo).

Ha studiato gli insiemi testuali della devozione, della comunicazione politica nella città moderna (le feste), i generi (dalla storia della filosofia al romanzo rosa), la genesi del racconto fiabesco nella cultura europea della prima modernità, le dinamiche dei rapporti tra canone e cultura subalterna nel genere della fiaba barocca, i processi storici di affermazione della lingua letteraria napoletana nella cultura europea dal barocco al contemporaneo, il ruolo dei linguaggi d'arte nella rappresentazione sociale e nella sollecitazione del mutamento, il rapporto tra istituzioni culturali e linguaggio politico a Napoli nel Seicento.

What is cultural innovation? How do we measure it?

Prof. Riccardo Pozzo

Università di Verona

ABSTRACT

'Social and Cultural Innovation' is a syntagma that is receiving increased usage among researchers, when it was chosen by the European Strategy Forum Research Infrastructures (2016) for the name of the working group that deals with research infrastructures primarily connected with Social Sciences and the Humanities. Innovation refers to the creation of new products and services by bringing a new idea to the market. Economic growth turns on infrastructures, which provide access to services and knowledge, e.g. by overcoming the digital divide. Globalization has made it clear that a most urgent objective is to work out policies of social and cultural innovation to the advantage of citizens – policies that aim at achieving changes in the regulatory environment that make societies both inclusive and reflective. The Faro Framework Convention on the Value of Cultural Heritage for Society of UNESCO (2007) encourages reflection on the role of citizens in the process of defining, creating, and managing a cultural environment in which communities evolve. The notions of inclusion and reflection are inspired by philosophical ideas referring to the role of deliberative communication of citizens in a modern public sphere aiming at mutual understanding. Jürgen Habermas (1973) has applied to society what G.W.F. Hegel had elaborated as the passage from the surface of being to the ground of essence, a passage that takes place, literally, by reflecting into the thing – like reflected light that illuminates something previously invisible, or creates a pattern not previously existing. It is now time to examine the implications of innovation for redefining the ways in which the Humanities have been envisioned, particularly to visualize the various ways in which philosophy engage with cultural processes in the past, present, and future.

'Social innovation' aims to directly address unmet social needs in new ways by developing or enhancing new products and services through the direct engagement of the people who need and use them, typically through a bottom-up process. It takes place when a new product or service answers positively to the following three questions: (1) Does it solve the problem? (2) Does it have a fair cost? (3) Is it universally accepted? An example of social innovation is the regional healthcare card of the Lombardy region in Italy. It was introduced in 1999 as a pioneer endeavour. It solved the problem of providing access to data; not only did it cost right, but it enabled substantial savings; and it was accepted without major opposition.

'Cultural innovation' sounds like an oxymoron, no doubt. It is not, though. It is something real that tops up social and technological innovation. In fact, integration requires spaces of exchange in which citizens engage in the process of sharing their experiences while appropriating common goods content. We are talking of public spaces such as libraries, museums, science centres, but also of any place in which co-creation activities may occur. At this level, social innovation becomes reflective and generates cultural innovation. Insisting on reflexivity helps to raise awareness for the importance of framing issues around engaging with science and society, identifying problems and defining solutions.

How can we measure 'cultural innovation'? The answer is, as a result of co-creation. Items of cultural innovation are: content providers such as museums, science centres and libraries, as well as processes triggered by issues such as cosmopolitanism, inclusiveness, mobility, migration, heritage and creativity. Measuring the impact is fundamental to improve societal acceptance of public investment in culture, because these measurements may provide a basis for aligning research and innovation with the values, needs and expectations of society. Yet, an accepted methodology is still lacking. The objective of this paper is to develop and test a methodology to measure the impact of technology transfer from basic research to industry application and its cultural and socio-economic effects.

BIOGRAPHY

Riccardo Pozzo è un filosofo e storico della filosofia italiano. Laureatosi in filosofia all'Università Statale di Milano nel 1983, è stato allievo di Mario Dal Pra, Wilhelm Risse e Norbert Hinske. Riceve il dottorato nel 1988 all'Universität des Saarlandes e la libera docenza nel 1995 all'Università di Treviri. Nel 1996 è nominato professore di "Kant and Hegel" alla Catholic University of America a Washington. Nel 2002 si trasferisce alla cattedra di Storia della filosofia dell'Università di Verona. Nel 2009 succede a Tullio Gregory alla direzione dell'Istituto per il Lessico Intellettuale Europeo e Storia delle Idee - CNR a Roma. Nel 2012 diventa direttore del Dipartimento Scienze Umane e sociali - Patrimonio Culturale - CNR. Nel 2008 diventa ambasciatore scientifico per l'Italia della Alexander von Humboldt-Stiftung e viene eletto nel comitato direttivo della Fédération Internationale des Sociétés de Philosophie, dove rappresenta la Società Filosofica Italiana. Nel 2012 entra nel Core Group dello Standing Committee for Humanities della European Science Foundation. Nel 2012 viene eletto all'Institut International de Philosophie.

Storico della filosofia e autore di ricerche sull'aristotelismo, la logica dal Rinascimento a Kant e Hegel, la filosofia del XX secolo, la storia delle idee e la storia delle università.

Ha lavorato in prospettiva interculturale per la FISP e in prospettiva interdisciplinare per l'Istituto per il Lessico Intellettuale Europeo e Storia delle Idee occupandosi del ruolo svolto dalla storia della terminologia di cultura per la formazione dell'identità intellettuale europea. L'oggetto della filosofia sono i testi, e l'ambito "intellettuale" è il linguaggio, con la

conseguenza che di quest'ambito più ampio la storia della filosofia non è più il modello, come voleva l'idealismo, ma solo una provincia. L'obiettivo è creare strumenti per una più rigorosa lettura storica dei testi filosofici e documenti del costituirsi del patrimonio della cultura europea. Specifica caratteristica dell'ILIESI è il ricorso a metodologie informatiche per la documentazione e il trattamento dei dati linguistici e testuali in francese, greco antico, inglese, italiano, latino, spagnolo e tedesco per produrre studi e testi critici, banche dati terminologiche, spogli lessicali, concordanze e lessici d'autore o di un'epoca.

AIUCD 2018
Long papers

A sustainable work flow for a multi-layer edition of the *Chronicon* by Romualdus Salernitanus

Paolo Monella

Università di Palermo, Italy - paolo.monella@gmx.net

ABSTRACT

My experimental scholarly digital edition of the *De nomine* by Ursus Beneventanus tested the feasibility of the edition model theorized by Orlandi 2010, based on three different layers of text representation (graphematic, alphabetic and linguistic) and on a table of all graphemes having distinctive value in the graphic system of a specific manuscript (“graphematic table of signs”). Its work flow, however, proved to be very time-consuming. This talk analyzes the specific practices of the Ursus edition that mostly slowed down the work flow and outlines possible solutions to be applied in my ongoing digital edition of the *Chronicon* by Romualdus Salernitanus. Those aspects include (a) the markup of abbreviations, which can be expedited by taking advantage of the systematic nature of ancient abbreviations, (b) the markup of ancient punctuation, which will be omitted altogether, and (c) the representation of the linguistic layer. The minimal units of the latter are inflected words regardless of any specific spelling. In the Ursus edition, each word was represented at this layer by a combination of lemma and morphological information (e.g.: ablative plural of lemma “*praepositio, -onis*”), by means of @lemma and @ana attributes of <w>. The values of those attributes were populated by the lemmatizer/PoS tagger *TreeTagger*, but needed to be reviewed manually. The Romualdus edition will fall back on the common practice of representing words at the linguistic layer by means of their “normalized” spelling. Lastly, while in the Ursus project the TEI-to-HTML transformation was performed dynamically by JavaScript, in the Romualdus edition a Python script will perform this task statically.

KEYWORDS

Digital scholarly edition, layers, Romualdus Salernitanus, Ursus Beneventanus, manuscript, TEI XML, lemmatization, PoS tagging, Python

1. GOAL OF THE TALK

This talk will first shortly analyze the main issue of my edition of the the *De nomine* by Ursus Beneventanus (<http://www.unipa.it/paolo.monella/ursus/>; see Monella 2006): the excessive time-intensiveness of the work flow. It will then provide strategies to expedite the production of the edition of the *Chronicon* by Romualdus Salernitanus (XII Century; see Garufi 1914, Matthew 1981, Zabbia 2004), which I have recently started to work on, in the framework the ALIM project (<http://alim.unisi.it/>).

2. METHODOLOGY

My experimental digital scholarly edition of the *De nomine* tested the feasibility of the edition model theorized by Orlandi 2010. The Romualdus edition will be based on the same methodological principles, including

1. the representation of the text on three different layers (graphematic, alphabetic and linguistic) to integrate “diplomatic” and “interpretive” edition (see Haugen 2004, Driscoll 2006, Huitfeldt 2006, Orlandi 2006 and 2010, Brüning et al. 2013, Monella 2014, Pierazzo 2015) and
2. a table of all graphemes having distinctive value in the graphic system of a specific manuscript (“graphematic table of signs”).

Further details on the edition model are in Monella 2016 and in the project documentation on the edition website.

3. RESULTS EXPECTED

The Ursus project aimed to be a proof of concept, an experiment to test those theoretical and practical issues that would only arise from a real-world application of Orlandi’s ideas. The Romualdus edition sets out the specific goal of finding a balance between methodological sophistication and work flow sustainability. In other words, it aims at implementing as many features of the Ursus edition model as possible within the project’s three years time frame.

4. ISSUES OF THE URSUS EDITION

Two specific issues of the Ursus project were (a) the extent of the text actually published, compared with the time needed, and (b) the fact that I could not fully review the encoding of the linguistic layer.

Extent of the edition

In two years I produced an edition of 11 folios of manuscript *Casanatensis* 1086, the *codex unicus* of the grammatical works by Ursus Beneventanus, on three layers (graphemes, alphabetic letters and inflected words): not much. Possible causes of such a long elaboration time include the following:

- I was the only contributor on the project while also working full-time as a school teacher;
- This edition model had never been applied before (except for Orlandi 2006);

- I wrote all the software without any formal training as a professional programmer;
- There was no OCR-ed base text to start from, because the work was unpublished – though it consisted in a paraphrased summarization of known grammatical sources;
- The manuscript was often very much faded out.

Review of the linguistic layer

In addition, by the end of the research project I did not get to fully review the “linguistic layer” of the edition. At this layer each word was represented by a combination of lemma and morphological information. For example, inflected word “prepositionibus” was represented as `<w ana="11C---02---" lemma="praepositio" n="praepositionibus" xml:id="w20673">`, where 11C---02--- means “Nominal, Positive, III decl, Plural Ablative, Feminine”. The values of the @lemma and @ana attributes were generated by lemmatizer/PoS tagger *TreeTagger* (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>) and were correct for about 93% of words. I had planned to review those values to correct misanalyzed words, but as of fall 2017 I could not complete this final stage of the work, which is not likely to be very short.

In any case, a simpler representation of each inflected word at the linguistic layer is in the @n attribute, providing its “normalized” transcription.

5. WHAT TOOK YOU SO LONG? AN ANALYSIS AND SOME POSSIBLE SOLUTIONS

I shall now analyze in detail the specific aspects of the Ursus edition that mostly slowed down the work flow, and discuss the solutions that I am planning to apply in the Romualdus edition:

Graphematic layer: punctuation

Issue. In the Ursus edition, I encoded the original grapheme of each punctuation sign and marked its value (strong or weak separation, highlighting of special words, pause in the intonation etc.) with an @n attribute. In the following example `<pc n="0">.</pc>`

the “0” value of @n means that the punctuation sign does not mark any syntactical separation or pause in the sentence.

This was, like many aspects of the Ursus edition model, an experiment to explore and question our notions of graphic system, grapheme, immediate “meaning” of a grapheme. However, it was quite laborious.

Solution. Having explored the feasibility and the methodological issues of such a detailed markup, I will simply skip the transcription of such paragraphematic signs in the Romualdus edition. Those signs can be ascribed to a separate “sub-system”, in Orlandi’s terminology, within the whole graphic system of the manuscript. Choosing which objects to encode and the manner with which to encode them is subjective and always legitimate, as long as the editor declares and discusses his or her choices in the project documentation.

Graphematic layer: abbreviations

Issue. For Ursus, at the graphematic layer I encoded each of the graphemes composing an abbreviation (e.g. the brevigraph “p” for “per”, or “p” plus macron for “pre”) and marked each component of that abbreviation with the specific markup (`<choice>/<abbr>/<am>/<expan>`) as follows:

```
<w ana="11C---02---" lemma="praepositionibus" n="praepositionibus" xml:id="w20673">
  <choice>
    <abbr type="superscription">p<am>̄</am></abbr>
    <expan>pre</expan>
  </choice>
  positioni
  <choice>
    <abbr type="after">b<am>;</am></abbr>
    <expan>bus</expan>
  </choice>
</w>
```

I created custom “markdown-like” conventions to speed up the typing. For example, to generate the above code I would simply type

, p, -, prae, positioni, b, :, bus,

where commas separate the elements of the abbreviation (base grapheme, abbreviation mark, alphabetic expansion). A Python script would parse the string above and generate the TEI XML code.

Nevertheless, the encoding of each abbreviation in the manuscript slowed down the transcription, especially since I was not able to start from an OCR text, but I was typing the transcription anew without the help of any previous print edition.

I should note that this issue is not specific to my edition model, but affects all projects transcribing the fairly ubiquitous abbreviations of ancient handwritten primary sources.

Solution. I am not encoding all paragraphs of the text at both the graphematic and the alphabetic layers. For some paragraphs, I am only providing an alphabetic transcription, which is not very difficult to create starting from the OCR of the Garufi 1914 edition.

The `@decls` attribute of `<p>` marks the paragraphs encoded at the alphabetic layer only: `<p decls="#a">`. The reference `#a` points to an `<editorialDecl>` element in the `<encodingDesc>` of the TEI header:

```
<encodingDesc>
  <editorialDecl xml:id="ag" default="true">
    <p>Paragraph encoded at the Graphematic and Alphabetic Layers.</p>
  </editorialDecl>
  <editorialDecl xml:id="a">
    <p>Paragraph encoded at the Alphabetic Layer only.</p>
  </editorialDecl>
</encodingDesc>
```

Those elements encoded at both the graphematic and alphabetic layers might be marked with `<p decls="#ag">`, but this is not necessary because of the `@default="true"` attribute above: this means that all paragraphs having no `@decls` attribute at all, default to `@decls="#ag"` (encoded at both layers).

In this way, while the edition model framework remains the multi-layered one of the Ursus edition, each section of each manuscript transcription can flexibly implement the model in full (all layers) or in part, based on specific needs, as well as on the project scope and time frame.

However, those paragraphs that are encoded also at the graphematic layer still need a faster and more efficient encoding strategy for abbreviations.

A simple solution comes from the fact that the abbreviation system of pre-modern handwriting... is systematic. Most of the times, a specific abbreviation (e.g. “p” or “p” plus macron) means a specific sequence of alphabetic letters (“per” or “pre” respectively). The practice that I am following in the Romualdus edition (and will document in detail in the documentation) takes advantage of the systematic nature of ancient abbreviations:

- I created a CSV table for each manuscript of Romualdus’ work (“table of standard abbreviation combinations”) mapping common abbreviation combinations (e.g. “p” plus macron) to their standard alphabetic value (“pre”). The alphabetic meaning of one-glyph brevigraphs such as “p” (“pro”) is already provided in the “graphematic table of signs”.
- In the source TEI XML of the transcription, when an abbreviation in a specific point of the text has its standard meaning/expansion I only encode its graphemes (e.g. “p0”, where “0” is the Unicode character chosen for the *encoding* – not the visualization – of the macron). In this case, I do not mark the abbreviation with `<choice>` / `<abbr>` / `<am>` / `<expan>` in the source code because the software (a Python script), based on the documentation, can easily check the CSV “abbreviation combinations” file and identify the string “p0” as a standard abbreviation to be expanded to “pre” at the alphabetic layer.
- If, instead, an abbreviation does not have a standard alphabetic value as mapped in the CSV “table of standard abbreviation combinations”, I encode the abbreviation in full with `<choice>` / `<abbr>` / `<am>` / `<expan>`. Those cases, however, represent a minority of the abbreviations actually found in a manuscript.

I consider this faster practice equivalent, in terms of information content, to the practice (followed for Ursus) of encoding all abbreviations with `<abbr>`, since the CSV table combined with regular expression (regex) matching software ensures that abbreviations are formally identified and represented at the graphematic and alphabetic layers.

Linguistic layer: lemma/PoS tagging or normalized spelling?

Issue. The new approach tested in the Ursus edition to identify inflected words at the linguistic layer (through `@lemma` and `@ana` attributes of `<w>`) had a shortcoming: the amount of work needed to review the output of *TreeTagger*, the lemmatizer/PoS tagger.

Falling back on the usual approach to representing of this textual layer, one might simply provide a “normalized” spelling of each word (the function of the `@n` attribute in `<w>` in the Ursus project), as in `<w n="usque">usq9</w>`. This certainly is a much more straightforward encoding process, to the point that it may be considered complete for the Ursus edition itself. But what would be the value of `@n` – the “normalized” transcription – of “prepositionibus” (written in the manuscript with the medieval spelling “-e-” instead of classical “-ae-”)? Choosing classical “*praepositionibus*” is disputable from the point of view of cultural history, while choosing “prepositionibus” would bring about those issues connected with alternative spelling, which affect all computational methods of textual analysis (search, indexing, collation, lemmatization etc.).

Solution. A Draconian shortcut might consist in disposing of the linguistic layer altogether. With medieval Latin texts, the alphabetic layer is probably sufficient enough to provide scholars with a readable text. But for textual searches, indexing, automatic collation, lemmatization and any other form of textual analysis, it proves inadequate. For the time being, I am

applying the “Draconian” solution of omitting the linguistic layer. If time will suffice, I will apply the “normalized spelling” solution (<w n=“usque”>usq9</w>) at a later stage of the project.

6. SOFTWARE AND LICENSES

In the Ursus project, a large JS script processed the TEI XML code and the CSV “graphemic table of signs”. The script manipulated the DOM of an HTML file dynamically in the browser and visualized the edition. However, the browsers required 7-10 seconds to load the page. In the Romualdus edition all such tasks will be performed statically by a Python script using the lxml library.

All files and software of the new edition will be, as it was in the Ursus project, open source and designed with interoperability and reuse in mind. As a blind reviewer of this abstract correctly pointed out, the Ursus software as such was “too project oriented to be reused by other scholars in other contexts”. This may be the case for the Romualdus software as well, as its tasks are narrow: all it has to do is generating an HTML visualization of the TEI XML transcription, based on the CSV tables of signs and of standard abbreviation combinations. My aim, in any case, is to allow scholars to read the Ursus and Romualdus software source code, so they can find ideas to write their own code, should they create other multi-layered editions following a similar model.

7. REFERENCES

- [1] Brüning, Gerrit, Katrin Henzel, e Dietmar Pravida. 2013. «Multiple encoding in genetic editions: the case of “Faust”». *Journal of the Text Encoding Initiative*, n. 4. doi:10.4000/jtei.697.
- [2] Driscoll, Matthew J. 2006. «Levels of Transcription». In *Electronic Textual Editing*, a cura di Lou Burnard, Katherine O’Brien O’Keefe, e John Unsworth. Modern Language Association of America. http://www.tei-c.org/About/Archive_new/ETE/Preview/driscoll.xml.
- [3] Garufi, Carlo Alberto, a c. di. 1914. *Romualdi Salernitani Chronicon (A.m. 130-A.C. 1178)*. Vol. 127. *Rerum italicarum scriptores*. Città di Castello: S. Lapi.
- [4] Haugen, Odd Einar. 2004. «Parallel Views: Multi-level Encoding of Medieval Nordic Primary Sources». *Literary and Linguistic Computing* 19 (1): 73–91. doi:10.1093/lc/19.1.73.
- [5] Huitfeldt, Claus. 2006. «Philosophy Case Study». In *Electronic Textual Editing*, a cura di Lou Burnard, Katherine O’Brien O’Keefe, e John Unsworth. Modern Language Association of America. http://www.tei-c.org/About/Archive_new/ETE/Preview/huitfeldt.xml.
- [6] Matthew, Donald J. A. 1981. «The Chronicle of Romuald of Salerno». In *The Writing of History in the Middle Ages: Essays Presented to Richard William Southern*, a cura di Ralph Henry Carless Davis, 239–74. Oxford: Oxford University Press.
- [7] Monella, Paolo. 2014. «Many Witnesses, Many Layers: The Digital Scholarly Edition of the Iudicium Coci et Pistoris (Anth. Lat. 199 Riese)». A cura di Fabio Ciotti. *Digital Humanities: Progetti Italiani Ed Esperienze Di Convergenza Multidisciplinare, Atti Del Convegno Annuale Dell’Associazione per l’Informatica Umanistica e La Cultura Digitale (AIUCD) Firenze, 13-14 Dicembre 2012*, 173–206. doi:10.13133/978-88-98533-27-5. http://digilab2.let.uniroma1.it/ojs/index.php/Quaderni_DigiLab/article/view/190.
- [8] Monella, Paolo. 2016. «Livelli di rappresentazione del testo nell’edizione del De nomine di Orso Beneventano». In *AIUCD 2016 Book of Abstracts. Digital editions: representation, interoperability, text analysis and infrastructures*, a cura di Federico Boschetti, 53–56. AIUCD. <http://www.himeros.eu/aiucd2016/c20.pdf>.
- [9] Mordenti, Raul. 2011. *Paradosis. A proposito del testo informatico*. Memorie lincee Scienze morali, storiche, filologiche IX. Roma: Accademia Nazionale dei Lincei.
- [10] Orlandi, Tito. 2006. «Edizione digitale sperimentale di Niccolò Machiavelli, De principatibus». <http://www.cmcl.it/~orlandi/principe/>.
- [11] ———. 2010. *Informatica testuale. Teoria e prassi*. Roma: Laterza.
- [12] Pierazzo, Elena. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham, Surrey, UK and Burlington, VT: Ashgate.
- [13] Robinson, Peter. 2014. «Exactly what are we transcribing?» *Textual Communities*. <http://www.textualcommunities.usask.ca/web/canterbury-theses/blog/-/blogs/exactly-what-are-we-transcribing->.
- [14] Sampson, G. 1990. *Writing Systems: A Linguistic Introduction*. Stanford University Press.
- [15] Stokes, Peter A. 2011. «Describing Handwriting, Part IV: Recapitulation and Formal Model». *DigiPal Blog*. <http://www.digipal.eu/blogs/blog/describing-handwriting-part-iv/>.
- [16] Zabbia, Marino. 2004. «Romualdo Guarna arcivescovo di Salerno e la sua cronaca». In *Salerno nel XII secolo. Istituzioni, società, cultura*, a cura di Paolo Delogu e Paolo Peduto, 380–98. Salerno: Provincia di Salerno - Centro studi salernitani «Raffaele Guariglia».

AMiGre: A unified framework for archiving and processing oral and written dialectal data

Eleni Galiotou¹, Angela Ralli²

¹ Technological Educational Institute of Athens, Greece - egali@teiath.gr

² University of Patras, Greece - ralli@upatras.gr

ABSTRACT

Asia Minor Greek constitutes an ideal case study for historical linguistics and language contact. It offers testimonies for tracing the evolution of the Greek language and displays numerous borrowings, due to its long-lasting contact with Turkish. Its dialects belong to a rich cultural and linguistic heritage but are faced with the threat of extinction. Consequently, the description and preservation of this invaluable heritage is of crucial importance, the sustainability and awareness of which can be done in a decisive way with the use of modern Information Technology. In this paper, we deal with the problem of archiving and processing digitized corpora of oral and written data of three dialects of Asia Minor Greek, Pontic, Cappadocian, and Aivaliot, that have been compiled in the frame of the research program AMiGre (THALIS, 2012-2015), cofinanced by the European Union and Greek national funds. The project aimed at: (a) providing a systematic and comprehensive study of Pontic, Cappadocian and Aivaliot, three Asia Minor dialects of common origin and parallel evolution, which are also in danger of extinction; (b) digitizing, archiving and processing a wide range of oral and written data, thus contributing to the sustainability and awareness of this longwinded cultural heritage. The project resulted in the scientific presentation of data from Asia Minor Greek to the academia, under the form of multimedia corpora, an electronic tri-dialectal dictionary and a number of publications. The corpora of oral and written data are stored in a multimedia database which allows the parallel visualization of raw and processed dialectal data as well as the coding of a large number of metadata. The search and retrieve subsystem provides a combined search at two levels of linguistic representation, phonological and morphological, access in metadata and combined search in both oral and written resources.

KEYWORDS

Asia Minor Greek, dialects, multimedia databases, electronic corpora

1. INTRODUCTION

In recent years, there has been a growing interest in the use of Information Technology for the study of language contact and dialectal change. In fact, the availability of dialectal data on electronic media and the development of computational tools has contributed in a decisive way to the sustainability and awareness of this invaluable cultural heritage. To this end, several attempts to build dialectal corpora are reported. Indicatively, a multimedia database was used in order to create a voice language map of Japanese dialects [12], and the Linguistic Atlas of Middle and South Atlantic States (LAMSAS) was created in order to archive and process dialectal data from the Atlantic coast of the United States [10]. As far as European languages are concerned, several attempts are also reported. Indicatively, the Scottish Corpus of Text and Speech (SCOTS) is the outcome of a project aiming at constructing a large-scale electronic corpus of oral and written texts of the languages of Scotland [1]. A set of Natural Language Processing tools for the processing of Swiss German oral dialectal data is described in [11]. Similarly, the Corpus Oral Dialectal (COD) – a corpus of contemporary Catalan – is provided in [13]. The DynaSAND (Dynamic Syntactic Atlas of Dutch Dialects), an on-line tool for the processing of Dutch syntactic variation ([3]) is enriched and enhanced with a web service interface to the DynaSAND corpus, so that the data from the corpus can be used for other applications as well [9]. The first attempt in Greece to combine Information Technology and Theoretical Linguistics in order to present both raw and processed dialectal material on digital space, with the use of the most up-to-date software programs was the “THALIS” project “Pontus, Cappadocia, Aivali: in search of Asia Minor Greek (AMiGre¹)”. The project resulted in the scientific presentation of Greek dialectal data to the academia under the form of a multimedia corpus, an electronic tri-dialectal dictionary [5] and a number of publications. In fact, the aim of this project was twofold: (a) to provide a systematic and comprehensive study of Pontic, Cappadocian and Aivaliot, three Greek dialects of Asia Minor of common origin and of parallel evolution that are faced with the threat of extinction; (b) to digitize, archive and process a wide range of oral and written data, thus contributing to the sustainability and awareness of this longwinded cultural heritage.

This paper, focuses on the issue of storing and retrieving dialectal corpora on electronic media. In particular, we present the design and development of a multimedia software and database for archiving and processing oral and written dialectal data which were compiled in the course of the AMiGre project [4].

¹ This research was co-financed by the European Union (European Social Fund–ESF) and Greek national funds through the Operational Program "Education and Life-long Learning" of the National Strategic Reference framework (NSRF) - Research Funding Program: “THALIS. Investing in knowledge society” through the European Social Fund. The project was directed by Prof. Angela Ralli and hosted at the *Laboratory of Modern Greek Dialects* of the University of Patras.

2. THE ORAL AND WRITTEN CORPORA

The corpora of the AMiGre project consist of both written and oral data, which have been collected through archival search and field work, respectively. The collection was granted the written permission of the owners of written material or that of the informants, depending on the case, and had the approval of the Ethics committee of the University of Patras. They contain 180 hours of recorded raw data (narratives) [8], and digitized texts of 1,000,000 words, compiled from primary written sources which were detected in various archives of the 19th and early 20th centuries [8]. The corpora constitute the Asia Minor Archive (AMiGre, URL: <http://amigre.upatras.gr>), which is stored at the server of the *Laboratory of Modern Greek Dialects* (LMGD) of the University of Patras (<http://lmgd.philology.upatras.gr>), the personnel of which assures its long-term maintenance and preservation.² From these corpora, all written texts and a subset of 15 hours (5 hours per dialect) of the oral data are stored in an archive accessed electronically at the address <http://amigredb.philology.upatras.gr>. This material was annotated as for some metadata, for instance, information is given about the source, the informant and the conditions of collection. Note that from the written material, a subset of 50.000 words and the 15 hours of the oral material have been further annotated, translated and linguistically analyzed as far as phonetics/phonology and morphology are concerned. All processed written and oral data, together with those of the rest of raw oral material are accessible only to the researchers of LMGD, through a specific application. It is worth mentioning that the corpora were processed by using ELAN³ for multimodal annotation and phonetically analyzed with the use of Praat,⁴ resulting in the representation of vowels, diphthongs, consonants and consonant clusters appearing on different layers (tiers). One more tier was added, that of morphological representation, consisting of word internal morphological segmentation, constituent recognition and categorization. Segments and consonants were transcribed with the IPA⁵ symbols, while morphological words and syllables with the use of the SAMPA alphabet [14].

An advanced software tool such as Labb-CAT⁶ which provides the user with the possibility to store audio or video recordings, text transcripts and other annotations should be able to deal with the variety of linguistic information and annotation types. Yet, the system in question could not deal with our basic requirements, that is, (a) annotations at different linguistic levels, and (b) combined search at both the oral and written corpora. Consequently, we opted for the design and implementation of a software which would be tailored to our needs [6, 7].

3. THE MULTIMEDIA DATABASE

Our system comprises four collections of data files: digital recordings of oral data (WAV files), initial annotations of oral data (TextGrid files – output of Praat), digitized pages from the original texts (Image files), and transcriptions of pages of written resources (Transcribed Written Sources). Editable elements are stored in three databases: (a) “Struct” database: implements the abstract hierarchical structure common to both oral and written resources; (b) “EAV” database: it follows the “Entity-Attribute-Value” schema [2], since it comprises an extendable set of entities and attributes; (c) “Inner” database: it also follows the EAV schema and contains word fragments and their annotations. Our system is built around two basic subsystems: “G. Oral” (Graphical User Interface for Oral resources) and “G. Written” (GUI for Written resources), which invoke a number of web-like applications related to the processing of oral and written resources respectively. These applications comprise Taggers at different levels (Phonological, Morphological, Syntactic, Semantic) –although only the first two were used for the AMiGre purposes– Preview of Image resources, Transcription and Annotation modules. It also comprises a module for storing and updating metadata. An example of an overview of written data is depicted in Figure 1. The interface is in the form of a triptych; the left panel depicts a transcript page of the document, the middle one shows the list of morphological words and the right one contains the annotations of the highlighted word. A similar approach is followed when oral data are treated. Finally, the “Search and Retrieve” module invokes all the web-like applications for a combined research in both oral and written data. It allows the user to combine multiple criteria, define restrictions (value, distance) and focus on a certain level (document, part, etc.). The form of the query follows the template which is depicted in Figure 2.

4. CONCLUSIONS

In this paper, we have presented a system of archiving and processing oral and written dialectal corpora in a unified framework. We have described the design and development of a multimedia software and database for the preservation and processing of both oral and written data from three Greek dialects of Asia Minor, Pontic, Cappadocian and Aivaliot. Our system enables the parallel visualization of raw and annotated data at different levels, as well as the coding of a large number of metadata. The search and retrieve subsystem provides (a) a combined search at two levels of linguistic representation (phonological, morphological), (b) access in metadata and (c) combined search in both oral and written resources.

² One of the products, the tri-dialectal dictionary (LEPOKAM) consisting of 7,500 dialectal lemmas of Pontic, Aivaliot and Cappadocian is already available to researchers, upon request and the permission of the director of the project, Prof. Angela Ralli (<http://lepokam.philology.upatras.gr>).

³ <http://tla.mpi.nl/tools/tla-tools/elan/>

⁴ <http://www.praat.org>

⁵ <http://www.langsci.ucl.ac.uk/ipa/ipachart.html>

⁶ <http://onzeiminer.sourceforge.net/>

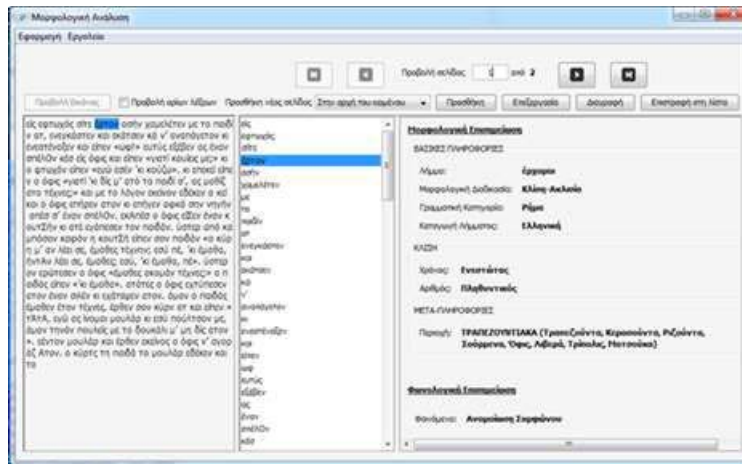


Figure 1. Triptych of written documents

Word/ token /phenomenon			Location				
<Value>	{Between, And, Or, Exact}	<Value>	<EAV_underscore>	<Attribute>	<Part distances>	<Word distances>	<Interval_no distances>

Figure 2. Query template

5. ACKNOWLEDGEMENTS

We wish to thank the following members of the research team of the AMiGre project, Prof. Nikitas Karanikolas (Department of Informatics, TEI of Athens), Assoc. Prof. Dimitris Papazachariou and Ass. Prof. Thodoris Markopoulos (Department of Philology, University of Patras), Prof. Nikolaos Pantelidis (University of Athens), Dr. Io Manolessou (Academy of Athens) and Dr. Nikos Koutsoukos (post-doctoral researcher at LMGD) who have helped significantly into setting the requirements for the design of the multimedia database.

6. REFERENCES

- [1] Anderson, J., Beavan, D., Kay, C. 2007. SCOTS: Scottish Corpus of Texts and Speech. In *Creating and digitalizing Language Corpora Vol.1*, Beal J. (ed.). Palgrave MacMillan Publication, pp. 17-34.
- [2] Anhøj, J. (2003), "Generic Design of Web-Based Clinical Databases", *Journal Medical Internet Research*, 4. DOI=<http://dx.doi.org/10.2196/jmir.5.4.e27>
- [3] Barbiers, S. et al. 2006. *Dynamic Syntactic Atlas of the Dutch dialects (DynaSAND)*. Amsterdam, Meertens Institute. <http://www.meertens.knaw.nl/sand/>
- [4] Galiotou, E., Karanikolas, N.N., Manolessou, I., Pantelidis, N., Papazachariou, D., Ralli, A., Xydopoulos, G. 2014. *Asia Minor Greek: Towards a Computational Processing*, *Procedia: Social and Behavioral Sciences*, 147, Elsevier, 458-466. DOI=<http://dx.doi.org/10.1016/j.sbspro.2014.07.138>
- [5] Karanikolas, N.N., Galiotou, E., Xydopoulos, G., Ralli, A., Athanasakos, K., Koronakis, G. 2013. Structuring a Multimedia tridialectal dictionary, In *Proceedings of the 16th Int. Conference on Text, Speech and Dialogue TSD 2013 (Plzeň, Sept. 1–5 2013)* LNCS vol. 8082, Springer, 509-518. DOI=http://dx.doi.org/10.1007/978-3-642-40585-3_64
- [6] Karanikolas, N.N., Galiotou, E., Ralli, A. 2014. Towards a unified exploitation of electronic dialectal corpora: Problems and perspectives, In *Proceedings of the 17th Int. Conference TSD 2014 (Brno, Sept. 8-12 2014)*, LNAI vol. 8 655, Springer, 257-266. DOI=<http://dx.doi.org/10.1007/978-3-319-10816-2>
- [7] Karanikolas, N.N., Galiotou, E., Papazachariou, D., Athanasakos, K., Koronakis, G., and Ralli, A., *Towards a computational processing of oral dialectal data*. PCI 2015, October 01 - 03, 2015, Athens, Greece. ACM 978-1-4503-3551-5, DOI=<http://dx.doi.org/10.1145/2801948.2801966>
- [8] Karasimos A., Galiotou E., Karanikolas, N., Koronakis, G., Athanasakos, K., Papazachariou, D., and Ralli, A., *Challenges of Annotating a Multi-Dialect, Multi-Level Corpus of Spoken and Written Modern Greek Dialects*. MGDLT6: 6th International Conference on Modern Greek Dialects & Linguistic Theory, 25-28 Sept. 2014, Patras, Greece.
- [9] Kunst, J.P., and Wesseling, F. 2010. *Dialect Corpora Taken Further: The DynaSAND corpus and its application in newer tools*. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation (Nov. 4-7, 2010)*, Tohoku University, 759-767. <http://www.aclweb.org/anthology/Y10-1088>
- [10] Nerbonne, J., Kleiweg, P.: *Lexical distance in LAMSAS*. *Computers and the Humanities*, 37(3), 339–357 (2003)
- [11] Scherrer, Y. 2010. *Natural Language Processing for Swiss German Dialects*. In: the 55th Annual Conference of the International Linguistic Association, New Paltz, NY (USA), April 2010. <http://archive-ouverte.unige.ch/unige:22810>
- [12] Ubul, A., Kake, H., Sakoguchi, Y., Kishie, S. 2015. *Research on Oral Map in Regional Dialect Using Google Map*. *Int. Jour. Comp. ch. 2 (2)*, 31-35. <http://ijcat.org/IJCAT-2015/2-2/Research-on-Oral-Map-in-Regional-Dialect-Using-Google-Map.pdf>
- [13] Valls, E., Nerbonne, J., Prokić, J., Wieling, M., Clua, E. and Lloret, M-R. 2012. *Applying Levenshtein Distance to Catalan Dialects. A Brief Comparison of Two Dialectometric Approaches*. *Verba* 39, 35-61.
- [14] Wells, J.C. (1997). 'SAMPA computer readable phonetic alphabet'. in Gibbon, D., Moore, R., and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin & New York: Mouton de Gruyter. Part IV, section B.

Approaches to the German reception of Verdi's »Messa da Requiem« through metadata analysis and »horizontal reading«

Torsten Roeder

University of Würzburg, Germany - torsten.roeder@uni-wuerzburg.de

ABSTRACT

My recently submitted dissertation deals with the German reception of Giuseppe Verdi's »Messa da Requiem« in the period of its first performances (1874–1878). Due to the scenic character of the music, the work was condemned by German critics as inappropriate. As a basis of the musicological analysis, a number of music critiques from music journals and local newspapers were compiled into a TEI corpus of 330 texts. Methodically, the work is linked to preceding research on dichotomy analysis and history of reception, and was refined using the digital material. I applied a method which I call »horizontal reading«, which means reading all occurring contexts of one single semantic unit in parallel, using a text database. The views through the corpus open up the possibility of working comparatively and identifying recurring patterns in the reception. While conventional research was limited to individual authors, the new method can be used to compare many different receptions to each other at the same time. Not only did general differences between the German Reich and Austria come to light, but also regional or local facets. The digital material which was used extensively for my thesis offers great potential for further analysis and is published freely to stimulate research.

KEYWORDS

reception, music history, music critique, text analysis, reading methods

1. INTRODUCTION

My recently submitted dissertation dealt with the German reception of the »Messa da Requiem« by Giuseppe Verdi in the years of its first performances (1874–1878). Due to the scenic character of its music, the work was repeatedly condemned by critics in the German Empire as inappropriate (»Verdi did not compose a requiem, according to the German style«). This is related to Wagnerism and German »Kulturkampf«, but also to the contemporary concept of musical »realism« and Romantic church music aesthetics [1]. Today, the ambivalence of that composition stills seems to float in a gray zone between spiritual and operatic music, which leads to discussions about appropriate interpretation. Reflecting on historical conceptions and concepts gives us the opportunity to question our current positions and their origins.

The groundwork for my research is a compilation of German-language articles on the »Messa da Requiem« from the 1870s. I intended to provide a broad basis for the investigation without pre-selecting only prominent authors. This is a unique feature in comparison to previous investigations [2]. The only two criteria for the selection of the texts were German language and a reference of the Messa da Requiem within the investigated period. Altogether, the final corpus consists of 320 texts from German music magazines and daily newspapers, transcribed into TEI and totaling 950,000 characters of text. The material was collected through online resources (about 250 texts, mostly from ANNO, SPO, Europeana and Wikisource) and archives (about 70 texts).

One of the basic principles of reception historiography is the necessity to go beyond facts and investigate into the backgrounds and conditions of reception. However, using newspapers as sources, we mostly do not know the authors of the texts, or only by acronym or pseudonym. Even when an author's name could be deduced, it is still difficult to find background information on that person. My approach here is to conduct a comprehensive research on a large corpus of text that allows to arrange the opinions and comments of the authors – or better: of their texts – in a spectrum, to form a broader picture of the diversity of opinions and ideas.

Two digital methods significantly supported my research. Metadata analysis helped to learn about the corpus' structure and its general characteristics (the data is represented in TEI, indexed in an XML database and was visualized with Google Chart API and OpenLayers retrospectively). Secondly, a reading approach which I call »horizontal reading« (please do not mix this up with distant reading), which helped to build a close-reading, narrative thread. After describing these methods and their results, I will close my paper with an outlook on how the resources could be utilized for further research.

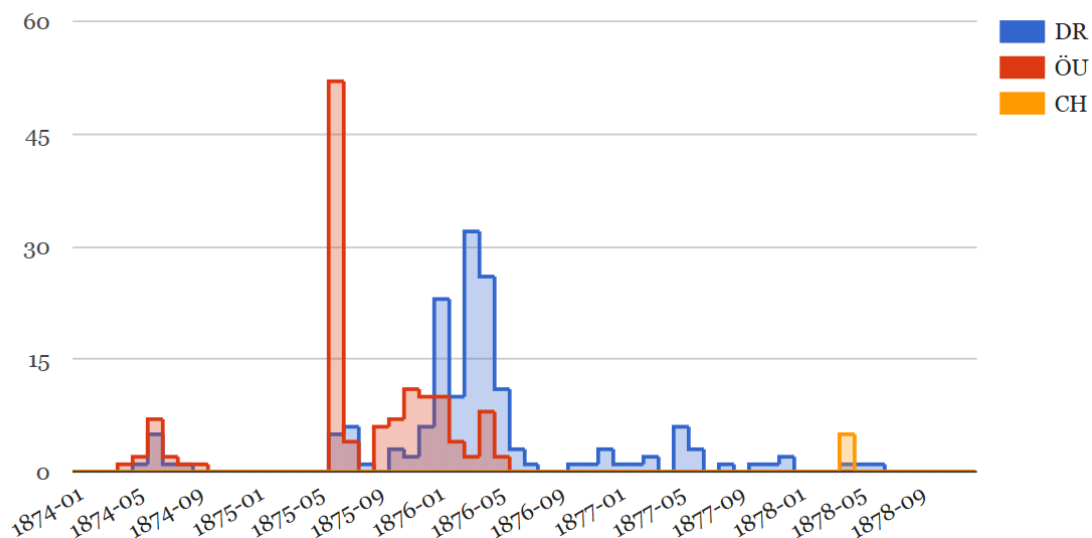


Figure 1: number of publications by time and country (DR: German Empire, ÖU: Austria-Hungary, CH: Switzerland)

2. METADATA ANALYSIS

The chronological overview (figure 1) displays the number of publications by time and country and distinguishes between three European countries with relevant German-speaking communities (German Empire, Austria-Hungary, Switzerland). The resonance is typically strongest in the months of the first performances, and was strongly amplified by the composer's presence in Vienna (May 1875), while Verdi's visit in Cologne (May 1877) did not produce a similar effect. In addition, it is visible that the press coverage in the German Reich, where opera houses and concert halls were spread widely over the country, stretches over a longer period than in Austria, where Vienna clearly dominated as cultural center. Finally, the diagram reflects the fact that performances in Germany started only half a year after the Austrian premiere. In mid-1874, there is a small reflection of the first performances in Milan and Paris, with a slightly higher interest in Austria. Concerning Switzerland, the number of recorded articles is still too low to be representative.

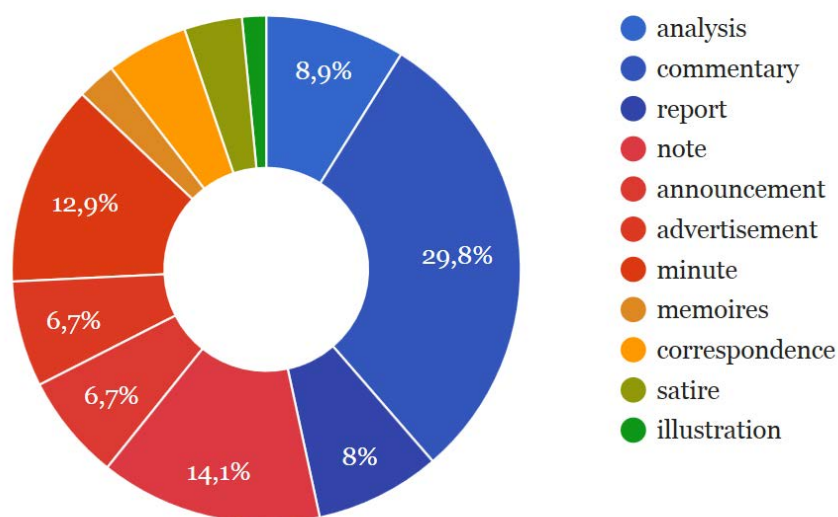


Figure 2: text categories and their proportional share of the corpus (by count)

Due to the open criteria, the corpus contains texts of very different genre and quality (figure 2). The texts cover a spectrum from advertisements to detailed reviews. Shorter texts contain mainly notes on events and served to reconstruct the performance history, while longer, more detailed texts contain relevant material for the investigation of the reception history itself. The latter texts (blue categories, about 150 in total) make up for 80% of the total amount of text.

[

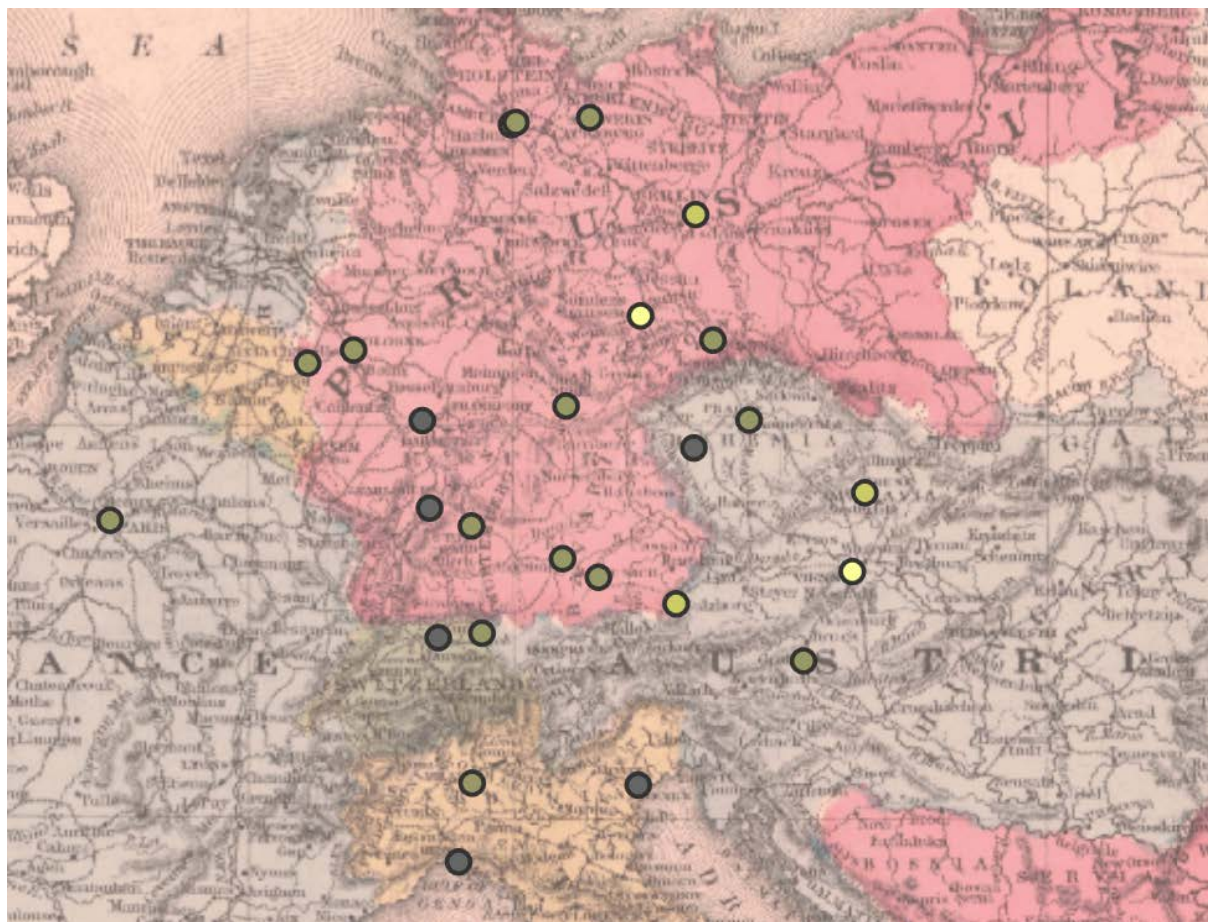


Figure 3: Publication places and their relative number (dark = 1, bright = 90). Note that a few publications from France and Italy appear here as well: these are articles that were repeatedly quoted in German newspapers.

This geographic view (figure 3) correlates the places of publication and the number of published articles. It becomes clear that Austria, again, is focused on Vienna, with 90 publications there, while the most productive place in Germany was Leipzig with 96 publications (both highlighted with the brightest dots). This is due to the fact that at least four important music journals resided in Leipzig. This points to very different situations of Austrian and German press.

Instead, comparing the mentioned places of performances in all articles (figure 4), the coverage is much broader, representing the most important venues (Berlin, Dresden, Hamburg, Cologne, Leipzig, Munich, Prague, Salzburg, Vienna) with a relatively high number of reports. The Vienna peak emerges again, while the large number of reports published in Leipzig were obviously not self-referring, but reported from other venues for the most part. (The low number of articles for the eastern German provinces is due both to weak availability of sources and lesser musical activity.)

Comparing Berlin, Leipzig and Vienna individually, it becomes clear that Vienna focused mostly on local events and marginally on events outside of Austria (mostly Italy and France), while ignoring the German Empire completely. Berlin, instead, published articles on events in the Empire and from European capitals as well, with only a few reports on Austria. Berlin and Vienna seem to have divided the German-speaking regions according to their territories. Leipzig, where most German music journals were published, demonstrated a broader scope that covered all German-speaking regions and many places in other European countries (indeed, for a relevant number of places, the only record in the corpus so far is an article published in Leipzig). This points to a very general tendency of newspapers to concentrate on local or national issues, while specialized press maintains an international perspective.

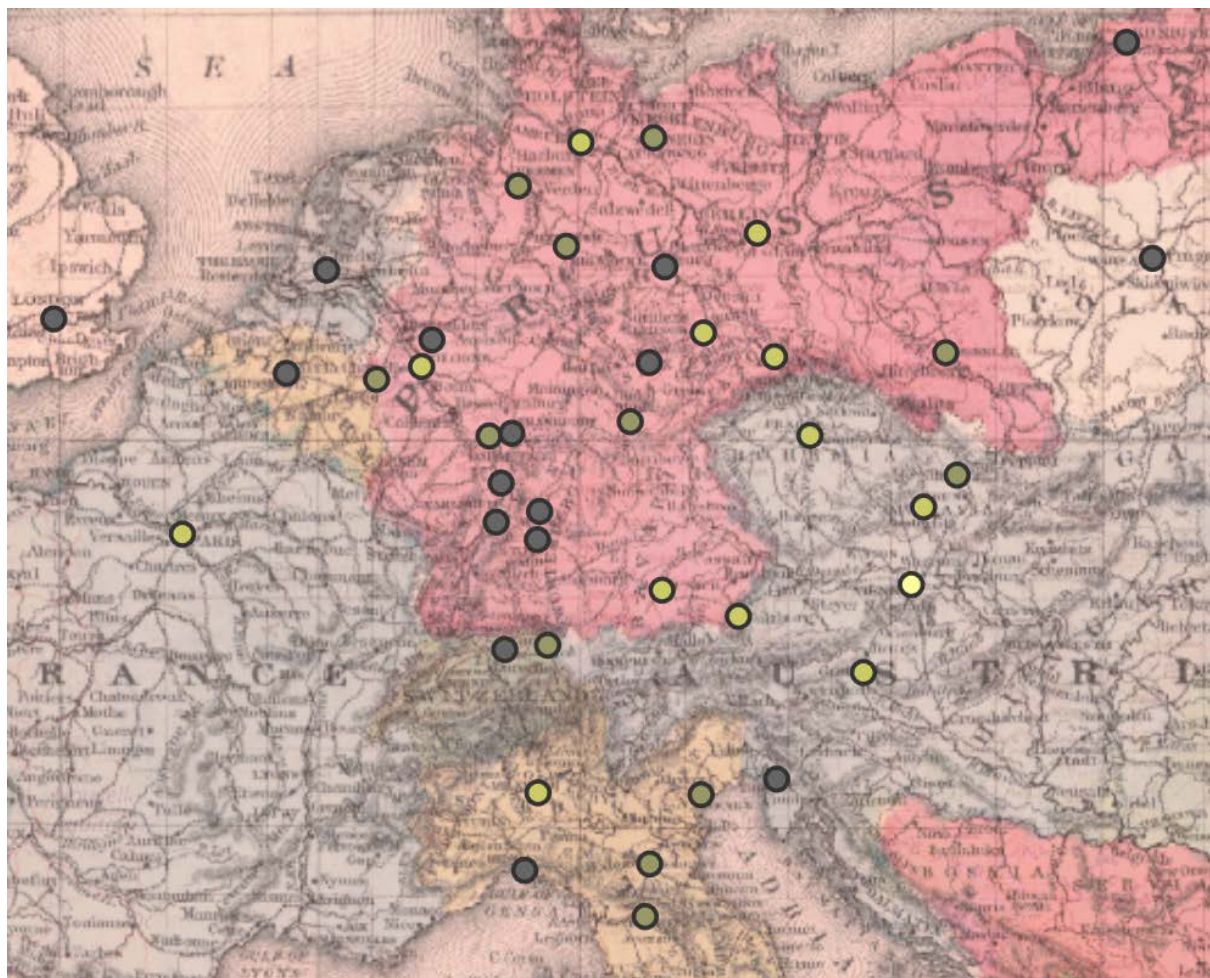


Figure 4: correspondence places and their relative number (dark = 1, bright = 90)

3. HORIZONTAL READING

For my text analysis, I used a method that kept a certain distance from the individual texts, and focused on common reference entities. Such a reference may be, for example, a composition or its part, or a place where a performance took place. After selecting one point of reference, the surrounding context (phrases) of all references in all texts from the corpus is extracted and grouped together into a parallel view, in a nutshell: reading all contexts of one single semantic entity in parallel. I call »horizontal reading«, in slight reference to »distant reading« [3]. Such a semantic »line of sight« through the corpus opens up the possibility of comparing different views and opinions directly to each other. Every individualized person, place, performance or composition could become a reference point for a horizontal reading. This requires technical preparations: Apart from the transcription work, the semantic entities in question must be assorted to standard identifiers. The present corpus contains a total of 8,000 reference points, covering about 530 unique persons, 142 performances, 96 localities and 135 works and making it a corpus of »smart data« [4].

To demonstrate my approach, I give a horizontal reading of the »Mors stupebit« from the »Dies irae« sequence and the introductory piece »Requiem aeternam«. Both pieces reflect different ideas of death: One was perceived more figurative and the other one more idealistic, which were appreciated very differently.

The »Mors stupebit«, in which Verdi depicted the amazement of death in front of the Last Judgment by a short rhythmical figure followed by a beat of the bass drum and a grand pause, was depreciated with categories like »realism« and »tone painting« and strongly criticized in almost all reports (see quotations in figure 5). Some critics tried to interpret this as a typical reflection of Italian temperament, while others tried to caricature the musical effect with horrifying descriptions.

The contexts of the »Requiem aeternam«, instead, give a totally different impression. The introductory piece thrilled the critics as well as the audience. The reports from Vienna described an intense, serious mood, and one critic from Munich tried to give an associative descriptions. Only the reports from Berlin and Dresden were much more skeptical, however: Although the piece was generally admired, the distribution of the text was boldly criticized.

»wenn er [...] den leibhaftigen Tod vor lauter Schrecken fast sprachlos dastehen läßt, so steht eine derartig übertriebene Tonmalerei jedenfalls nicht im richtigen Verhältnisse zu Charakter der Messe, selbst wenn diese nicht unmittelbar den kirchlichen Zwecken dienen soll« (J. G. Wörz, Wien)

»Das Stutzen und Staunen des Todes scheint mir hier denn doch nur zu derb realistisch ausgedrückt« (F. Stetter, München)

»mit [...] den humoristisch wirkenden Schlägen der großen Trommel« (Norddeutsche Allgemeine Zeitung, Berlin)

»Den Eindruck einer Tongrimmasse macht das wiederholte Bum! der großen Trommel bei dem aus dem Zusammenhange gerissenen Worte: »Mors« (stupebit).« (Kreuzzeitung, Berlin)

»beim »mors« stockte das Athmen des Tonkörpers, wie es handgreiflicher gar nicht zu machen war« (A. Dörffel, Leipzig)

»Bei uns im Norden wird man bei diesem drastischen Realismus sicherlich keinen eine Gänsehaut hervorrufenden kalten Schauer empfinden, sondern vielmehr ausrufen: »Bange machen gilt nicht!«« (E. Naumann, Dresden)

»gar zu äußerlich und auf leicht entzündbare italienische Gemüther berechnet« (National-Zeitung, Berlin)

»Manche seiner Effecte [...] erfordern geradezu scenischen Apparat [...] so trocken und nüchtern im Concertsaale mit brillanter Beleuchtung ausgehört, streift die Stelle [...] hart an der Grenze des Lächerlichen, weil nicht in Harmonie mit der Umgebung.« (A. Guckeisen, Köln)

»Frappant wirken bei Verdi die nachschlagenden Pulse — ein Klangeffect, den er (der Erste) in glücklicher Weise aus Beethovens neuer Symphonie herübergenommen.« (A. W. Ambros, Wien)

Figure 5: contexts for entity »Mors stupebit« (from the second part of the *Messa da Requiem*), extract; here not translated or transcribed, but given as an image, to demonstrate the idea of a context.

In the result, it becomes very clear which idea of death was preferred in German music aesthetics. While critics were emotionally appalled of the »Mors stupebit«, the »Requiem aeternam« just hit the perfect, consoling tone. Concerning ideas about death and eternity, German aesthetics rejected concepts such as »realism« and preferred a puristic, spiritualising effect in music, which was promoted by the restoration movement in church music. It is remarkable that this conception was only visible in the area of concert criticism, since the »Messa da Requiem« was completely ignored by contemporary church music journals.

Furthermore, horizontal reading also reveals general differences between the German Empire and Austria, as well as regional and local differences. For example, reports from Berlin were extremely sensitive concerning the structure of fugue compositions, while critics in Vienna and in Cologne were more open to the scenic effects. Through comprehensive comparison it is possible to identify and locate general tendencies, even if the individual background of the authors is not known in each individual case.

4. SUMMARY

Metadata analysis and horizontal reading are two ways to approach a text corpus that consists of heterogeneous texts by mostly unknown authors. They strongly support the creation of analytic narrations. It is possible to conduct single studies on each part of the composition, to explore ideas and concepts in music criticism, or to look at the situation in certain venues. Partially, these ideas were pursued in my dissertation, but there are many more facets of the reception history still to be investigated. The digital material is published freely to stimulate further analysis. Can positive and negative critiques be distinguished by stilometric methods? Can network visualizations of comparisons to give an idea of the relevance of canonical works? What would a MEI encoded full score of the work look like if it was connected section-wise with the various comments by the critics? How can the data be effectively connected to other corpora or databases?

5. REFERENCES

- [1] KIRSCH, Winfried: *Kirchenmusikreform, Cäcilianismus und Palestrina-Renaissance*, in: Wolfgang Hochstein und Christoph Krummacker (eds.): *Geschichte der Kirchenmusik*, vol. 3, Laaber: Laaber, 2013, p. 56–71.
- [2] KREUZER, Gundula: »Oper im Kirchengewande«? Verdi's Requiem and the Anxieties of the Young German Empire, in: *Journal of the American Musicological Society* 58,2 (Summer 2005), p. 399–450.
- [3] MORETTI, Franco: *Distant Reading*, Konstanz: Konstanz University Press, 2016.
- [4] SCHÖCH, Christof: Big? Smart? Clean? Messy? Data in the Humanities, in: *Journal of Digital Humanities* 2, no. 3 (Summer 2013), p. 2–13.
- [5] SCHWEIKERT, Uwe: *Messa da Requiem*, in: Anselm Gerhard and Uwe Schweikert (eds.): *Verdi-Handbuch*, Kassel: Bärenreiter; Stuttgart/Weimar: Metzler, 2013 (2., überarb. und erw. Auflage), p. 557–565.
- [6] SPONHEUER, Bernd: Über das »Deutsche« in der Musik. Versuch einer idealtypischen Rekonstruktion, in: Hermann Danuser and Herfried Münkler (eds.): *Deutsche Meister – böse Geister? Nationale Selbstfindung in der Musik*, Schliengen: Edition Argus, 2001, p. 123–150.

Building A Digital Ottoman/Turkish Serial Novel Archive

Ali Serdar¹, Reyhan Tutumlu²

¹ Özyeğin University, Turkey - ali.serdar@ozyegin.edu.tr

² Sabancı University, Turkey - rtutumlu@sabanciuniv.edu

ABSTRACT

As a part of “History of Serial Novels in Turkish Literature (1831-1928)” project 302 newspapers and journals published in Arabic script in the period 1831-1928 has been examined. 569 indigenous and 780 translated serial novels published in those periodicals has been recorded, digitalized and submitted to the Ottoman/Turkish serial novel archive. In this paper as one, but of a prominent outcome of the project, formation of the digital archive, the methods used during archive studies and digitalization process are going to be explained. Since both serial form and Ottoman/Turkish literature had unique characteristics, digital archive needed to be organized according to those features. Developing of this kind of a specialized digital archive is also contributed to Ottoman/Turkish cultural heritage and cultural memory, since the period and the genre that had been examined are nearly out of the scope of Turkish literary history. As a consequence of project many novels and novelists that had been forgotten are rediscovered and reintroduced to Turkish literature. Digitalization also gives the opportunity to make statistical analysis of the literary period, to generate graphs of indigenous and translated serial novels, to analyze the period in terms of gender. This archive can be seen as one of a modest contribution to the growing area of digital humanities.

KEYWORDS

Ottoman/Turkish serial novel, digitalization, archive, cultural heritage, cultural memory

1. INTRODUCTION

Serial is a name given to the publication of literary works, which are printed as episodes or chapters in periodicals such as newspapers and journals and many different literary genres, like travel writings, memoirs, history articles etc. had been published in serial form. But the main focus of our study is serial novel which beginning from the late 18th century but especially during the 19th century had become a dominant trend and had formed a literary tradition of its own especially in Western literatures. Novel as a genre was established in late 19th century in Ottoman/Turkish society, and unlike the birth of novel in western literature, published and spread through serials in periodicals. Considering the significance of Ottoman/Turkish serial novel tradition a project titled “History of Serial Novels in Turkish Literature (1831-1928)”¹ had been conducted and a digital serial novel archive has been constructed.

In this paper as one, but of a prominent outcome of the project, formation of the digital archive, the methods used during archive studies and digitalization process are going to be explained. Especially archive’s contribution to Ottoman/Turkish cultural heritage and cultural memory is going to be discussed. Finally, in a broader sense archive’s potential impact on digital humanities studies as a specialized digital archive is going to be referred.

2. METHODOLOGY

“History of Serial Novels in Turkish Literature (1831-1928)” project begun in May 2014 and ended in May 2017, in this course of three years 302 newspapers and journals published in Arabic script between 1831, the date which the first Turkish newspaper was established and 1928, the date in which Turkish alphabet had changed, has been examined and 569 indigenous and 780 translated serial novels published in those periodicals has been recorded. 12 libraries, collections and archives which are settled in Istanbul, Ankara and Izmir were used for archive studies.

Since hundreds of periodicals had been published in that period criteria for selecting the periodicals for analysis set as follows: (1) Turkish periodicals using the Arabic-based Ottoman alphabet²; (2) periodicals that relate to literature in one way or another, e.g. by (2a) including literary texts, (2b) featuring literature in their tags, and (3) publishing at least 10 issues. So 302 periodicals are selected and examined according to those criteria.

In terms of determining whether a text could be read as a novel, we encountered the problem of distinguishing between long stories, stories, and novellas. There are some cases in which periodical publishers had lightened our way by putting a descriptive note about the genre of serial published, like “*İkdam*’ın edebî tefrikası” (“Literary serial of *İkdam*”) or “*Son Saat*’in roman tefrikası” (*Son Saat*’s novel serial). Although it is not a hundred percent evidence those notes are directive and mostly resolved the problem. If there is not such a note, we check the author’s name and consider her/his literary works, and again this is not a hundred percent evidence. In both cases of course textual analysis considering whether if the text has a plot, story and characters are also checked out, since different genres of prose, like travel memoirs, political

¹ Our project is supported by TÜBİTAK (The Scientific and Technological Research Council of Turkey) (project number 113K497). The web site of the project is <http://tefrikaroman.ozyegin.edu.tr/ANASAYFA?lang=en>

² Newspapers published in Greek, Armenian, Kurdish, and Arabic reflected the languages spoken in the Ottoman Empire. Some newspapers were Turkish but used the Greek or Armenian alphabet. For example, the Greek newspaper *Telegraphos tou Vosporou*, published in 1843 in Istanbul, serialized translated novels such as *Monte Cristo* and works by Eugène Sue and others (Strauss, 2003, 43).

articles, and essays about daily life were also serialized. Once we decided that a serial text is a fiction, rather than using a qualitative method depending on a strict definition of the novel genre, because different readers and critics may categorize the same work in different genres we counted it as novel if it runs longer than seven serial pages³.

Also a digital archive has been constructed in which the citation information of recorded serial novels, full text of novels which are outside of copyright law is uploaded. In the case of serial novels which are bounded by copyright law while the full citation information is displayed, digital copies of only first and last episodes are uploaded. The database is accessible to researchers since January 2017 (<http://ereseach.ozyegin.edu.tr/xmlui/handle/10679/888>).

Digital archive titled “Türk Edebiyatında Tefrika Roman Tarihi” is available on Özyeğin University Library’s academic, open-access repository titled eResearch@Ozyegin. “Repository runs on Dspace, open source software originally developed by MIT (Massachusetts Institute of Technology) in the USA. It is compatible with OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). All items deposited in eResearch@Ozyegin is indexed by Google Scholar, OpenDOAR and ROAR, harvested by OpenAIRE. It uses Dublin Core as metadata elements and the metadata is OpenAIRE and DRIVER compatible” (<http://ereseach.ozyegin.edu.tr/page/about>).

Since “serial novel” is a specific genre, compared to other items like article, conference paper, book chapter etc. submitted to dspace repository some new features related with the entries had to be added to collection of serials. For example, serial novels have a beginning and end dates. In addition in late Ottoman society three different calendars (Hicri [The Hagira], Rumi [Julien] and Miladi [Gregorian]) were used simultaneously. While every item is submitted by converting the dates to Gregorian, original dates are also recorded to items catalogue records. So the process of submission is organized for those features, and users can find those details in collection items.

Another issue that has been solved is the standardization of digital copies of serial novels. Some of the archives which have been examined had digital copies of periodicals, but the type of data may vary: “bitmap”, “tiff”, “jpeg”, “djvu” or “pdf”. In the course of examining a periodical once a serial is detected, only the page that the serial published is copied. If archive is not digitalized the photos of the pages are taken. Once the serial is gathered regardless of the type of the data, every file is converted to “pdf”, combined in a single file and submitted to the collection as “full text” of the serial novel. In the course of examination two information forms were also filled, one is for the examined periodical and the other one is for the serial novel, in order to follow the dates and number of issues, but especially for the missing issues of both periodicals and serials. Recording the missing issues is especially crucial since they can be found in different archives, and in some cases a serial novel is completed using three or four different archive.

A different problem is about the “stalking” of uncompleted serial novels, since in some cases for many varying reasons a periodical had stopped publishing the serial novel, but the serialization of novel might be completed in a different periodical in a different time. For example Halit Ziya Uşaklıgil’s *Kırık Hayatlar* (Broken Lives) serialized in 1901 in both *Servet-i Fünun* and *İrtika* but left unfinished, however re-serialized and completed in *Vakit* in 1922. In those cases every “edition” of serial novel is submitted to collection for the sake of giving the possibility to the researchers to make the comparison of the different editions.

3. CULTURAL HERITAGE

Although politically a radical change had lived with the establishment of Republican regime, one cannot talk about a solid rupture from Ottoman culture immediately. It is also true for literature, since the Ottoman novelists are the founder of Turkish novel. However after the establishment of Republican regime in 1923 there had been an alphabet shift in 1928 (called “alphabet revolution”), and Turkish alphabet had changed from Arabic based to Latinized based. It is called “based” because both alphabets are not purely Arabic or Latin and shift of alphabet had some consequences, since generations that had learned Latinized based alphabet could not read texts that had been published before 1928. This is the cultural aspect of shift, which by the project fill the gap in terms of history of Turkish novel. By the project a period and a huge sample of periodicals examined, and serial novels in those periodicals are rediscovered. Moreover a tradition which resides in periodicals are preserved and classified by digitalization.

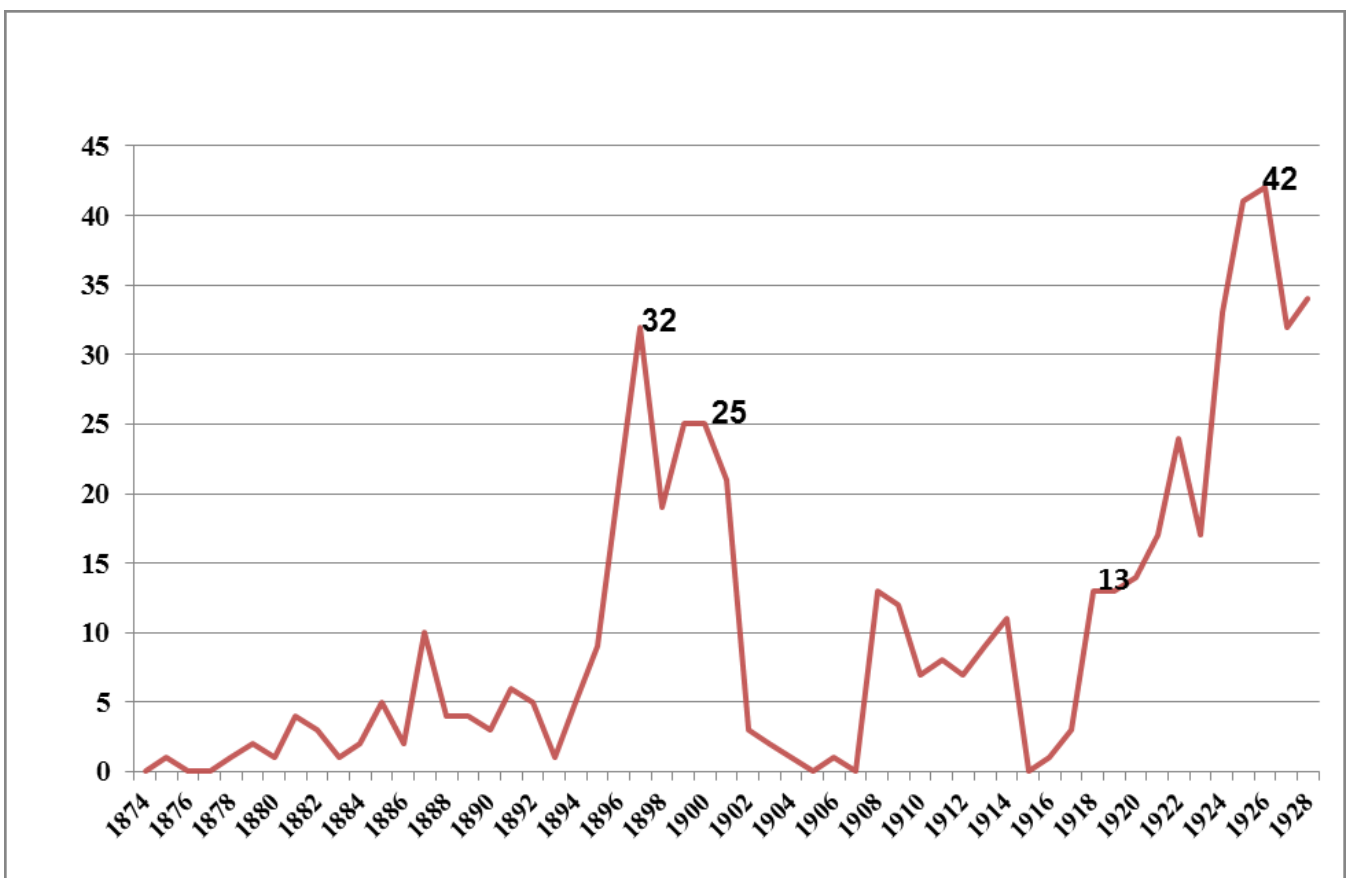
4. CULTURAL MEMORY

One of the aims of the project is to discover the novels and novelists which are residing in the pages of periodicals, sinking into oblivion and add their names to the literary history. As mentioned above 569 indigenous novels detected and 239 of those serial novels are never mentioned in literary histories, anthologies or in any resource related with Turkish literature. With the discovery of forgotten novels and novelists the scope of Turkish literature is enhanced. Moreover the novels and novelists once red and disputed by Ottoman/Turkish literary public, since they are published in daily newspapers, weekly or monthly magazines are quite on the scene and even some of which are popular at that time, can be re-introduced to the actual readers/society by transliterating and re-printing. So by the project at least for the genre of serial novel it can be asserted that a refreshment of cultural memoir of the Turkish readers and researchers can be expected.

³ In general, an episode counts only a quarter page of a newspaper, but sometimes a newspaper devotes two quarter pages, or in some cases a magazine devotes three full pages to an episode and serial may end in five episodes. So in order to overcome this relativity we determine the lower limit as seven serial pages, which is nearly equivalent of 35 book pages in Latin alphabet.

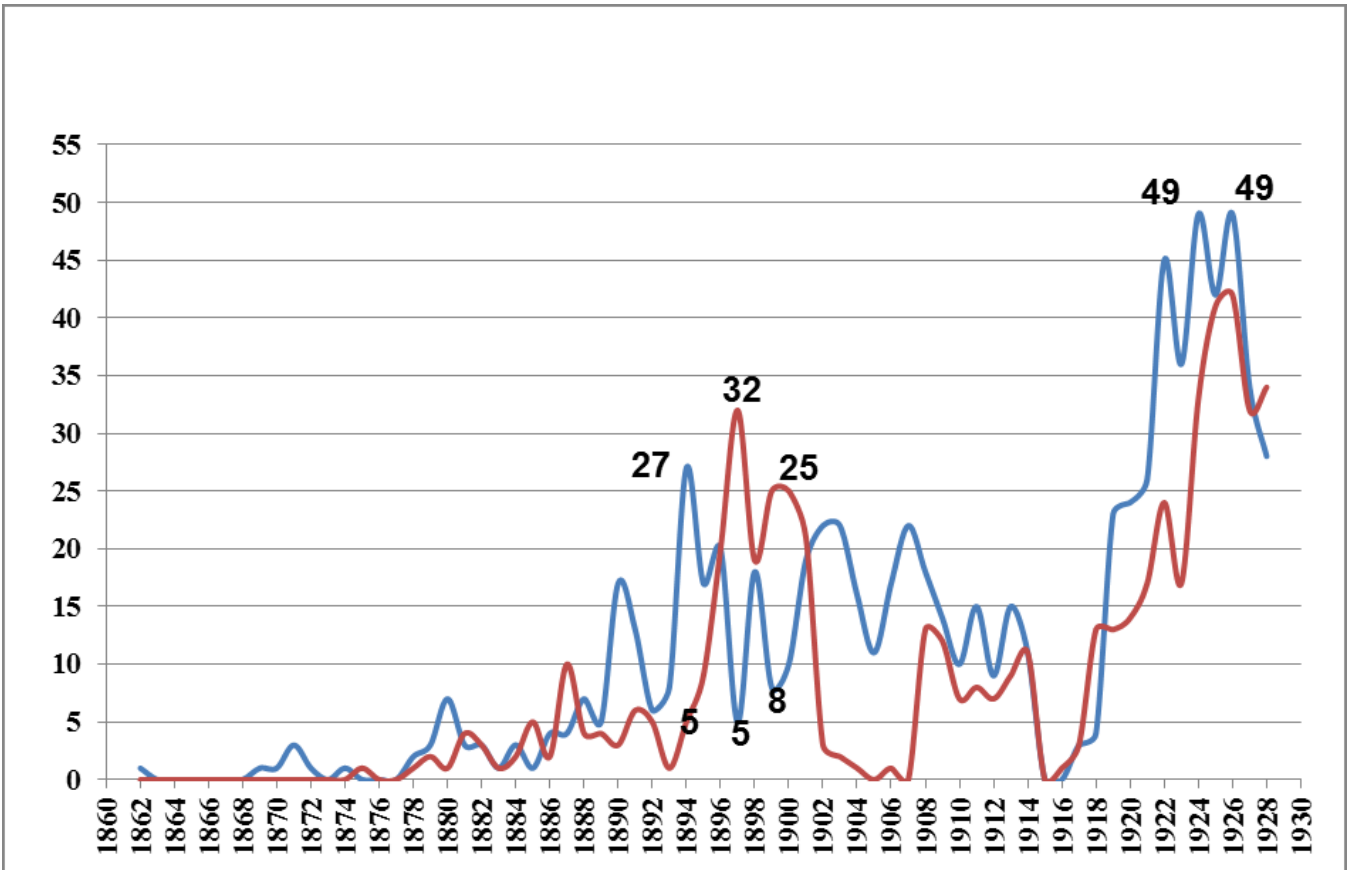
5. LITERARY HISTORY AND ANALYSIS

With the digitalization of serial novels collecting the data gets more systematic and easier to examine. Since we had a vast data, we have adopted Franco Moretti’s “distant reading” approach. Moretti (2013) states that the close reading method that has been the prominent approach in literary criticism for many decades ought to be limited to a certain set of texts: “But the trouble with close reading (in all of its incarnations, from the new criticism to deconstruction) is that it necessarily depends on an extremely small canon. [...] And if you want to look beyond the canon (and of course, world literature will do so: it would be absurd if it did not!), close reading will not do it. It is not designed to do it, it is designed to the opposite.” (48). What lies behind the Moretti’s approach are the “great unread” texts that are unclassified and therefore excluded from the literary canon, and in our case the great unread turns out to be the Ottoman/Turkish serial novels residing in the pages of periodicals. Moretti’s approach suggests rather than analyzing this mass of texts by close reading, which is not functional and suitable in this case, but proposes the method of distant reading: “[W]e know how to read texts, now let’s learn how *not* to read them. Distant reading: where distance, let me repeat it, *is a condition of knowledge*: it allows you focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems” (48-49). In this regard, by the data we have gathered, we can deduce some of the characteristics of the Turkish serial novel, figure out the basic trends like frequency of serialized novels per year (Graph 1) and correlate this frequency with social, political, economic and cultural factors.



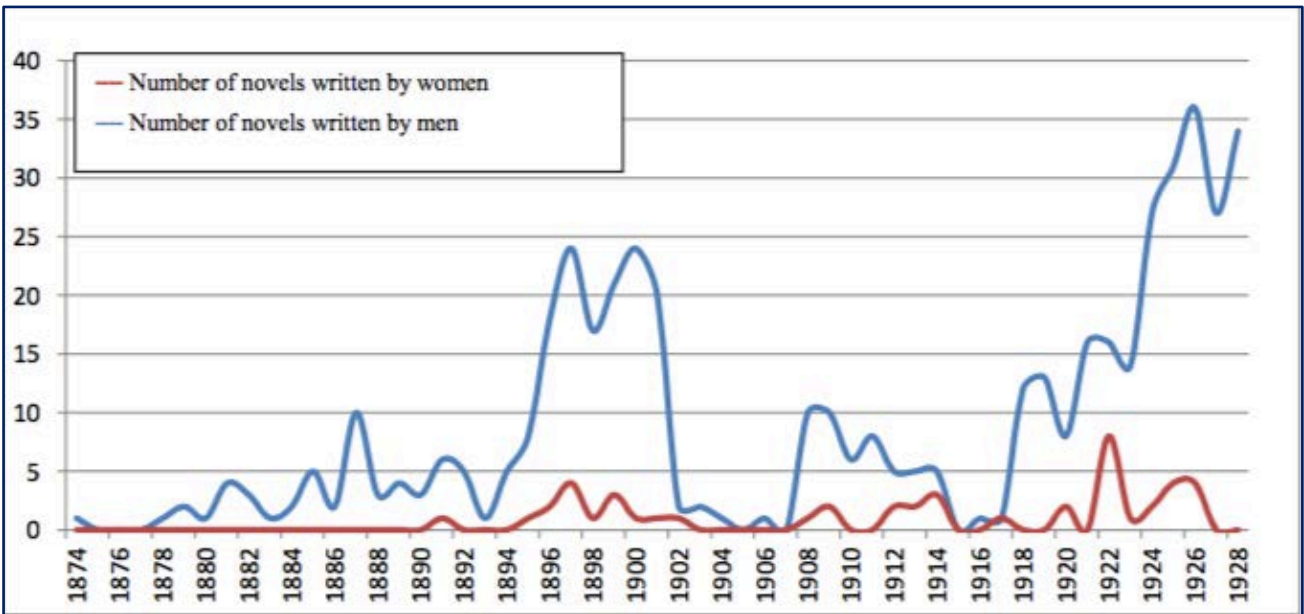
Graph 1. Frequency of serialized indigenous novels by year.

Since a similar graph has been created for translated novels and by that way comparison and analysis of indigenous and translated novels can be done (Graph 2):



Graph 2. Comparison of frequencies of indigenous and translated serial novels by year

A graph that is based on the frequencies of gender has also been created, by which distribution of men and women novelists can be analyzed (Graph 3).



Graph 3. Frequency of number of novels written by men and women novelists

All those up and downturns of frequencies displayed in these graphs gives us the chance to interpret the literary period under examination and all those frequencies closely relates literature with the cultural, historical, economic and political

history of Ottoman/Turkish society. Since the frequencies correlates with the basic moments of such phenomenon like censorship, political pressures, and reel politics like war and political turmoil⁴.

6. CONCLUSION

As a part of a broader project a digital archive of the Ottoman/Turkish serial novel had been built. With this archive one of a substantial cultural heritage of Ottoman/Turkish society is being preserved, classified and analyzed. Nearly an unforgotten tradition of serial novels rememorized and reintroduced to literary public. This archive can be seen as one of a modest contribution to the growing area of digital humanities.

7. REFERENCES

- [1] Moretti, F. 2013. *Distant Reading*. London: Verso.
- [2] Strauss, Johann. 2003. "Who Read What in the Ottoman Empire." *Arabic Middle Eastern Literatures* 6.1: 39-76.

⁴ Since the description and discussion of such an extensive analysis is not the focus of this study, those issues are not discussed in detail. Hence, if demanded, in the discussion period of the presentation they can be referred and extended.

Creating a Community on Film-induced Tourism through a Crowdsourcing Approach

Andrea Micheletti¹, Daniel Zilio², Nicola Orio³

¹ University of Padua - Department of Cultural Heritage, Italy - andrea.micheletti@unipd.it

² University of Padua - Department of Cultural Heritage, Italy - daniel.zilio@unipd.it

³ University of Padua - Department of Cultural Heritage, Italy - nicola.orio@unipd.it

ABSTRACT

In this paper, we present an experiment to create a social community on film-induced tourism, which is visitation of places related to film productions. The first part of the paper briefly reviews the state of the art on crowdsourcing, because the tasks identifying the location where a movie scene has been filmed does not require skilled experts. Yet, general users need to be motivated for completing these tasks, which is the reason why this paper pays attention to the narcissistic act of exhibitionism aimed at social recognition. In the final part, the paper introduces a project that aims at enriching a digital collection of information about places appearing in movies and other audio-visual artefacts. It also includes a brief illustration about the obtained results and the exploited strategy.

KEYWORDS

Crowdsourcing; Film-induced Tourism; Social communities; Motivation.

1. INTRODUCTION

In a scenario where it is increasingly important to diversify the tourist offer, the development of thematic paths such as the ones of film-induced tourism plays a central role [3]. Film-induced tourism has been defined by Sue Beeton as “visitation to sites where movies and TV programmes have been filmed as well as to tour to production studios, including film-related theme parks” [1]. This phenomenon is rising interest in destination management and destination marketing, because movies and TV programmes add immaterial and intangible value to locations. The intangible value dimension goes beyond the concept of a real place, which cannot be suitably recognized and represented using tools for traditional visits. Furthermore, Francesco Di Cesare noted that film-induced tourism: is largely independent from seasonal trends; it conveys tourists from overcrowded sites to new and less explored ones; and eventually can be suitable for a substantial re-branding of a certain area. A number of initiatives has been already undertaken at the international level by public and private bodies, which developed movie maps and movie tours or exploited the success of a particular movie as a tool for destination branding [6]. There are many international success stories where movies have been exploited to attract tourists, such as *The Lord of the Rings*, *Harry Potter* and *Games of Thrones* sagas. Also Italy has some interesting cases: *The Passion of the Christ* in Basilicata, *Inspector Montalbano* in Sicily, *Don Matteo* in Umbria and *Un passo dal cielo* in Trentino [12,14].

A first approach with this topic dates back to 2014 [13]. The aim was the creation of a database to represent the relationship between movie excerpts and locations with a double goal: promoting film-induced tourism and raising interest towards audiovisual productions, their language and their relation with the landscape. One critical aspect was that a restricted number of users – e.g. scholars in film studies – were involved in providing all the information, from film analysis to location identification. Yet, only part of this information required the competences of trained experts while most of the effort was spent in the less specialized task of annotating multimedia items with geographical information [18]. In this paper we propose an approach to solve the problem of information gathering through a crowdsourcing approach.

2. ABOUT CROWDSOURCING

The arrival of the Web 2.0 and social media started a technological emancipation, where more minds work together to achieve a common goal. As mentioned by famous Gestalt psychologist Kurt Koffka: “the whole is other than the sum of the parts”. A direct consequence of all these changes is the birth of a new generation called “generation C”: connected, communicating, content-centric, computerized, community-oriented, always clicking [9]. Any individual with the appropriate technology can now produce his or her online media and provide images, text, and sound about whatever he or she chooses [4]. Everyone is a potential author of contents available to everybody. Such a definition replaces the one-to-many model of traditional mass communication with the possibility of a many-to-many communication. Amazon *Mechanical Turk*, for example, is a well-known crowdsourcing platform which has demonstrated its utility in producing labelled image sets of adequate reliability for a variety of generic content labels in natural images, including objects, actions, and scenes [11].

The term crowdsourcing was indeed originally coined as a [10] variant of the term (and the practice of) outsourcing: “it is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual

skills, while they will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken” [7].

This definition highlights some important aspects of crowdsourcing:

- voluntary undertaking: timing and procedures are self-defined
- focused approach: activities are result-oriented
- remote interaction: activities are based on remote communication systems
- recognition of achievement: participation lays on the recognition of a material or immaterial achievement

These traits implicate some benefits such as: smaller economic risk, evaluation of intellectual capital, public engagement and innovation. Yet, there are also many possible menaces, such as: amateurishness, emphasis of traditional format and production of low-quality contents.

3. MOTIVATIONS

These considerations suggested that crowdsourcing is likely to be a part of a more complex process that is an investment in social recognition. People are encouraged to participate in crowdsourcing activities to see themselves objectify in other people’s mind. In his book “The Selfish Gene” [5], Richard Dawkins introduced the term meme to describe a cultural unit that is transmitted through replication. According to this perspective, imitation plays a central role in transmitting knowledge and the act of imitating is functional to the way people gain information from the environment. At the same time, imitation of a cultural model helps improving self-esteem and gaining reputation within a group of peers.

In order to encourage participation to a crowdsourcing activity, we envisage an interaction with simple and engaging quests, which implements a reputation mechanism. This process with the increasing development of social networks, is more and more linked to an investment in social capital. In this contest, “social capital is the actual or potential resources which are linked to a durable network of more or less institutionalized relationships of mutual acquaintance or recognition” [2]. The quest for gaining a social capital requires a continuous search for consensus [17].

A community of undergraduate students, for instance the class of Design and Management of Cultural Tourism that has been involved in our initial tests, does not differ from the general trend previously outlined. In this case, participants were encouraged to carry out the activities to gain points that were added to the final grade for the course in Computer Tools for Tourism, but their involvement went beyond the initial requirements. This identification with a group, based on doing activities just for the pleasure of it with no hidden agenda, is one major aspect of new digital media. Moreover, sharing with the group both what we have and what we know is becoming a pleasure without the expectation of reciprocity, an openness based on selfless approach. This sense of belonging, which is shared also by on-line communities that have been called emotion communities, is balanced by an opposite drive: consumption and participation as a narcissistic act of exhibitionism to be socially recognized [8,15].

4. A CROWDSOURCING PLATFORM FOR FILM-INDUCED TOURISM

We develop a web portal called Cinemaps where registered users can contribute to recognizing movie excerpts that were filmed outside and of identifying the location where they have been filmed. Cinemaps envisage a simple and intuitive interface, which implements a reputation mechanism. The focus was on the creation of a database to represent the relationship between movie excerpts and locations.

Our crowdsourcing approach is based on three tasks:

- discover and insert into the database system audiovisual productions that are relevant for a geographical area
- geolocate the precise location of the movie scenes on a digital map
- approve or refine the list of movies and their localization to create a community of users based on the concept of reputation

5. A CROWDSOURCING COMMUNITY FOR FILM-INDUCED TOURISM

The crowdsourcing activity through the social network started investigating the potential interest of a passionate community. We build some focused posts into the Facebook groups related to a small geographical area (Padua and its Province and the Po Delta). In these posts, we ask information about movie and the location where they were filmed. The reaction was positive, as shown in Table 1, with many suggestions which go beyond the initial questions. Starting from these results we have created a Facebook group called “I Film del Veneto” and a Facebook page called “Cinemaps Veneto”, aimed at building a community of film-addicted users rooted in the territory.

The results are as follows:

People Reached	Page Followers	Post engagements	Community members
41.271	231	4.176	926

Table 1: Principal parameters of Facebook campaign.

To create an effective Facebook campaign, we started planning a strategy based on:

- **objective:** move users interested on film-induced tourism from Facebook to Cinemaps web portal
- **mission:** bringing People in the places they love to enhance the films they love

- **target:** people who live in Veneto and which love films
- **themes and contents to provide:** we had design a Photoshop layout with three section: place, movies, riddle. The aim of the post was to recognize movie or location requested in a riddle to engage people as a sort of quiz-game. Furthermore, we had built a lot of tips and curiosity about Veneto films and locations
- **competitors:** the main competitors are listed in Table 2

Page Name	Followers	Likes
Cineturismo	322	6
Veneto Film Tour	1253	10
Veneto Film Network	820	10
Films A Venezia	499	8
Padova Films Commission	850	8
Film4tourism Padova	450	2
Veneto Film Commission	1498	3
I Cineamatori Del Davinotti	478	5
Italymovietour	354	1
Vicenza Film Commission	2500	4
Cineturismo (Gruppo Pubblico)	184	1
Filmartstudium	1662	1267
Cinema&turismo	120	1

Table 2: List of the main competitors.

6. CONCLUSIONS

We created and are managing a community of film enthusiasts centred on the relation between films and territory. The goal is to promote a crowdsourcing approach for the identification and geolocation of experts from movies and TV series that were filmed in Veneto. At the same time, we expect that the community will be an active promoter of its own territory and of the audiovisual productions that have been filmed in. The final goal is to collect, manage and distribute the information to promote film-induced tourism initiatives.

7. REFERENCES

- [1] Beeton, S. (2005). *Film-Induced Tourism*. Channel View Publications. Clevedon.
- [2] Bourdieu, P. (1985). *Handbook of Theory and Research for the Sociology of Education*. In *The Forms of Capital*. Greenwood, New York: Richardson, J.C. (ed.).
- [3] Censis (2008). *L'auto e i turismi tematici. Rapporto Turismo*. Retrieved 09 2017, from: <http://bit.ly/2zH54zf>
- [4] Croteau, D. (2003). *Media society: industries, images, and audiences / David Croteau, William Hoynes*. Thousand Oaks, California: Pine Forge Press.
- [5] Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- [6] Di Cesare, F., & Rech, G. (2007). *Le produzioni cinematografiche, il turismo, il territorio*. Roma: Carocci.
- [7] Estellés, E. (2012). *Towards an integrated crowdsourcing definition*. *Journal of Information Science*, v.
- [8] Fabbris, G. (2010). *La società post-crescita: consumi e stili di vita*. Milano: Egea.
- [9] Friedrich, R., Peterson, M., & Koster, A. (2011). *The Rise of Generation C How to prepare for the Connected Generation's transformation of the consumer and business landscape*. Retrieved settembre 21, 2017, from: <http://bit.ly/2BPlyah>
- [10] Howe, J. (2006). *The Rise of Crowdsourcing*. *Wired Magazine*.
- [11] Larson, M., Soleymani, M., & Eskevich, M. (2012). *The community and the crowd: Multimedia benchmark dataset development*. 19, 15-23.
- [12] Lavarone, G. (2016). *Cinema, media e turismo. Esperienze e prospettive teoriche del film- induced tourism*. Padova: Padova University Press.
- [13] Lavarone, G., Orio, N., Polato, F., & Savino, S. (2016). *Modeling the concept of movie in a software architecture for film-induced tourism*. IRCDL 2015. Springer, Cham.
- [14] Martello, G. (2016). *Il Location Placement come strategia di promozione del territorio*. Master's degree diss., University of Padua.
- [15] Micheletti, A. (2016). *Motivating and involving users through gamification: a proposal*. In M. Agosti, M. Bertini, S. Ferilli, S. Marinai, & N. Orio, IRCDL 2016 (Vol. 612, pp. 116-125). Springer, Cham.
- [16] Peterson, M. A. (2003). *Anthropology & Mass Communication: Media and Myth in the New Millennium*.
- [17] Wellman, B., & Wortley, S. (1990). *Different Strokes from Different Folks: Community Ties and Social Support*. *American Journal of Sociology*, 96(3), 558-588.
- [18] Zilio, D., Micheletti, A., & Orio, N. (2017). *Crowdsourcing for Film-Induced Tourism: An Approach to Geolocation*. IRCDL. Springer, Cham.

Dai giacimenti culturali ai repository digitali: il Museo virtuale dell'informatica archeologica

Paola Moscati

Istituto di Studi sul Mediterraneo Antico - CNR, Roma, Italia - paola.moscati@isma.cnr.it

ABSTRACT

L'intervento intende soffermarsi sulla nozione di patrimonio culturale e ripercorrere storicamente l'evoluzione del concetto di archivi digitali, a partire dagli anni Ottanta del XX secolo per giungere ai giorni nostri. Con particolare riferimento alla ricerca archeologica, si prende avvio dall'espressione "giacimenti culturali" per giungere al nuovo concetto di repository digitale, che interpreta al meglio le potenzialità della rete e la nuova sfida promossa dall'open access. In questo contesto s'inserisce il progetto di ricerca internazionale su "La storia dell'informatica archeologica" che è stato promosso dall'Istituto di Studi sul Mediterraneo Antico del CNR (ISMA) d'intesa con il Centro Linceo Interdisciplinare "Beniamino Segre" dell'Accademia Nazionale dei Lincei. Scopo del progetto è ripercorrere le principali tappe dello sviluppo dell'informatica archeologica, a partire dagli anni Cinquanta del XX secolo, e realizzare una raccolta di risorse digitali consultabili in rete. Il "Virtual Museum of Archaeological Computing" (<http://archaeologicalcomputing.lincoln.it/>) costituisce l'aspetto comunicativo del progetto e, attraverso la storia delle applicazioni informatiche e dei suoi protagonisti, pone in luce le implicazioni teoriche e le ricadute metodologiche che derivano dall'incontro tra scienza e archeologia. Itinerari multimediali accompagnano la navigazione tradizionale e in essi la narrazione si concentra sulla storia di progetti, istituzioni, eventi e tecnologie, rintracciandone le origini e seguendone l'evoluzione fino ai nostri giorni. Tra le risorse online ad accesso aperto, che consentono di effettuare approfondimenti, vi è il riferimento costante al repository della rivista «Archeologia e Calcolatori» (<http://www.archcalc.cnr.it/>), che è stata fondata nel 1990 e che dal 2005 ha aderito all'Open Archives Initiative, adottando pionieristicamente il protocollo OAI-PMH.

PAROLE CHIAVE

Archeologia, Informatica archeologica, Repository digitali, Musei virtuali.

1. PATRIMONIO CULTURALE: GIACIMENTO E RISORSA

L'art. 15, c. 1, della Legge 28 febbraio 1986, n. 41 così disponeva: «E' autorizzata la spesa di lire 300 miliardi per l'anno 1986 e di lire 300 miliardi per l'anno 1987, di cui il 50 per cento riservato al Mezzogiorno, da destinarsi alla realizzazione di iniziative volte alla valorizzazione di beni culturali, anche collegate al loro recupero, attraverso l'utilizzazione delle tecnologie più avanzate, e alla creazione di occupazione aggiuntiva di giovani disoccupati di lungo periodo, secondo le disposizioni del presente articolo. Il Ministro per i beni culturali e ambientali, d'intesa con il Ministro del lavoro e della previdenza sociale, definisce entro il 31 marzo 1986 un programma che dovrà concernere le seguenti aree d'intervento prioritarie: patrimonio archeologico, patrimonio architettonico e urbanistico, patrimonio librario, patrimonio letterario e linguistico, patrimonio storico archivistico, arti figurative e arti minori».

Nasce così nel 1986 l'iniziativa "Giacimenti culturali", che prevede, per la prima volta in Italia, l'interazione tra enti pubblici e consorzi di imprese private nell'intento di realizzare progetti per la catalogazione e la gestione delle risorse culturali. Tra i partner privati emergono colossi quali ITALSIEL, ITALTEL, ENIDATA, INFRASUD, FIAT ENGINEERING, IBM ITALIA, Olivetti, per citarne solo alcuni (Rapporto 1989). Tra i progetti, sicuramente il più noto è il Progetto Neapolis, che fa il giro del mondo grazie alla mostra Rediscovering Pompeii, inaugurata a New York all'IBM Gallery of Science and Art nel 1990 (Rediscovering Pompeii 1990).

Nel 1988, nel volume "Le isole del Tesoro. Proposte per la riscoperta e la gestione delle risorse culturali" pubblicato dall'IBM Italia, Umberto Eco si sofferma sull'espressione "giacimento culturale" (Eco 1988). Egli la considera come una metafora che si riferisce non solo a un bene nascosto che deve essere scoperto (o riscoperto) ma che implica anche una sua conservazione e un suo "sfruttamento", nel senso che esso deve essere reso fruibile ed entrare a far parte di un flusso di comunicazione che permetta di restituirlo alla collettività, rivalutandone la funzione originaria. Infatti, è proprio negli anni Ottanta del XX secolo che si afferma l'idea che l'automazione delle procedure di archiviazione e di gestione della documentazione relativa al patrimonio culturale possa al contempo costituire un potente strumento di ricerca e di diffusione della cultura, che risponda a nuove strategie di comunicazione e di fruizione dei contenuti.

Si tratta di quell'attività di "valorizzazione" che negli anni Novanta Oreste Ferrari definì come «ogni azione che porti alla percezione e a una più sviluppata presa di coscienza collettiva dei valori intrinseci e dei significati di identità socio-culturale delle opere archeologiche, artistiche e monumentali» da perseguire non solo con gli strumenti didattici tradizionali ma anche grazie a un'editoria che si avvale delle risorse della tecnologia informatica, come i CD ROM e, «in un non lontano futuro», di Internet (Ferrari 1998).

2. DAI DEPOSITI AI REPOSITORY

Alla diffusione del web corrisponde quella dei "depositi culturali digitali", collezioni di documenti in formato digitale raccolti per finalità collegate alla produzione, conservazione e fruizione di informazioni relative ai beni culturali. Rispetto agli archivi tradizionali, basati su un modello di disseminazione policentrica del patrimonio (Ferrari 1991; Ferrari 1996),

gli archivi informatici hanno caratteristiche e funzionalità gestionali e conservative proprie: criteri per la predisposizione di contenuti digitali, norme per l'acquisizione delle risorse, metodi e procedure per l'archiviazione e la gestione dei depositi, tecnologie e piattaforme applicative per lo scambio e la trasmissione delle informazioni. Ai fini di una definizione condivisa di deposito digitale e in vista della conservazione delle memorie digitali, la questione della qualità e della normalizzazione dei contenuti risulta di primaria importanza, così come l'attenzione verso gli standard di rappresentazione e di descrizione e i protocolli di trasmissione dei dati.

Nel corso del XXI secolo, il concetto di deposito si evolve in quello più attuale di biblioteca e museo digitali o virtuali, che riassumono in sé il fenomeno sempre più diffuso della transizione verso il digitale da parte delle istituzioni culturali (Galluzzi 2010). Ma è il "repository" il vero erede del "giacimento culturale", in quanto interpreta al meglio le potenzialità della rete e la nuova sfida promossa dall'open access. "Oggetto socio-tecnologico", come è stato definito da Guédon (2009), esso racchiude in sé risorse rappresentate nel loro contesto di provenienza e nella loro tipologia di appartenenza, ma le arricchisce di significati e le valorizza grazie a una rete di relazioni che evidenziano nessi concettuali tra i dati e che consentono agli utenti di interagire in modo differenziato con i contenuti accessibili online.

3. IL MUSEO VIRTUALE DELL'INFORMATICA ARCHEOLOGICA E IL REPOSITORY DI «ARCHEOLOGIA E CALCOLATORI»

In questo contesto si inserisce il progetto di ricerca internazionale su "La storia dell'informatica archeologica" che è stato promosso dall'Istituto di Studi sul Mediterraneo Antico del CNR (ISMA) d'intesa con il Centro Linceo Interdisciplinare "Beniamino Segre" dell'Accademia Nazionale dei Lincei. L'idea è nata in occasione del Convegno internazionale "La nascita dell'informatica archeologica" (Moscati 2009), svoltosi all'Accademia nell'ottobre 2008 (Comitato Ordinatore: Tito Orlandi, Paolo Sommella, Edoardo Vesentini, Paola Moscati), e il progetto ha preso avvio nel 2010. Scopo del progetto è ripercorrere le principali tappe dello sviluppo dell'informatica archeologica, a partire dagli anni Cinquanta del XX secolo, quando l'archeologia ha accolto procedimenti d'indagine scientifici e tecnici, adattandoli alle specifiche esigenze della propria ricerca (Moscati 2014; Moscati 2015).

Attraverso una rete internazionale di contatti – che si è potuta giovare delle collaborazioni già attivate nell'ambito della rivista «Archeologia e Calcolatori», pubblicata dal 1990 e riferimento per l'intero settore a livello internazionale, e nel corso dei Convegni organizzati presso il CNR e l'Accademia Nazionale dei Lincei – in questi anni si è operato per realizzare una raccolta di risorse digitali relative alla storia dell'informatica archeologica e per renderle consultabili in rete. È così nato il "Virtual Museum of Archaeological Computing" (<http://archaeologicalcomputing.lincei.it/>; Figura 1), che costituisce l'aspetto comunicativo del progetto e che, attraverso la storia delle applicazioni informatiche e dei suoi protagonisti, pone in luce le implicazioni teoriche e le ricadute metodologiche che derivano dall'incontro tra scienza e archeologia. Il progetto museale è condotto d'intesa con l'Osservatorio Tecnologico per i Beni Culturali del MiBACT, grazie alla scelta del CMS open source Museo&Web (Natale, Saccoccio 2010) e con il MIUR, che lo ha selezionato tra i progetti intesi alla diffusione della cultura scientifica (L. 6/2000).



Figura 1. Homepage del "Virtual Museum of Archaeological Computing"

Tra le principali risorse già disponibili in rete: le biografie e le bibliografie dei "pionieri" delle applicazioni informatiche all'archeologia; alcuni documenti inediti provenienti dall'archivio di Jean-Claude Gardin, conservato a Nanterre presso la Maison René Ginouvès (Moscati 2013; Moscati 2016); i Contributi del Centro Linceo interdisciplinare "Beniamino Segre", che dagli anni Settanta hanno dedicato specifica attenzione all'informatica applicata alla ricerca archeologica; gli Atti del

“III International Symposium on Computing and Archaeology”, svoltosi a Roma nel 1995 e pubblicati nel n. 7 della rivista «Archeologia e Calcolatori» (Moscati 1996); la “Bibliografia di informatica archeologica” relativa agli anni Novanta, che raccoglie oltre 2600 titoli.

Itinerari multimediali accompagnano la navigazione tradizionale e in essi la narrazione si concentra sulla storia di progetti, istituzioni, eventi e tecnologie, rintracciandone le origini e seguendone l’evoluzione fino ai nostri giorni (Figura 2). Tra le risorse online ad accesso aperto, che consentono di effettuare approfondimenti, vi è il riferimento costante al repository della rivista «Archeologia e Calcolatori» (<http://www.archcalc.cnr.it/>), che dal 2005 ha aderito all’Open Archives Initiative, adottando pionieristicamente il protocollo OAI-PMH, e che è inserita nella Directory of Open Access Journals (Barchesi 2006; Caravale, Piergrossi 2012; Caravale, Piergrossi 2015).

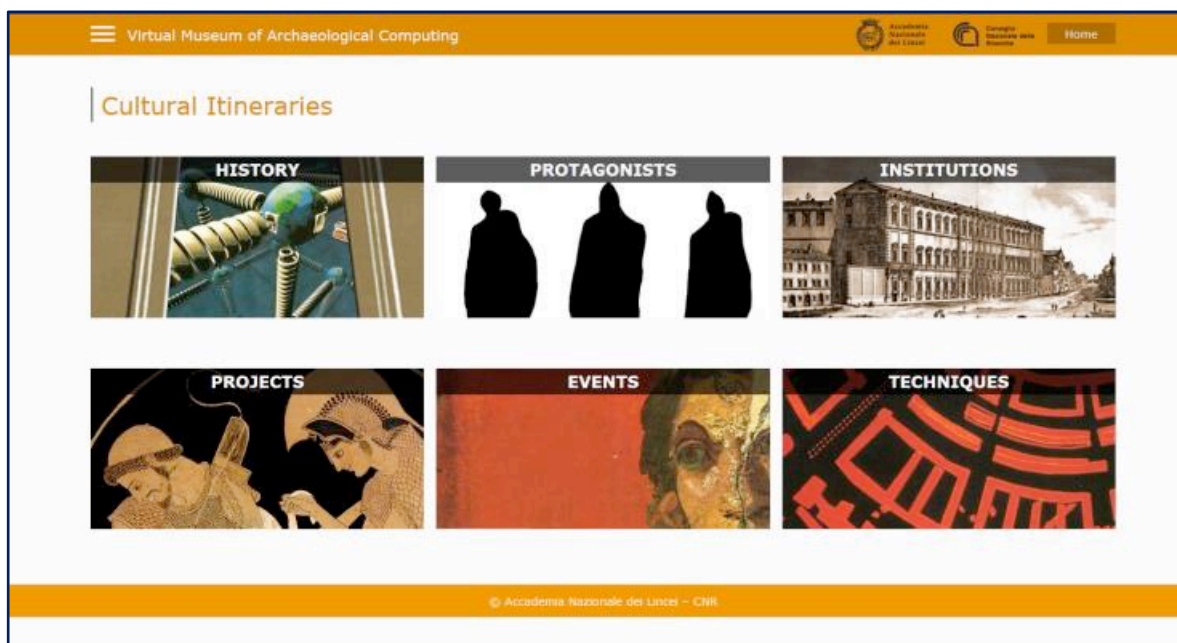


Figura 2. Cultural Itineraries

Oltre agli aspetti tecnici che caratterizzano il repository (Moscati 2012) e alle modalità scelte per la diffusione online delle risorse (ALLEA 2015) sono interessanti alcuni aspetti quantitativi che si riferiscono ai numeri della rivista finora pubblicati e ai Supplementi. Più di 800 articoli, oltre 1000 autori, più di 13.000 pagine e una classificazione rispondente a 20 soggetti principali, suddivisi tra applicazioni informatiche e settori della ricerca archeologica.

Gli articoli sono archiviati anche nel database SOLAR (Open-Access Scientific Literature Archive e Repository: <http://eprints.bice.rm.cnr.it/>), che partecipa a sua volta al più ampio e ambizioso progetto lanciato dal CNR dopo la firma della Dichiarazione di Berlino nel 2012 con il nome di Science & Technology Digital Library (<http://stdl.cnr.it/>). Una recente convenzione con l’Istituto Centrale per il Catalogo Unico consentirà a breve alla rivista di essere inserita tra le risorse del portale CulturalItalia, mentre il repository e la bibliografia di informatica archeologica compaiono già in DARIAH-IT tra i contributi in kind messi a disposizione dall’ISMA (<http://it.dariah.eu/sito/partner/isma/>).

4. CONCLUSIONI

Anche l’informatica archeologica, disciplina recente e in continua evoluzione, ha bisogno di mantenere viva la memoria delle sue radici e di valorizzare il suo patrimonio di risorse. A differenza di quanto avvenuto nella seconda metà del Novecento grazie al vivace dibattito sulla teoria e sui metodi di ricerca da adottare nell’incontro tra le “due culture”, oggi c’è meno interesse ad approfondire alcune questioni di rilevante importanza, legate ad esempio al significato rivoluzionario della rappresentazione della realtà attraverso modelli e procedure di codifica o all’importanza del campionamento sistematico delle informazioni per giungere alla fase interpretativa dei dati. Attratti dai risultati della scienza e della tecnica, spesso si perde di vista l’obiettivo finale della ricerca archeologica che consiste nella ricostruzione storica e nell’interpretazione del passato. A volte – cosa ancor più grave – gli archeologi non vengono neppure coinvolti in questo trionfo della tecnologia, dove tutto sembra essere realizzabile senza l’intervento degli esperti del settore.

Questo dunque è il significato del nostro progetto interdisciplinare, che s’inserisce in una linea di ricerca che ha avuto origine al CNR oltre trent’anni fa. Certamente esso richiede uno sforzo di coordinamento internazionale per non disperdere un patrimonio teorico e metodologico di conoscenze e renderlo disponibile a studiosi, studenti e a un pubblico più ampio.

5. BIBLIOGRAFIA

- [1] ALLEA E-Humanities Working Group Report. *Going Digital: Creating Change in the Humanities*. Berlin 2015: ALLEA. https://www.rd-alliance.org/sites/default/files/attachment/Going%20Digital_ALLEA%20May%202015.pdf.
- [2] *Archivi informatici per il patrimonio culturale, Atti del Convegno Internazionale (Roma 2003)*. Contributi del Centro Linceo Interdisciplinare "Beniamino Segre" 114. Roma 2006: Accademia Nazionale dei Lincei. http://archaeologicalcomputing.lincoln.it/attachment/114_2006.pdf.
- [3] Barchesi, C. 2006. «Archeologia e Calcolatori»: nuove strategie per la diffusione di contenuti in rete sulla base dell'OAI-PMH. In: *Archeologia e Calcolatori*, 16: 225-241. http://www.archcalc.cnr.it/indice/PDF16/Barchesi_16_2005.pdf.
- [4] Caravale, A., Piergrossi, A. 2012. *Archeologia in rete. Le riviste open access: risorse e prospettive*. In: *Archeologia e Calcolatori*, 23: 187-207. http://soi.cnr.it/archcalc/indice/PDF23/11_Caravale_Piergrossi.pdf.
- [5] Caravale, A., Piergrossi, A. 2015. *Archaeological open access journals: The case of «Archeologia e Calcolatori»*. In: Giligny et al. 2015: 275-263.
- [6] Eco, U. 1988. *Osservazioni sulla nozione di giacimento culturale*. In: *Le isole del tesoro. Proposte per la riscoperta e la gestione delle risorse culturali*. Milano: IBM Italia: 13-42.
- [7] Ferrari, O. 1991. *La catalogazione dei beni archeologici e le tecnologie informatiche*. In: *Archeologia e Calcolatori*, 2: 13-17. http://www.archcalc.cnr.it/quotations/ac_02_Ferrari.pdf.
- [8] Ferrari, O. 1996. *Archeologia e calcolatori nella prospettiva poli-disciplinare della tutela*. In: Moscati 1996: 805-808. http://archaeologicalcomputing.isma.cnr.it/itineraries/wp-content/uploads/2016/05/07_AC_V2S07.pdf.
- [9] Ferrari, O. 1998. *Beni culturali*. In: *Enciclopedia del Novecento*. II Suppl. Roma: Istituto della Enciclopedia Italiana Treccani.
- [10] Galluzzi, P. 2010. *Museo virtuale*. In: *XXI Secolo. I, Norme e Idee*. Roma: Istituto della Enciclopedia Italiana Treccani. [http://www.treccani.it/enciclopedia/museo-virtuale_\(XXI-Secolo\)/](http://www.treccani.it/enciclopedia/museo-virtuale_(XXI-Secolo)/).
- [11] Gardin, J.-C. 1991. *Le calcul et la raison. Essais sur la formalisation du discours savant*. Paris: EHESS.
- [12] Giligny, F., Djindjian, F., Costa, L., Moscati, P., Robert S. (eds.). *Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Paris 2014)*. Oxford: Archaeopress.
- [13] Guédon, J.-C. 2009. *It's a repository, it's a depository, it's an archive...: open access, digital collections and value*. In: *Arbor: Ciencia, Pensamiento y Cultura*, 85 (737): 581-595. <http://arbor.revistas.csic.es/index.php/arbor/article/viewFile/315/316>.
- [14] Huggett, J. 2013. *Disciplinary issues: challenging the research and practice of computer applications in archaeology*. In: G. Earl et al. (eds.), *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (Southampton 2012)*. Amsterdam: Amsterdam University Press: 13-24. <http://eprints.gla.ac.uk/92301/>.
- [15] *I modelli nella ricerca archeologica, Atti del Convegno Internazionale (Roma 2000)*. In: Contributi del Centro Linceo Interdisciplinare "Beniamino Segre" 107, Roma 2003: Accademia Nazionale dei Lincei. http://archaeologicalcomputing.lincoln.it/attachment/107_2003.pdf.
- [16] Moscati, P. (ed.). 1996. *III International Symposium on "Computing and Archaeology" (Roma 1995)*. In: *Archeologia e Calcolatori*, 7. http://archaeologicalcomputing.lincoln.it/index.php?en/254/symposium-rome-1995-_-proceedings.
- [17] Moscati, P. (ed.). 2009. *La nascita dell'informatica archeologica, Atti del Convegno Internazionale (Roma 2008)*. In: *Archeologia e Calcolatori*, 20. http://www.progettocaere.rm.cnr.it/databasegestione/open_block_pages.asp?IDyear=2009-01-01.
- [18] Moscati, P. 2012. *Archeologia e Calcolatori*. In: N. Palazzolo (ed.), *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze 2011)*. Collectanea Graeco-Romana 10. Torino: 75-79.
- [19] Moscati, P. 2013. *Jean-Claude Gardin (Parigi 1925-2013). Dalla meccanografia all'informatica archeologica*. In: *Archeologia e Calcolatori*, 24: 7-24. http://www.archcalc.cnr.it/indice/PDF24/01_Moscati.pdf.
- [20] Moscati, P. 2014. *Parcours culturels pour une histoire de l'informatique appliquée à l'archéologie*. In: L. Costa, F. Djindjian, F. Giligny (eds.), *Actes des 3èmes Journées d'Informatique et Archéologie de Paris - JIAP 2012 (Paris 2012)*. In: *Archeologia e Calcolatori*, Suppl. 5: 9-17. http://soi.cnr.it/archcalc/indice/Suppl_5/01_Moscati.pdf.
- [21] Moscati, P. 2015. *Towards a history of archaeological computing*. In: Giligny et al. 2015: 9-15.
- [22] Moscati, P. 2016. *Jean-Claude Gardin and the Evolution of Archaeological Computing*. In: F. Djindjian, P. Moscati (eds.), *Dossier: Jean-Claude Gardin (1925-2015)*. *Les nouvelles de l'archéologie*, 144: 10-13. <http://nda.revues.org/3457>.
- [23] Natale, M. T., Saccoccio, R. 2010. *Museo and Web: un kit pratico per le istituzioni culturali che vogliono realizzare un sito web di qualità*. In: *Archeologia e Calcolatori*, 21: 27-47. http://www.archcalc.cnr.it/indice/PDF21/02_Natale_Saccoccio.pdf.
- [24] *Rapporto sui progetti ex art. 15 Legge 41/1986*, Catalogo della Mostra. Roma 1989.
- [25] *Rediscovering Pompeii*, Catalogo della Mostra. Roma 1988: L'Erma di Bretschneider. <http://archaeologicalcomputing.lincoln.it/index.php?en/80/exhibitions>.

Exploiting ICT to Create Smart Visit Experiences at Cultural Heritage Sites

Carmelo Ardito¹, Paolo Buono¹, Maria Costabile¹, Danilo Caivano¹,
Giuseppe Desolda¹, Rosa Lanzilotti¹, Maristella Matera², Antonio Piccinno¹

¹ Università degli Studi di Bari Aldo Moro, Bari, Italy - {name.surname@uniba.it}

² Politecnico di Milano, Milano, Italy - maristella.matera@polimi.it

ABSTRACT

Information and communication technologies have a great potential to increase awareness and appreciation of cultural heritage. In the last decade, at the Computer Science Department of the University of Bari (Italy), we have been working to the design, development and evaluation of software systems supporting people during their visits to historical sites and museums, with the goal of improving the overall user experience. We focused on pervasive games to offer an engaging way to experience the cultural heritage site. This article briefly describes some games we developed and presents some of our new recent approaches. Specifically, we are investigating possible use of the Internet of Things (IoT) technology, as well as Mixed Reality, in order to ensure a more engaging visit experience and to increase the appropriation of CH contents by visitors.

KEYWORDS

Internet of Things, Serious games, Mashup

1. INTRODUCTION

Cultural heritage assets keep alive the history of a territory and of its inhabitants. Italy has one of the richest patrimony of historical sites and Cultural Heritage (CH) in the world. Information and Communication Technologies (ICT) can effectively provide new ways to support people's understanding of CH and make visits to CH sites more engaging.

In this article we briefly illustrate the work we have performed and the systems that we have developed in the CH application domain together with some ongoing research. Specifically, in Section 2 we report about serious games mainly addressing pupils of elementary and middle school, while in Section 3 and Section 4 we describe novel approaches for improving the experience of people visiting CH sites, which exploit novel ICTs such as Internet of Things and Mixed Reality, respectively. Section 5 concludes the article.

2. SERIOUS GAMES FOR VISITING CH SITES AND LEARNING MORE ABOUT HISTORY

It is well known that pupils make up a large proportion of visitors to archaeological sites. However, traditional visits generate little interest in young visitors. New approaches must be proposed in order to capture student's attention and to engage them. Gameplay showed to be successful. Pervasive games break away from the usual static paradigm of room play and go towards a more dynamic and social experience, by combining game and physical reality.



Figure 1. Two different phases of Explore!, a pervasive serious game for visiting archaeological parks: play and debriefing.

In the last ten years, we have created serious games on different devices, not only to engage pupils in a more enjoyable visit to CH sites, but also to support them to reflect and deepen knowledge acquired through their active involvement during visits to these sites and followup activities at school. These games have been developed by setting up multidisciplinary teams that included experts in the CH domain, e.g., archaeologists, directors and employees of archaeological parks, software developers and HCI experts, schoolteachers, as well as end users, i.e., children and other

types of visitors. We developed Explore!, a mobile learning framework incorporating a pedagogical technique in the form of the excursion-game, and a reusable technological architecture, which can be adapted to different historical sites by inserting appropriate content [1, 2]. We experimented it with several classes of middle schools and fifth grade of elementary schools visiting the archaeological park of Egnazia in the Apulia region. Special attention has been devoted to the design of a soundscape, which improves players' navigation in degraded physical environments and enrich their overall experience [3]. The left image of Figure 1 shows pupils deciding the game strategy; the backpack brought by a team member contains two speakers that reproduce ancient sounds while moving in the site. The right image of Figure 1 shows the debriefing phase after the visit, during which students consolidate the acquired knowledge by discussing together on the paths they followed in the archaeological park and the items they found.

Other serious games aiming at fostering visitors' learning of history have been proposed on large interactive displays. For example, History-Puzzle requires that pupils to complete puzzles of historical monuments/objects they saw during the visit to a cultural heritage site [22]. Time-Voyager is another serious game proposed on large interactive displays that invites pupils to organize photos of ancient artefacts according to a chronological order corresponding to the different historical eras [4, 23] (see Figure 2).



Figure 2. Pupils playing Time voyager, a serious game on large interactive displays.

3. MASHUP AND IOT TECHNOLOGIES TO ENHANCE THE VISIT EXPERIENCE

The discussions with the stakeholders involved in the multidisciplinary teams and various field studies provided inspiration about new ways of using current technology to support the work of the professional guides in organizing and conducting the visits, with the double goal of enhancing the visitors' experience on the site and extending the overall experience beyond the specific visit. In the last years, also in collaboration with researchers of the "Politecnico di Milano", we have been developing EFESTO, a platform that, according to End-User Development (EUD) [5-7] and participatory design approaches [8], supports end users, not technology skilled, to create personalized visits to CH sites [9]. End users become both information consumers and producers, and they are actively involved in the management of CH information. The platform, based on mashup technology [10], implements a new composition paradigm to allow end users to extract contents from heterogeneous (personal or third-party) sources, and compose Interactive Workspaces that satisfy their situational information needs and can be ubiquitously executed on different devices. The platform is general and is flexible enough to be adopted in different contexts of use [11]. In the CH domain, we have used to support the work of professional guides when accompanying visitors to archaeological parks, also allowing guides and visitors to collaborate among them to create new visit experiences [12].

The Internet of Things (IoT) has recently emerged as a technology able to enhance the access to CH collections. It promotes the use of smart objects, i.e., physical devices connected to the Internet, equipped with sensors and/or actuators and embedded software. Visitors of CH sites can touch, manipulate and interact with smart objects, for example to receive personalized information during the visit. Such interaction fosters the appropriation of CH content [13]: as recognized by some recent works [14]. Indeed, physical manipulation is an effective supplementary channel for the visitor to gather and store information, since they are enabled to activate real-world knowledge. In the CH domain, research has especially focused on technical aspects related to the development of sensor and actuator infrastructures [13]. A few works have studied how non-technical users can configure smart objects to define smart visit experiences. To fill this gap, we implemented our End-User Development approach within the EFESTO platform [15], in order to support users without skills in computer programming to configure the behavior of smart objects by creating Event-Condition-Action (ECA) rules [15, 24]. The platform empowers professional guides to enhance visit experiences by managing multiple synchronized smart objects. A peculiarity of this new version of EFESTO is the adoption of a model, called 5W, that introduces some specification constructs (Which, What, When, Where, Why) to build ECA rules coupling multiple events and conditions exposed by smart objects, and to define temporal and spatial constraints on rule activation and actions execution. Once a professional guide has defined the behavior of certain smart objects placed in the CH site, visitors are

enabled to acquire more tangible CH content by interacting with the surrounding environment and the smart objects included in it.



Figure 3. a) the guide introduces a visit to the archaeological park of Egnazia by using the EFESTO platform on a multitouch screen. b) the guide and visitors sharing information through the EFESTO platform deployed on the guide's tablet and visitors' smartphones.

4. EXPLOITING MIXED REALITY FOR CH SITES

Technological advances are fostering the proliferation of low-cost devices, including head-mounted displays, that make more real and natural the human-computer interaction experiences, as for example Virtual Reality (VR), Substitutional Reality (SR), Augmented Reality (AR), and Mixed Reality (MR). VR consists in an immersive multimedia reality that reproduces an environment simulating a physical presence in real or imaginary interactive worlds [16]. SR is a class of VR where every physical object surrounding a user is paired, with some degree of discrepancy, to a virtual counterpart [17]. AR is a live, direct or indirect view of a physical, real-world environment whose elements are augmented by virtual contents, which are not anchored to the real worlds and are not able to respond to each other [18]. Mixed Reality merges real and virtual worlds to yield new hybrid environments where physical objects and interactive digital artifacts are blended [19-20]. In other words, in VR digital environments shut out the real world, in AR digital content is on top of the real world, in MR digital content interacts with the real world. While MR is widely considered as one of the most promising technology, its potential is still unexploited. In the last few years, we have been working on the design, development and evaluation of MR applications based on Microsoft HoloLens1. HoloLens is a holographic computer built into a headset that lets users see, hear, and interact with holograms within a real environment. The real world is used as a 3D canvas in which to position and show virtual objects. Mid-air hand gestures is the most used technique to interact with holograms. MR devices are very appealing. However, we observed several issues related to interaction. In particular, as happened in the past with other novel technologies, companies and researchers spent most of their effort on improving MR hardware, giving much less attention to the design of heuristics, frameworks, guidelines and models driving the creation of MR worlds and the interaction with them. MR devices require specifically designed interaction techniques, but designers improperly adapt those of other traditional devices. For example, when the multi-touch displays appeared [21], users were asked to use the fingers as a mouse pointer, losing the potentialities offered by the simultaneous use of multiple fingers or multiple hands. During the years, it became evident the need of designing ad-hoc interaction mechanisms. A similar story is happening MR-devices, which are often implemented by converting traditional mouse interaction in mid-air gestures. In Cultural Heritage, MR could reveal a fruitful approach in a situation like the following.

The stored content can be exploited to rebuild location using MR and HoloLens. The tourists, through such technologies, could interact with the artifacts retrieved in situ or in other places, but relevant for the location, in order to study them in their original or expected context. Examples are lost items or reproduction of items exhibited in other places. This is a research that we aim to investigate in the near future.

5. CONCLUSION

In this paper we have described some research we have performed in the last decade. We have also presented new approaches that may have great potential to enhance the experience of visitors of CH sites by allowing them either to manipulate smart objects available at the site, exploiting IoT technology, or by permitting the interaction with virtual objects through Mixed Reality. Our ongoing and future research aims to improve the support CH sites experts in designing smart visit experiences. To this aim, we are currently investigating novel natural interaction paradigms to enrich smart objects with custom attributes, which can be used to create ECA rules in an even simpler and more powerful way.

6. REFERENCES

- [1] Costabile M. F., De Angeli A., Lanzilotti R., Ardito C., Buono P., and T. Pederson. 2008. Explore! possibilities and challenges of mobile learning. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 145-154. DOI: <https://doi.org/10.1145/1357054.1357080>
- [2] Ardito C., Buono P., Costabile M. F., Lanzilotti R., Pederson T., Piccinno A. (2008). Experiencing the Past through the Senses: An M-Learning Game at Archaeological Parks. *IEEE Multimedia*. vol. October-December 2008, pp. 76-81. ISSN: 1070-986X
- [3] Ardito C., Costabile M. F., De Angeli A., and Lanzilotti R. (2012). "Enriching exploration of archaeological parks with mobile technology," *ACM Trans. Comput.-Hum. Interact.*, vol. 19, pp. 1-30, 2012.
- [4] Ardito C., Costabile M. F., and Lanzilotti R. (2010). "Gameplay on a Multitouch Screen to Foster Learning about Historical Sites". *International Congerence on Advanced Visual Interfaces*, Rome, Italy, 2010.
- [5] Ardito C., Buono P., Costabile M. F., Lanzilotti R., and Piccinno A. (2012). "End users as co-designers of their own tools and products," *Journal of Visual Languages & Computing*, vol. 23, pp. 78-90, 2012.
- [6] Costabile M. F., Fogli D., Mussio P., and Piccinno A. (2007). "Visual Interactive Systems for End-User Development: A Model-Based Design Methodology," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, pp. 1029-1046, 2007.
- [7] Lieberman H., Paternò F., and Wulf V., Eds., *End User Development*. Springer, 2006.
- [8] Schuler D., *Participatory design: principles and practices*. Hillsdale N.J.: L. Erlbaum Associates, 1993.
- [9] Ardito C., Costabile M. F., Desolda G., Lanzilotti R., Matera M., Piccinno A., and Picozzi M. (2014). "User-driven visual composition of service-based interactive spaces," *Journal of Visual Languages & Computing*, vol. 25, pp. 278-296, 2014.
- [10] Daniel F. and Matera M., *Mashups - Concepts, Models and Architectures*. Springer Series: Data-Centric Systems and Applications, XIX, 319. 2014.
- [11] Ardito C., Bottoni P., Costabile M. F., Desolda G., Matera M., and Picozzi M. (2014). "Creation and Use of Service-based Distributed Interactive Workspaces," *Journal of Visual Languages & Computing*, vol. 25, pp. 717-726, 2014.
- [12] Ardito C., Bottoni P., Costabile M. F., Desolda G., Matera M., Piccinno A., and Picozzi M. (2013). "Enabling End Users to Create, Annotate and Share Personal Information Spaces," in *End-User Development - IS-EUD 2013*. vol. LNCS 7897, Dittrich Y., Burnett M., Mørch A., and Redmiles D., Eds., ed Berlin Heidelberg: Springer, 2013, pp. 40-55.
- [13] Mighali V., Fiore G. D., Patrono L., Mainetti L., Alletto S., Serra G., and Cucchiara R. (2015). "Innovative IoT-aware Services for a Smart Museum," presented at the International Conference on World Wide Web, Florence, Italy, 2015.
- [14] Uckelmann D., Harrison M., and Michahelles F. (2011). "An Architectural Approach Towards the Future Internet of Things," in *Architecting the Internet of Things*, Uckelmann D., Harrison M., and Michahelles F., Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1-24.
- [15] Desolda G., Ardito C., and Matera M. (2017). "Empowering End Users to Customize their Smart Environments: Model, Composition Paradigms, and Domain-Specific Tools," *ACM Trans. Comput.-Hum. Interact.*, vol. 24, pp. 1-52, 2017.
- [16] Burdea G. C. and Coiffet P. *Virtual reality technology*. London: Wiley-Interscience, 1994.
- [17] Simeone A. L. and Velloso E. (2015). Substitutional reality: bringing virtual reality home. *XRDS* 22, 1 (November 2015), 24-29. DOI: <https://doi.org/10.1145/2810044>
- [18] Azuma R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and virtual environments* 6(4), pp. 355-85.
- [19] Milgram P. and Kishino F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems* 77(12), 1321-9.
- [20] Ohta Y. and Tamura H. 2014. *Mixed Reality: Merging Real and Virtual Worlds*. Springer Publishing Company, Incorporated.
- [21] Ardito C., Buono P., Costabile M. F., and Desolda G. (2015). Interaction with large displays: a survey. *ACM Computing Survey* 47(3), pp. 1-38.
- [22] Ardito C., Lanzilotti R. (2011). An EUD Approach to the Design of Educational Games. *International Journal of Distance Education Technologies*, vol. 9(4); pp. 25-40, ISSN: 1539-3100, doi: 10.4018/IJDET.2011100103.
- [23] Ardito C., Costabile M. F., Lanzilotti R., Desolda G. (2013). Integrating traditional learning and games on large displays: an experimental study. *Educational Technology & Society*, vol. 16(1); pp. 44-56.
- [24] Ardito C., Buono P., Desolda G., Matera M., From smart objects to smart experiences: An end-user development approach, *International Journal of Human-Computer Studies*, ISSN 1071-5819, <https://doi.org/10.1016/j.ijhcs.2017.12.002>.

Get Ready, Get Set, Curate: Understanding the ‘Everyday Curator’

Michele Reilly,¹ Santi Thompson²

¹ University of Arkansas Libraries, U.S.A., - reilly@uark.edu

² University of Houston Libraries, U.S.A., - sathompson3@uh.edu

ABSTRACT

While conducting a variety of studies on the reuse of digital images across the web, the presenters Michele Reilly and Santi Thompson found that social media users were collecting, organizing, and sharing images in a manner that was not unlike the tasks performed by archivists, librarians, and other cultural heritage professionals -- that there was something else occurring beyond ‘reuse’.

The researchers believed that the Pinterest users were, knowingly or not, becoming everyday curators. To quantify these actions the researchers closely analyzed the Pinterest users and the ways in which they were organizing images.

This presentation focuses on understanding the characteristics of those who are engaged in this everyday curation process. Using an existing Pinterest dataset, the presenters developed a rubric for understanding these “everyday curators.” This presentation will: define curators; review the methodology used to devise and classify the everyday curator rubric; discuss the results of applying this rubric to the Pinterest dataset; and conclude with implications for digital humanists and cultural heritage professionals.

KEYWORDS

Social Media, Pinterest, Everyday Curators, Digital Images, Social Curation

1. INTRODUCTION

In a recent episode of CBS Sunday Morning, correspondent Faith Salie [1] explored the changing perceptions of “curation” over time, noting that in contemporary society curation has saturated our popular culture -- a “le deluge” as Salie puts it. She exclaims “Nowadays everyone’s a curator; no Ph.D. in Art History required. Got an opinion? Got content? Curate!” The proliferation of the world wide web and, with it, ways to communicate and share experiences virtually through social media sites, has changed the ways that people interact with online cultural heritage materials. As the accessibility of the web increases over time, an individual’s potential ability to “curate” occurs from nearly anywhere - making it easier for everyday people to engage in the curation process. These amateur curatorial practices may include how users transform digital objects as a means of self-expression, political, social, and personal commentary, cultural products, and artistic license. Recognizing how everyday users curate digital images enhances the digital humanists’ understanding of the contemporary cultural landscape.

This presentation focuses on understanding the characteristics of those who are engaged in this everyday curation process. Using an existing Pinterest dataset, presenters Michele Reilly and Santi Thompson developed a rubric for understanding these “everyday curators.” This presentation will: define curators; review the methodology used to devise and classify the everyday curator rubric; discuss the results of applying this rubric to the Pinterest dataset; and conclude with implications for digital humanists and cultural heritage professionals.

2. WHAT IS A CURATOR? EXPANDING THE DEFINITION OF CURATORS

Wikipedia, which identifies multiple types of curation (as well as curators), defines content curation as the act of “collecting and sorting of content” [2]. Targeted more closely to the archival profession, the Society of American Archivists (SAA) have defined an archivist as “an individual responsible for organising, prioritising and editing information and objects to records of enduring value, according to the principles of provenance, original order, and collective control to protect the materials’ authenticity and context” [3]. Munoz and Renear define curation as “gathering material, making it discoverable by describing and organizing it, placing it in a context of related information, supporting its use for diverse intellectual purposes, and ensuring its long-term survival” [4]. These definitions highlight the importance of the curatorial lifecycle, which encompasses selecting and preserving content for the aim of access.

Archivists, librarians, and others in cultural heritage professions have traditionally been the ideal benchmark of a curator. The advent of technology and the web have begun to obscure this idea. Milena Droumeva notes that the role individuals play in media production “shifts the critical discourse on aesthetic expression from the realm of specialised expertise to general practice” [5] She connects the personal devices used by ordinary users as a driver for their expanded curatorial roles. “The act of curation is filtered through the aesthetic and technological capabilities of the smartphone, a device that has become co- constitutive of our routine sensorial encounters with the world” [5]. Social curation, as described by Joyce Seitzinger, is “the discovery, selection, collection and sharing of digital artefacts by an individual for a social purpose such as learning, collaboration, identity expression or community participation” [6]. Rhema Linder, Clair Snodgrass, and Andruid Kerne’s write that “everyday ideation,” a concept similar to social curation, is the “ongoing processes in which [everyday] curators look for, find, organize, and return to meaningful information as a means of provoking and forming ideas that address practical and emotional needs,” [7]. Adding on the idea of everyday curators, O’Neill argues that “everyone within an online social circle” is impacting on each other through “finding, filtering and recommending” content

[8]. Through this networked activity, O'Neill believes that social archives emerge that are "less official, but no less formal" than traditionally-curated collections [8].

3. THERE'S SOMETHING ELSE GOING ON AROUND HERE

While conducting a variety of studies on the reuse of digital images across the web, the presenters found that social media users were collecting, organizing, and sharing images in a manner that was not unlike the tasks performed by archivists, librarians, and other cultural heritage professionals -- that there was something else occurring beyond 'reuse'.

The researchers believed that the Pinterest users were, knowingly or not, becoming everyday curators. To quantify these actions the researchers closely analyzed the Pinterest users and the ways in which they were organizing images.

To conduct this study, the researchers had to complete a two-step process to develop and classify their dataset. First, the presenters had to develop the everyday curator rubric. To do this, they drew upon Sophia Liu's Socially distributed curation model and Rebecca O'Neill's Citizen curator model to understand the behavior of social media users.

Liu developed the socially distributed curation model based on the premise that social media sites were increasingly becoming "the new vehicle for... sharing memories and stories that are perceived to be historically valuable as well as personally meaningful" [9]. Liu's model emphasizes the activities conducted by a diverse range of participants, as opposed to specific roles, such as "archivist" or "digital preservation officer." To develop this model, Liu "synthesized the meaning of curators and curation in the social web context" [9]. Her model addresses the following seven curatorial activities: (1) Preserve and maintain, (2) Collect and archive, (3) Categorize and organize, (4) Edit and verify, (5) Craft a story and synthesize, (6) Exhibit and juxtapose, and finally (7) Guide and converse [9].

In her 2014 article "Recognizing the Citizen Curator," O'Neill envisioned, in the not so distant future, an information universe where everyday people were acquiring, organizing, and sharing digital objects in new and personalized ways. These everyday people emerged as "citizen curators" because they could leverage the Web to perform many of the functions once believed to be done primarily by "professional curators" [8]. O'Neill situates the citizen curator model around four distinct characteristics: (1) self-directed, (2) self-selected expert, (3) object based, and (4) information based [8].

The presenters synthesized these two models, extracted social media actions, and cross-walked these actions into some of the roles of traditional curators. From this process, they determined that the everyday curator exhibited five characteristics that they could potentially measure using their existing dataset (see detailed information below in table 1).

Second, they coded Pinner's reuse of digital images against these five characteristics. After coding, they eliminated Pinterest entries that were posted by organizations or corporations because these entities did not reflect the curatorial practices of individuals.

Characteristics	Definition
Identity	Anyone who selects, organizes, and distributes digital content [1]
New metadata description or tag	Terms describing objects pinned on pinboards. These terms not necessarily align with archival or library best practices, but could be the products of "crowd-sourced metadata" [1]
Personal Narrative	A narrative, in story or other forms, from a user's life that reflects their own experiences.
Remixing	The transformation of an object to create a new product that has traces of the original content. An example of this would be a meme.
Reorganizing	Placing objects in a new organizational scheme or topic than its original intent .

Table 1

4. DISCOVERING THE EVERYDAY CURATOR

The presenters evaluated for each characteristic in the rubric and found:

- *Identity*: A majority of pinners do not self-identify (64%) by writing a descriptive narrative about themselves in the "about you" portion of their Pinterest profile;
- *New metadata description or tag*: A majority of pinners (62%) do not change the metadata from the original;
- *Personal Narrative*: While a majority of pinner's boards (74%) reflect a diverse set of interests, only a quarter of pinners either state that they are telling a personal narrative or have boards that reflect a personal narrative;
- *Remixing*: One in five pinners (21%) shared remixed images;
- *Reorganizing*: Over two-thirds of pinners (72%) created a new organizational schema for reused images.

The active, conscious decision to organize pins into themed pin boards reflects the most prominent characteristic of the everyday curator. Examples from the data show the range of organizational schema. Some examples include photographs of slaves taken by Mathew Brady (which were originally organized into a lesson plan about Matthew Brady and his work found at the Library of Congress' Teaching with Primary Resources website) and were reorganized by pinners into pinboards specifically addressing slavery (not Matthew Brady). Other pinners curated images into pinboards that were not

closely related to the original organizational scheme. One pinner included images from the Great Depression in a pinboard dedicated to denim fashion. This reorganizing behavior has the potential to give the pinner the opportunity to tell a story, whether their own or someone else's.

The presenters will elaborate on the other characteristics during the AIUCD 2018 conference presentation.

5. WHY SHOULD WE CARE ABOUT THE EVERYDAY CURATOR?

The growing participation of everyday curators in curatorial activities means that the social web has offered everyday people the opportunity to comment on, critique, and ultimately change the personal, cultural, social, and political landscapes in ways that may or may not be intentional by the process of interacting, transforming, and curating digital cultural heritage materials. Digital humanists and cultural heritage professionals will need to be aware of how everyday curators are telling their stories and their perceptions of themselves on social media platforms [10]. As Julia Kirby writes, "The advent of social media has given us a chance to deploy our cognitive power very differently: actively participating in the creation of content rather than passively receiving it, and doing so in concert with others, not individually" [11].

Digital humanists and cultural heritage professionals will also need to recognize how everyday curators are repurposing digital projects to construct new meanings and dialogues. They should view the narratives derived from social media platforms as legitimate reflections on culture and should be preoccupied with compiling these narratives, analyzing them, and valuing them as viable cultural artifacts.

6. REFERENCES

- [1] Salie, F. (2017, January 8). "Faith Salie on a life well curated." CBS Sunday Morning. Accessed January 8, 2017. Retrieved from: <http://www.cbsnews.com/news/faith-salie-on-a-life-well-curated/>
- [2] Wikipedia contributors, "Curation," *Wikipedia, The Free Encyclopedia*, <https://en.wikipedia.org/w/index.php?title=Curation&oldid=733565355> (accessed August 8, 2016).
- [3] Society of American Archivist, "Manuscript Curator" *Glossary of Archival and Records Terminology*, <http://www2.archivists.org/glossary/terms/m/manuscript-curator> (accessed August 8, 2016)
- [4] Muñoz, T., & Renear, A. H. (2011). Issues in humanities data curation. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/30852/MunozRenear_2011_IssuesinHumanitiesDataCuration.pdf?sequence=2&isAllowed=y.
- [5] Droumeva, M. (2015). Curating Everyday Life: Approaches to Documenting Everyday Soundscapes. *Media - Culture Journal*, 18,4
- [6] Seitzinger, Joyce. "Curate Me! Exploring online identity through social curation in networked learning." In *9th International Conference on Networked Learning*, pp. 7-9. 2014.
- [7] Linder, Rhema, Clair Snodgrass, and Andruid Kerne. "Everyday ideation: All of my ideas are on Pinterest." In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 2411-2420. ACM, 2014.
- [8] O'Neill, Rebecca. "Recognising the Citizen Curator." In *Virtual Cultural Heritage in Ireland, The State-of-the-Art and Visions for the Future* (2014): https://www.academia.edu/10279201/Recognising_the_Citizen_Curator
- [9] Liu, Sophia B. "Socially distributed curation of the Bhopal disaster: a case of grassroots heritage in the crisis context." *Heritage and Social Media: Understanding and Experiencing Heritage in a Participatory Culture*, Cambridge, UK: Routledge (2012).
- [10] Good, Katie Day. "From scrapbook to Facebook: A history of personal media assemblage and archives." *New Media & Society* 15, no. 4 (2013): 557-573.
- [11] Kirby, Julia. (2010). Brains to spare. (Cognitive Surplus: Creativity and Generosity in a Connected Age) (Brief article) (Book review). *Harvard Business Review*, 88(7 8), 166.

Invading the Italian Literary Polysystem. A Distant Reading Case Study of Sienkiewicz's "Quo vadis"

Jan Rybicki¹, Katarzyna Biernacka-Licznar², Monika Woźniak³

¹ Uniwersytet Jagielloński, Kraków, Poland - jkrybicki@gmail.com

² Uniwersytet Wrocławski, Wrocław, Poland - katarzyna.biernacka-licznar@uwr.edu.pl

³ Università degli Studi di Roma "La Sapienza", Italy - moniwozniak@gmail.com

ABSTRACT

This paper explores and discusses the phenomenon of the Italian fortunes of *Quo vadis* (1896), the historical romance set in Ancient Rome by the Polish literary Nobel Prize winner (1905) Henryk Sienkiewicz (1846-1916). The problem is addressed through stylometric analysis of word frequencies in the numerous Italian translations of the novel, set within the context of a sizeable (1000 texts) corpus of original Italian literature. Frequencies of most frequent words and medium-frequent words are compared using the well-known Burrows procedures of Delta and Zeta; word frequency strings will be compared through cluster analysis, and these will be visualized by network analysis. The results show the extent to which a translated work of literature - in fact, its many different translations - becomes part of the receiving culture, and to what extent it remains a "foreign object" discrete from the national literature it invades. This distant-reading/macroanalytic/quantitative approach will be discussed in terms of Even-Zohar's polysystem theory. In this way, a single case study will also serve as a reconnaissance into a possible marriage of the visions of culture of Even-Zohar and those of Moretti and Jockers.

KEYWORDS

literary translation, literary polysystem(s), distant reading, stylometry, network analysis.

1. INTRODUCTION

In 2011, the Forum Romanum hosted an exhibition devoted to the times of Emperor Nero. Interestingly, its illustrative material relied quite heavily on Polish academic painting of the turn of the 19th and the 20th centuries, in which various Neronian themes – mostly that of persecution of early Christians – seemed to have a niche of their own. It is quite obvious that this (Ancient) Roman interest among such painters as Siemiradzki or Styka was due to the international success of their compatriot's bestselling novel *Qui vadis* (1896) [15].

Henryk Sienkiewicz was already a major literary voice in his stateless country¹, but it was only his Roman novel that gained him enough respect to become the fifth-ever winner of the literary Nobel Prize (1905). This was followed by a veritable explosion in terms of numbers of translations into various languages, soon followed by theatre, opera and film adaptations; the first cinema version appeared as early as in 1901; two silent films were shot in Italy before 1924; and that interest did not subside even after the 1951 Hollywood production. This widespread reuse of Sienkiewicz's material – apart from the literary value of his work – must have been at least partially enhanced by the fact that the author, officially a Russian subject, was not protected by copyright laws of the time. Adaptations as well as translations could be made quite freely and without the writer's consent – and without royalties. No wonder, then, that, in many countries, several different translations vied for the public's attention. "Several" does not even begin to describe the situation in Italy, where as many as 60 different renditions can be still found today – and that in more than five hundred editions. In the first two years of Sienkiewicz's Italian invasion (1899-1900), as many 6 different translations, including two anonymous ones, were already available to the readers [2].

The particular Italian interest in *Quo vadis* is scarcely a mystery. Not only was the novel set in the Italian capital and not only did it deal with a subject already very present in Italian culture old and new; the book's (and its author's) brand of conservative Catholicism must have appealed to some of the most influential circles of the country. What is more, the novel – perhaps paradoxically – was equally praised by some of Italy's progressive critics, who saw, in Sienkiewicz's persecuted Christians, the struggle of their contemporary revolutionary and working-class movements, and who liked to read his depiction of Imperial Rome's decadence as a diatribe against the existing power structure [8]. At a time when the 19th-century lengthy historical romance was slowly becoming a genre of the past, the Italian success of *Quo vadis* was something of a phenomenon [10]. This is why it deserves a closer examination also from a quantitative, stylometric, distant-reading perspective.

2. MATERIAL AND METHODS

A corpus of Italian versions of Sienkiewicz's novels was collected as part of a greater project on the reception of Sienkiewicz in Italy; most of the texts had to be scanned and OCRed from copies in numerous libraries across Italy and Poland. It was supplemented with Italian translations of Sienkiewicz's other novels, a much smaller list. A collection of 400 19th- and 20th-century Italian novels by 95 authors was used to represent the Italian literary system; it was more or less balanced in terms of book per author ratio to avoid natural grouping of overrepresented authors, and this proved successful

¹ At the time, ethnic Polish lands were partitioned by Austria-Hungary, Prussia, and Russia.

in trial network analysis, with the exception of Salgari's adventure novels, which grouped outside the main network in irrespective of how many of these were used in the analysis (the same limitation had to be applied to the number of Sienkiewicz's works compared with this broader context). The control "native Italian" corpus yielded a strong authorial signal. It was also quite devoid of marked chronological evolution: quite understandably, since the corpus was dominated by a single genre, that of realist fiction, even if historical romances and adventure novels were also a sizeable presence. The method applied was that described by Eder [4] and applied to other literary corpora by Rybicki [12,13]. Basing on Burrows's Delta procedure [3], a list of most-frequent words is produced for the entire corpus. These words are then counted in the individual texts, and their frequencies are compared in text pairs to produce a matrix of distance measures; in this study, the distances were established by means of the modified Cosine Delta [14], which is now seen as the most reliable version [7]. The distance matrix then undergoes Cluster Analysis (Ward's Hierarchical clustering), resulting in grouping the texts into "clusters" of greatest similarity. The procedure is performed by means of stylo [5], a stylometric package for R [11]. The results are visualized by means of network analysis, applying the Force Atlas 2 gravitational algorithm [9] in Gephi [1].

3. RESULTS

Irrespective of the specific content of the Sienkiewicz corpus (i.e. of the specific translations of *Quo vadis* used in the experiments), the results were quite repetitive and consistent. As can be seen in Figure 1, the several versions of *Quo vadis* (in this case, some of the most successful ones, including those by Baldini, 1927; Fabietti, 1929; Foulques, 1900; Garosci, 1950; Minardi, 1906; Racovska and Fabietti, 1952; Valera, 1900; Verdinois, 1899; and two anonymous translations of 1899 and 1900) form a compact cluster (red) in the top-right quadrant of the graph. Translations of other Sienkiewicz texts (also marked red) are far away and mingle with the greater native Italian corpus. Interestingly, this visualization contains no trace of the fact that one of the *Quo vadis* translations was penned by Paolo Valera (1850-1926), one of the *veristi*: his own writings (violet) cluster at bottom-centre, and his translation of the Polish books bears no stylometric affinity to any of his novels.

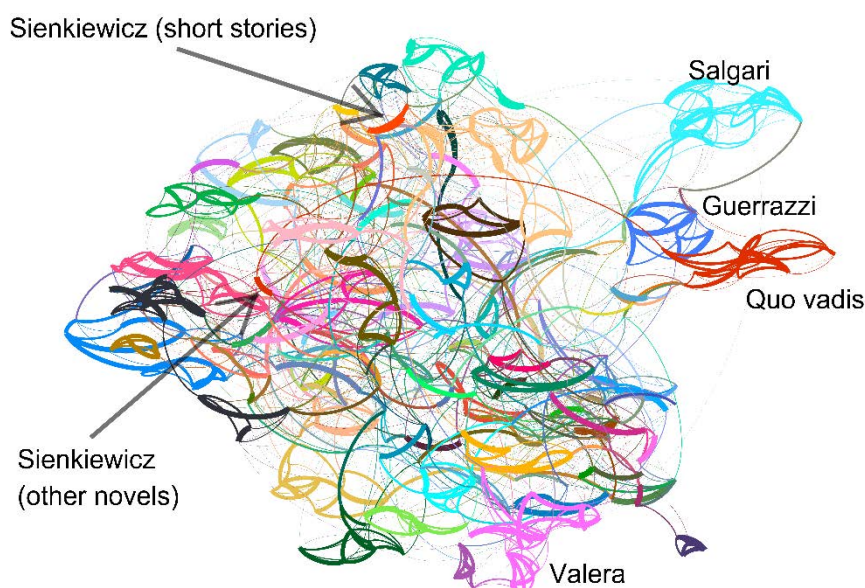


Figure 1. Network analysis of 400 Italian novels and Italian translations of Sienkiewicz, based on a comparison of frequencies of 100-2000 most frequent words.

Only two Italian authors remain in the vicinity of the invading Polish novel: the prolific and popular adventure writer Emilio Salgari (1862-1911), and the historical novelist and major *Risorgimento* figure, Francesco Guerrazzi (1804-1873).

4. DISCUSSION

Itamar Even-Zohar discusses the position of translated literature within a literary polysystem and identifies two possible situations. When the native system is not yet crystallized, or "weak", or in crisis, translated literature may assume a central, innovative, primary position. When literature in translation or a single translated work enters a well-established literary tradition, it usually remains at its periphery and does not participate in shaping the native polysystem unless it becomes strongly naturalized [6]. The position of Sienkiewicz's *Quo vadis* in the Italian literary polysystem of the 19th and 20th centuries seems to represent the latter possibility. The native Italian texts create a consistent whole – even if, in this case, the consistency is in the seemingly very "simple" matter of most-frequent-word frequencies. *Quo vadis* refuses to enter (or is refused entry) into its midst; instead, it seems to be suspended between two Italian subsystems: of historical romances produced by a writer of a previous generation, yet with a nationalist motivation very similar to Sienkiewicz's, and – to a lesser extent – of adventure novels, the work of a writer, Sienkiewicz's approximate contemporary, whose Italian

popularity was just as tremendous as Sienkiewicz's Polish popularity still is. This is interesting: the combination of "historical romance" and "adventure novel" could very much be attached to much of Sienkiewicz's *oeuvre*, but the latter not necessarily to *Quo vadis*; and yet that is exactly the position that the Polish bestseller seems to capture and hold in its foray into Italian literature. What is even more interesting, this early reconnaissance seems to suggest that distant reading may serve as an illustration of the application of polysystem theory to literary translation. If this is confirmed by other similar studies, this could provide scholars with an exciting way to describe interactions between literary polysystems that are such an important part of the global cultural heritage.

This does not mean that there are no unresolved issues at a number of levels. One methodological snag is that the sheer number of Italian translations of the Polish bestseller makes it difficult to find a native Italian corpus that would be large enough – and contain enough very prolific writers – to counterbalance the inescapable bias in the content of the most-frequent-word lists towards the lexicon of the most numerous subgroup. While one way to alleviate this – the piecemeal introduction of the *Quo vadis* group into the Italian corpus – has been used with some success here, this is certainly a limitation of what could be a population-based rather than a sample-based study.

5. ACKNOWLEDGEMENTS

This research was made as part of the project: "Miejsce Quo vadis? w kulturze włoskiej. Przekłady, adaptacje, kultura popularna" (0136/NPRH4/H2b/83/2016), funded by Poland's National Program for Advances in the Humanities (NPRH).

6. REFERENCES

- [1] Bastian, M., Heymann, S., and Jacomy, M. 2009. "Gephi: an open source software for exploring and manipulating networks." *Proceedings of the International AAAI Conference on Weblogs and Social Media*, San Jose, Ca.
- [2] Berti, G. de, and Galletti, E. 2016. "La fortuna di *Quo vadis* in Italia nel primo quarto del Novecento." In Woźniak, M., Biernacka-Licznar K., *Quo Vadis. Da caso letterario a fenomeno di massa. Ispirazioni - adattamenti - contesti*. Ponte Sisto, Roma, 50-59.
- [3] Burrows, J.F. 2002. "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17: 267-287.
- [4] Eder, M. 2017. "Visualization in stylometry: Cluster analysis using networks." *Digital Scholarship in the Humanities* 32(1): 50-64.
- [5] Eder, M., Kestemont, M., and Rybicki, J. 2016. "Stylometry with R: A package for computational text analysis." *The R Journal* 8(1): 107-121.
- [6] Even-Zohar, I. 1990. "The Position of Translated Literature within the Literary Polysystem." In *Polysystem Studies [= Poetics Today]* 11(1): 45-51.
- [7] Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., and Vitt, T. 2017. *Digital Scholarship in the Humanities* 32 (sup. 2): 4-16.
- [8] Giergielewicz, M. 1990. *Henryk Sienkiewicz. A Biography*. Hippocrene Books, New York.
- [9] Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. 2008. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9(6): e98679. DOI=10.1371/journal.pone.0098679.
- [10] Marinelli, L. "'Quo vadis.' Traducibilità e tradimento," *Europa Orientalis* 3: 131-146.
- [11] R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL=<https://www.R-project.org/>.
- [12] Rybicki, J. 2014. "Pierwszy rzut oka na stylometryczną mapę literatury polskiej," *Teksty drugie* 2: 106-128.
- [13] Rybicki, J. 2016. "Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies," *Digital Scholarship in the Humanities* 31(4): 746-761.
- [14] Smith, P. and Aldridge, W. 2011. "Improving authorship attribution: Optimizing Burrows' Delta method." *Journal of Quantitative Linguistics*, 18(1): 63-88.
- [15] Woźniak, M. 2016. "Quo vadis: da caso letterario a fenomeno di massa. Dove ci ha portato Sienkiewicz?" In Woźniak, M., Biernacka-Licznar K., *Quo Vadis. Da caso letterario a fenomeno di massa. Ispirazioni - adattamenti - contesti*. Ponte Sisto, Roma, 6-15

L'analisi delle reti come strumento a supporto della ricerca storica. Da Ceprano a Benevento (1266)

Alessandro De Troia¹, Vito Ricci²

¹ Independent Researcher, Master Degree in Computer Science, Università di Bologna - alessandrodetroia@gmail.com

² Università di Bari, Unità operativa Statistiche di Ateneo - vito.ricci@uniba.it

ABSTRACT

Applicare le metodologie dell'analisi delle reti sociali all'ambito della ricerca storica può fornire una misura quantitativa quanto qualitativa a supporto degli storici. Nella fattispecie si vuole presentare uno studio in cui si è analizzata una vasta gamma di documenti relativi agli anni precedenti la famosa Battaglia di Benevento avvenuta nel febbraio del 1266 in cui Manfredi, Re di Sicilia, perse la vita a causa di un presunto tradimento dei baroni che dovevano supportarlo in battaglia. Il grafo delle relazioni creato a partire dalla vastissima documentazione online (es. Regesta imperii) e non (es. Registri Angioini) ha permesso di ricostruire una fitta rete di comunicazioni tra le figure principali del periodo: il Papa Clemente IV (documenti dal 1265 al 1266), Manfredi (documenti dal 1257 al 1266) e Carlo I d'Angiò (1265 al 1267). Infine la lunga serie di personalità citate nella documentazione ha consentito, grazie all'uso del software Ucinet, una puntuale applicazione di alcune euristiche proprie dell'analisi di rete permettendo dapprima un'analisi su diversi piani semantici e infine un bilancio sul grafo complessivo dei livelli di Degree e Betweenness su quei nodi che storicamente sono individuati come i personaggi chiave della disfatta di Manfredi. I risultati mostrano come ad alti livelli di Betweenness e Degree il soggetto è tra quelli indicati come traditore dagli storici, mentre coloro che hanno un basso grado di Betweenness nonostante Degree alta, pare siano coloro che gli storici identificano come fedelissimi al Re e che effettivamente morirono nella battaglia o furono imprigionati e uccisi qualche mese dopo. Uno sviluppo futuro potrebbe prevedere misure predittive sul grafo o l'aggiunta della rete dei matrimoni tra le varie famiglie coinvolte nell'analisi.

PAROLE CHIAVE

social network analysis, military history, ucinet

1. INTRODUZIONE

Da diversi anni l'analisi di rete è diventata uno strumento a supporto degli storici per l'individuazione di strutture e comportamenti complessi. A prescindere dall'approccio strutturalista o connessionista modellare in un grafo una serie di informazioni permette allo studioso di avere a disposizione una serie di euristiche [1] [2] che permettono analisi quantitative e qualitative del modello. Questa metodologia ha iniziato a prendere piede anche nel campo delle *Digital Humanities* in riferimento alla Storia con diversi studi, basti pensare a Padgett [3] [6], Tieken e van Ostade [4], Mahoney e Rueschemeyer [5], McLean [6] e i più recenti Engl [7] e Fernández-Aceves [8]. L'analisi proposta in questa sede esamina e modella un grafo citazionale a partire dalla documentazione (anche online) di un range di anni ben definito¹ adoperando alcune misure (*Degree Centrality* e *Betweenness Centrality*) al fine di valutare la centralità di alcuni nodi.

2. IL CONTESTO STORICO

Dopo la morte dell'Imperatore Federico II Hohenstaufen² avvenuta nel dicembre del 1250, il Regno di Sicilia vede contrapporsi il Papato e la dinastia sveva. La contesa troverà in Manfredi³, Re di Sicilia dal 1258, il campione ghibellino, da una parte e i papi, Urbano IV prima e Clemente IV poi, dall'altra. Lo scontro scaturirà nella discesa del fratello del Re di Francia Carlo I d'Angiò che fu convinto dai papi ad intraprendere l'impresa della conquista del *Regnum Siciliae* che avvenne dopo una lunga campagna il cui epilogo vide la totale sconfitta di Manfredi a Benevento nel febbraio del 1266. Le cronache [11] [12] [13] e persino Dante nel XXVIII canto dell'*Inferno* (vv.16-18)⁴ riferiscono del tradimento di alcuni baroni del Regno i quali avrebbero abbandonato proprio sul campo di battaglia Manfredi. Il giovane Re decise comunque di buttarsi nella mischia ma fu inevitabilmente ucciso. Lo storico Pasquale Cafaro nel 1952 [14] analizzò la situazione politica del periodo individuando nella perdita di consensi del sovrano svevo la causa dell'abbandono di una parte della nobiltà regnicola.

3. METODOLOGIA

La documentazione raccolta fa riferimento sostanzialmente a tre fonti principali:

- I. Regesta Imperii Online⁵
- II. Alcuni fatti riguardanti Carlo I di Angiò dal 6 di agosto 1252 al 30 di dicembre 1270 [15]

¹ Il primo documento è del 1245 mentre l'ultimo è del 1268.

² Si rimanda al recente studio dello Stürner [9] per approfondimenti in merito.

³ Lo studio del Pispisa [10] è di fondamentale importanza per la figura del Re di Sicilia.

⁴ "...a Ceperan, là dove fu bugiardo ciascun Pugliese".

⁵ <http://www.regesta-imperii.de/regesten/suche.html> - Per la parte relativa alla ricerca dei regesti. Consultata il 5 ottobre 2017.

III. Codice Diplomatico del regno di Carlo I e II D'Angiò [16]

Le motivazioni che hanno spinto alla scelta di II e III sono puramente legate alla lingua italiana che ha facilitato la creazione del legame semantico tra i soggetti coinvolti nella comunicazione⁶. Per I invece il lavoro è stato più arduo data la mancanza di annotazioni in lingua inglese o italiana ma d'altro canto ha permesso di automatizzare parte del processo interrogando la base di dati con l'ausilio di algoritmi. Nella Tabella 1 un riassunto sui dati raccolti in riferimento ai personaggi più importanti. Si noti come l'esigua quantità di documenti relativi a Manfredi è in parte dovuta alla scarsità di testimonianze nelle tre fonti analizzate. La quantità di record e gli intervalli temporali in cui questi sono distribuiti permettono di ottenere un'analisi statisticamente rilevante del periodo in oggetto.

Soggetto Produttore	Elezione - Morte	Documenti	%	Primo Documento	Ultimo Documento
Clemente IV	1265-1268	198	47,03	1265	1266
Manfredi	1258-1266	84	19,95	1257	1266
Carlo I D'Angiò	1265-1285	139	33,02	1265	1267
		421	100,00		

Tabella 1 – Schema riassuntivo di parte della documentazione raccolta

Ciascun documento, partendo dalle definizioni del Pratesi [17], è stato suddiviso in:

- a) Cronaca
- b) Mandato (ordine, comunicazione)
- c) Concessione (Privilegio concesso, Revoca, Scomunica, Revoca di Scomunica)
- d) Lettera (rapporto epistolare)

I soggetti presenti in ciascun documento sono stati classificati come:

- Soggetto Principale
- Destinatario Generico
- Destinatario Positivo
- Destinatario Negativo
- Testimone
- Citazione Generica
- Citazione Positiva
- Citazione Negativa

Un legame quindi si viene a formare nel momento in cui un soggetto principale (indicato nella fonte) effettua una comunicazione o ha una relazione con un altro soggetto magari citandone altri. Nelle Tabelle 2 e 3 alcuni esempi di creazione di legami.

Data: 14 Marzo 1267	ID 51
Re Carlo scrive al castellano del castello di Nocera, il milite Radulfo de Faiello, al quale affida anche la custodia di Elena vedova di re Manfredi, nelle carceri di quel castello rinchiusa.	
CarloIDAngio	SoggettoPrincipale
RadulfoDeFaiello	Destinatario
ElenaDucas	CitazioneNegativa

Tabella 2 – Esempio di creazione di un legame

⁶ La scelta poteva coinvolgere anche i volumi dei Registri Angioini editi dall'Accademia Pontaniana.

Data: 23 Settembre 1265	ID 9
Carlo I d'Angiò scrive a Radulfo de Zandino castellano di Arsoli ordinandogli di consegnare quella fortezza a Guglielmo Carpentiero da lui eletto a castellano.	
CarloIDAngio	SoggettoPrincipale
RadulfoDeZandino	Destinatario
GuglielmoCarpentiero	DestinatarioPositivo

Tabella 3 – Esempio di creazione di un legame

Per i livelli semantici *Destinatario* e *Citazione* è stata effettuata un'analisi di correlazione statistica che in questa sede non verrà illustrata perché poco rilevante ai fini del presente studio⁷.



Figura 1 – Analisi delle correlazioni statistiche tra i livelli di relazione e semantici individuati

La matrice ottenuta invece dalla trasformazione *multiplex* dei vari piani semantici è rappresentata dal grafo vincolato⁸ con legami binari nella Figura 2.

La raccolta è di circa 450 testimonianze documentarie e di 550 soggetti censiti. In letteratura⁹ alcune misure di centralità vengono riferite come indicatore del “potere” che ciascun nodo esercita nel grafo. Il *grado* come misura di connessione: maggiore è il grado, maggiore è il potere dell’attore, in quanto può disporre di maggiore libertà nella scelta dei legami o in maniera prospetticamente diversa, è meno dipendente dagli altri. Questa misura inoltre può essere calcolata sia per reti a legami orientati (*in degree* e *out degree*) sia per reti a legami vincolati (*degree centrality*). La *betweenness centrality* invece come misura del ruolo di connettore di altri nodi, assumendo una funzione di *broker*: più è alta questa misura e maggiore sarà il potere posseduto dal nodo; in sostanza, il nodo mediatore avrà comunque parte nella relazione tra due nodi in quanto non superabile. Mediante il software UciNet il calcolo di entrambe le misure ha permesso di valutare il potere che tutti i soggetti coinvolti potevano esprimere nella fitta rete di comunicazione tra i tre principali soggetti analizzati.

⁷ Si faccia riferimento alla Figura 1 per una panoramica.

⁸ Un grafo a matrice simmetrica.

⁹ Si veda per esemplificazione [1] al cap. 5 e [2] al cap. 3.

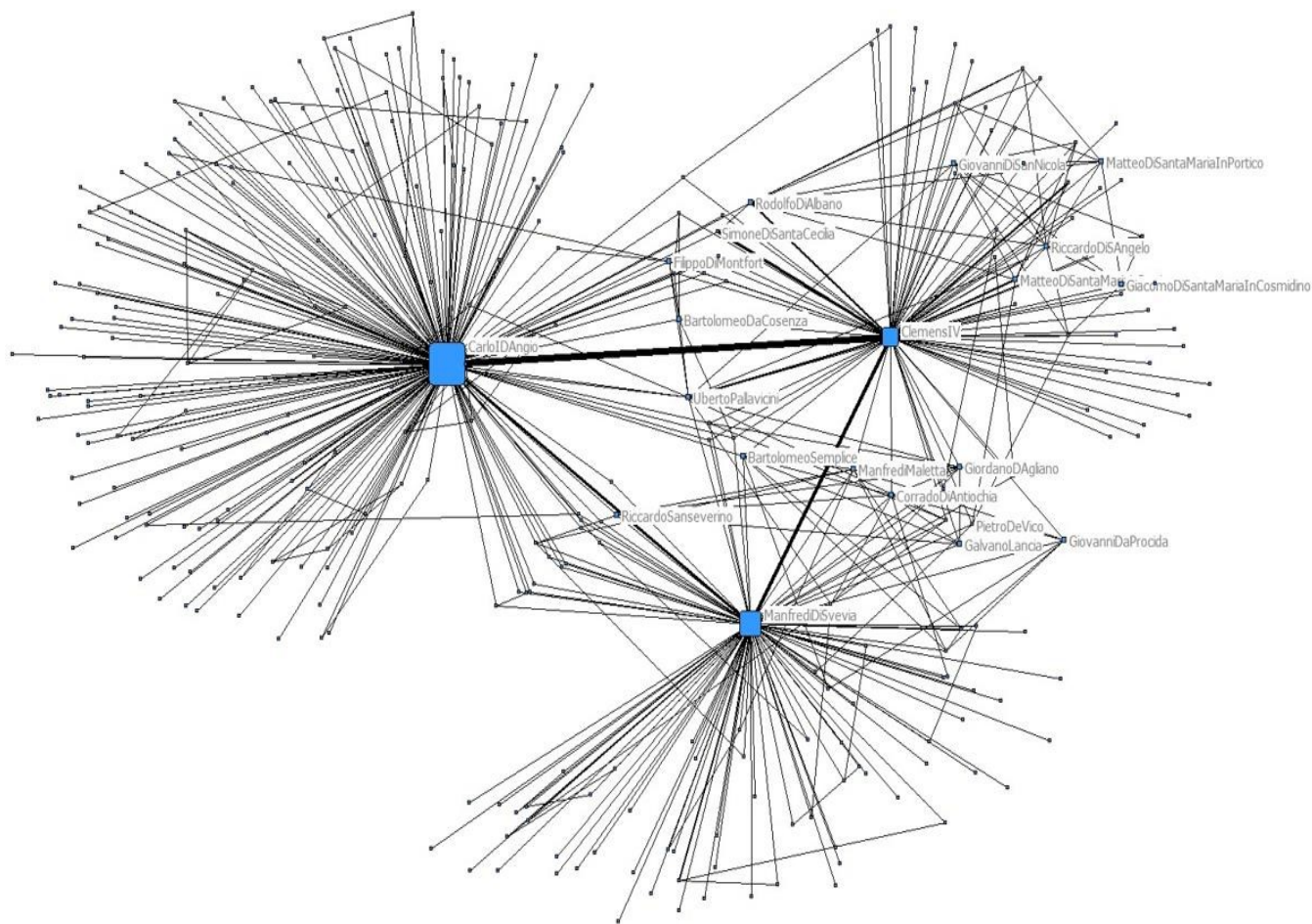


Figura 2 – Grafo risultante dall’analisi delle relazioni. In blu i 3 soggetti principali: Carlo I d’Angiò, Manfredi di Svevia e Clemente IV

4. RISULTATI

Nella Tabella 4 sono stati inseriti i soggetti individuati da Cafaro come vicini a Manfredi che da una parte pare abbiano effettivamente tradito come dimostrato da documenti successivi al periodo preso in esame¹⁰, alcuni soggetti di cui gli storici non riescono a dare un giudizio definitivo¹¹ e i restanti rimasti fedeli alla causa dello svevo¹². Generalmente ad un alto livello di *Degree* e *Betweenness* pare corrispondere una figura che tradì il sovrano. Più in generale come mostrato nella Figura 3 ad alti livelli di *Betweenness* è associata una figura il cui giudizio degli storici è negativo o sospeso. Una zona in cui la *Betweenness* è piuttosto bassa mentre il *Degree* è medio alto denota un cluster di soggetti che vengono identificati come fedeli, mostrando come sia sostanzialmente la prima misura ad indicare un potere più forte rispetto alla seconda.

<i>Personaggio</i>	<i>Giudizio degli Storici</i>	<i>Degree</i>	<i>Betweenness</i>
UbertoPallavicini	«In Lombardia, messi alla difesa dell’Oglio, aprirono il varco al nemico»	33	677,41
ManfrediMaletta	Consegnò il tesoro del Re all’angioino	17	298,94
GalvanoLancia	Restò fedele alla causa ghibellina e sostenne l’imperatore Corradino	14	8,36
GiordanoDAgliano	Ferito e fatto prigioniero, morì di fame in carcere	14	24,31
GiovanniDaProcida	Dopo la battaglia scappò dal Regno. Non abbandonò mai la causa ghibellina	14	46,53
CorradoDiAntiochia	Si sottomise a Carlo d’Angiò il quale lo imprigionò fino al 1267	12	50,57
PietroDeVico	«Si sottomise al Pontefice, e fece pace con Clemente e la stessa Pisa»	10	101,56
TommasoIIDAquino	Si recherà a Roma, ricevuto da Clemente IV, ottiene da re Carlo il permesso di andare nuovamente a Roma. Resterà fedele all’Angiò	8	78,44
RiccardoSanseverino	«Morì di crepacuore qualche mese dopo la battaglia»	7	117,50

Tabella 4 – Schema riassuntivo dei personaggi

¹⁰ Manfredi Maletta e Pietro De Vico.

¹¹ Uberto Pallavicini, Riccardo Sanseverino e Tommaso II D’Aquino.

¹² Galvano Lancia, Giordano d’Agliano, Giovanni da Procida, Corrado di Antiochia.

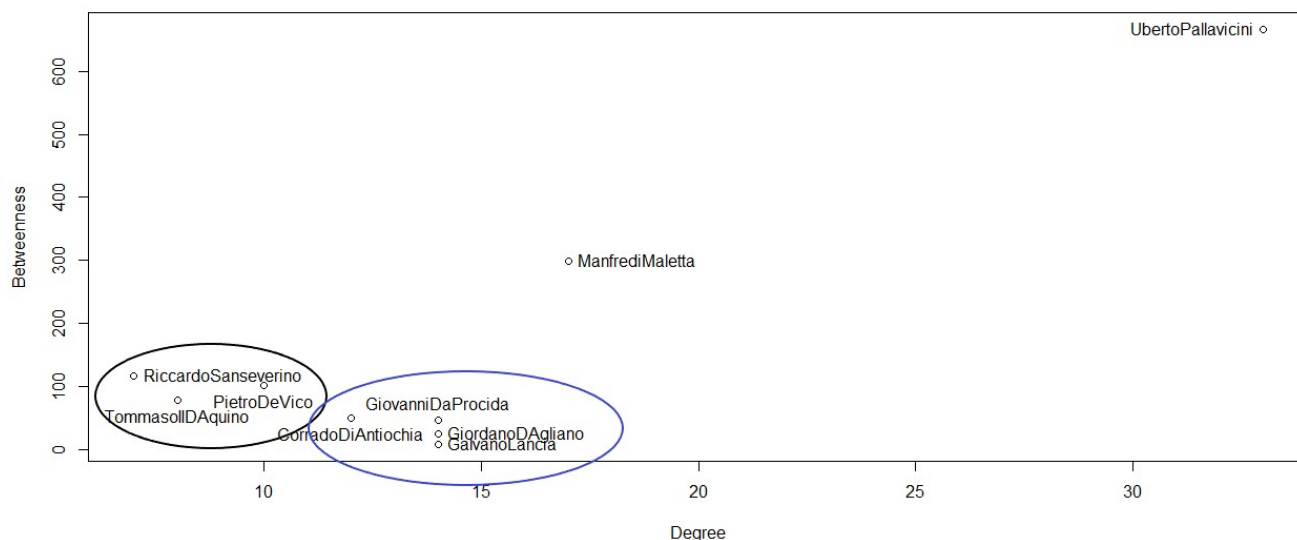


Figura 3 – Grafico di posizionamento dei soggetti principali analizzati dagli storici rispetto a Degree e Betweenness

5. CONCLUSIONI

Il processo di raccolta dei dati diventa fondamentale quando in fase di modellazione si ha la necessità di definire un piano semantico sulle relazioni. La maggior parte del tempo è stata impiegata proprio nella costruzione del dataset e nella raccolta dei documenti in parte automatizzata per la parte dei Regesta Imperii e in parte completamente a mano per le restanti fonti. Il dataset ha permesso comunque di effettuare alcune misure strutturali e di centralità ottenendo dei risultati soddisfacenti sotto il profilo qualitativo. Ulteriori misure sia strutturali che connessioniste potrebbero essere applicate su questo dataset anche per effettuare misure relative ai gruppi e predittive di un eventuale momento di rottura delle relazioni. Infine, a partire dai Registri angioini¹³, si potrebbe anche inserire un ulteriore livello con le genealogie e i matrimoni intercorsi nel periodo preso in esame.

6. RINGRAZIAMENTI

Si vuole ringraziare il prof. Marco Ruffino, docente dell'Università di Bologna, per i preziosi consigli durante i corsi di *Teoria delle decisioni nei processi cognitivi* e *Analisi delle reti sociali*.

7. BIBLIOGRAFIA

- [1] Scott, J. 1997. *L'analisi delle reti sociali*. A Cura di Amatore E., Carocci, ISBN 9788843023851.
- [2] Salvini, A. 2017. *L'analisi delle reti sociali. Risorse e meccanismi*. Pisa University Press, Pisa, ISBN 9788867417674.
- [3] Padgett, J. F. 1994. *Marriage and elite structure in Renaissance Florence; 1282-1500*. In: Paper delivered to the Social Science History Association.
- [4] Tieken, I., van Ostade, B. 2000. *Social Network Analysis and the History of English*. In: European Journal of English Studies, Vol. 4, n. 3, pp. 211-216, ISSN 1382-5577.
- [5] Mahoney, J., Rueschemeyer, D. 2003. *Comparative Historical Analysis in the Social Science*. Cambridge University Press, ISBN 9780511803963.
- [6] Padgett, J. F., McLean, P. D. 2006. *Organizational Invention and Elite Transformation: The Birth of Partnership Systems in Renaissance Florence*. In: American Journal of Sociology, Vol. 111, n. 5 (March), pp. 1463-1568, ISSN 0002-9602.
- [7] Engl, R. 2014. *Das Ende muslimischen Lebens im mittelalterlichen Südtalien. Netzwerkanalytische Überlegungen zu einer hundertjährigen Forschungsfrage*. In: Gesellschaftliche Umbrüche und religiöse Netzwerke. Analysen von der Antike bis zur Gegenwart, Transcript, Bielefeld, pp. 119-154, ISBN 9783837625950.
- [8] Fernández-Aceves H. 2017. Social network analysis and narrative structures: measuring communication and influence in a Medieval source for the Kingdom of Sicily. In: *Intersticios Sociales*, Volume 7, Issue 14, pp. 122-154, ISSN 2007-4964.
- [9] Stürmer, W. 2009. *Federico II e l'apogeo dell'impero*. Salerno Editore, Roma, ISBN 9788884026798.
- [10] Pispisa, E. 1991. *Il regno di Manfredi: proposte di interpretazione*. Sicania, Messina, ISBN 9788872680315.
- [11] Delle Donne, F., a cura di. 2014. *Andreas Ungarus. Descriptio victoriae Beneventi*. Istituto Storico Italiano per il Medioevo, Roma, ISBN 9788898079155.
- [12] De Rosa, F., a cura di. 2014. *Saba Malaspina. Storia delle cose di Sicilia: (1250-1285)*. Ciolfi, Cassino, ISBN 8886810660.
- [13] Aquilecchia, G. a cura di. 1997. *Giovanni Villani. Cronica con le continuazioni di Matteo e Filippo*. Einaudi, Torino, ISBN 9788806150327.
- [14] Cafaro, P. 1952. *Se i pugliesi furono bugiardi a Ceprano*. In: Archivio Storico Pugliese, Anno V – Fasc. I-IV pp. 243-250, ISSN 0392-0054.
- [15] Minieri Riccio, C. 1874. *Alcuni fatti riguardanti Carlo I di Angiò dal 6 di agosto 1252 al 30 di dicembre 1270, tratti dall'archivio Angioino di Napoli*. Napoli.
- [16] Del Giudice, G. 1863. *Codice Diplomatico del regno di Carlo I e II D'Angiò*. 3 vol., Napoli.
- [17] Pratesi, A. 1999. *Genesi e forme del documento medievale*, Jouvence, Milano, ISBN 9788878012608.

¹³ Ad oggi sono stati pubblicati circa 60 volumi del perduto Archivio Napoletano di cui si auspica presto una digitalizzazione.

Le lettere di Bellini: dalla carta al Web

Angelo Mario Del Grosso¹, Salvatore Cristofaro², Maria Rosa De Luca³,
Emiliano Giovannetti¹, Simone Marchi¹, Graziella Seminara³, Daria Spampinato²

¹ Istituto di Linguistica Computazionale - CNR, Italia - {angelo.delgrosso, emiliano.giovannetti, simone.marchi}@ilc.cnr.it

² Istituto di Scienze e Tecnologie della Cognizione - CNR, Italia - {salvatore.cristofaro, daria.spampinato}@istc.cnr.it

³ Università di Catania, Italia - {mdeluca, g.seminara}@unict.it

ABSTRACT

Nel contesto del progetto “Museo Virtuale della Musica BellinInRete” sarà reso fruibile, attraverso un processo di acquisizione, codifica e pubblicazione digitale, un corpus di lettere di Vincenzo Bellini, compositore catanese del XIX secolo. L’edizione digitale delle lettere belliniane sarà consultabile in rete e, inoltre, sarà integrata in un percorso museale interattivo in allestimento presso il Museo Civico Belliniano di Catania. Inoltre, il progetto prevede l’analisi e l’organizzazione del patrimonio belliniano del Museo, secondo i più recenti e accreditati standard e raccomandazioni per la metadattazione e fruizione dei dati di musei, biblioteche e archivi. Il patrimonio in esame viene reso fruibile attraverso un museo digitale multicanale che racconta i contenuti dei documenti.

PAROLE CHIAVE

Digital Edition, Vincenzo Bellini, Virtual Museum, Music, Computational Philology, Semantic Web, Software Design

1. INTRODUZIONE

Nella sua più ampia articolazione, il progetto si prefigge l’obiettivo di rinnovare e creare un cambiamento duraturo nella fruizione e nella valorizzazione del Museo Civico Belliniano attraverso modalità digitali e sviluppando, al contempo, un nuovo punto di vista comunicativo del patrimonio materiale e immateriale. Questo lavoro prevede, innanzitutto, l’analisi e l’organizzazione del patrimonio belliniano del Museo secondo i più recenti e accreditati standard e raccomandazioni per la metadattazione e fruizione dei dati di musei, biblioteche e archivi. Il patrimonio in esame viene reso fruibile attraverso un museo digitale multicanale che racconta i contenuti dei documenti. Inoltre, è previsto il riallestimento scenografico del museo, evocando i vari ambienti dei teatri dell’epoca, in cui il visitatore viene immerso con l’aiuto della narrazione della “vita in quattro atti” del Maestro, affidata a tecnologie audiovisive strettamente integrate con l’allestimento scenografico. BellinInRete è un progetto altamente multidisciplinare che si avvale delle competenze musicologiche presenti presso l’Università di Catania e dell’esperienza nel settore delle Digital Humanities dell’Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR di Catania e dell’Istituto di Linguistica Computazionale “A. Zampolli” (ILC) del CNR di Pisa.

Ad oggi è stato definito lo schema di codifica del carteggio ed è stato realizzato un prototipo di fruizione dell’edizione digitale delle lettere autografe di Bellini. L’edizione digitale del corpus epistolare è realizzata attraverso una codifica TEI-XML, con particolare attenzione alla gestione dei metadati relativi alla corrispondenza codificati mediante l’utilizzo del modulo *correspDesc*. Accanto alla codifica dei dati testuali sono state analizzate le tecniche standard per la descrizione catalografica e bibliografica del carteggio, come, ad esempio, l’aderenza allo schema UNIMARC per la registrazione degli elementi bibliografici. In tale scenario le buone prassi del Semantic Web (quali il collegamento di informazioni presenti nei documenti a dataset pubblici autorevoli, come VIAF¹, Geonames², CIDOC-CRM³ e RISM⁴) sono state giudicate le più adatte a rispondere ai requisiti ermeneutici del progetto. L’edizione digitale accosterà la riproduzione facsimilare di ciascuna lettera autografa con la sua trascrizione digitale, realizzata con criteri conservativi nel rispetto della forma del dettato della missiva originale.

2. STATO DELL'ARTE

L’approccio seguito è simile a quello adottato con successo in altre iniziative, dove i corpora epistolari sono stati codificati con le linee guida della TEI [5]. Tra l’altro, uno Special Interest Group di TEI (Correspondence SIG), dal 2008 al 2014, ha lavorato proprio allo sviluppo del tagset *correspDesc* per la codifica della corrispondenza [10]. Il web service CorrespSearch⁵ fa uso di questa risorsa e consente l’indicizzazione di lettere per mittente, destinatario, data, luogo di invio e ricezione attraverso il protocollo Correspondence Metadata Interchange Format (CMI) [7].

Tra i progetti più vicini agli scopi di BellinInRete citiamo innanzitutto WeGa⁶, un progetto tedesco che coinvolge l’edizione digitale di opere e scritti del musicista Carl Maria von Weber. Nel progetto Digital Archive of Letters in Flanders⁷ (DALF),

¹Virtual International Authority File, <https://viaf.org/>

²<http://www.geonames.org/>

³<http://www.cidoc-crm.org/>

⁴Répertoire International des Sources Musicales, <http://www.rism.info/>

⁵<http://correspsearch.net/index.xql?id=about&l=en>

⁶<http://weber-gesamtausgabe.de/de/Index>

⁷<http://ctb.kantl.be/project/dalf/index.htm>

sono state codificate 1500 lettere estratte dalla rivista di letteratura fiamminga “Van Nu en Straks”. Molti tra gli elementi di marcatura prodotti in seno al progetto DALF sono stati riutilizzati per l’edizione digitale del progetto delle lettere di Van Gogh⁸. “In Mozart’s Words”⁹ è un progetto di raccolta, conservazione e valorizzazione della corrispondenza del musicista Wolfgang Amadeus Mozart; il progetto, grazie ad una accurata codifica digitale del contenuto delle lettere, consente di scoprire ed approfondire via Web le esperienze più intime della vita del compositore austriaco.

L’ILC ha partecipato al progetto “Clavius on the Web” il cui scopo principale è stato quello di digitalizzare, trascrivere, codificare e annotare a vari livelli alcune lettere indirizzate al matematico gesuita Cristoforo Clavio¹⁰ [1]. Seppur non incentrato sull’edizione digitale di lettere, si reputa importante citare il progetto Proust’s drafts prototype¹¹, nel quale sono proposte interessanti soluzioni per la codifica di edizioni genetiche. Infine, per gli aspetti di catalogo e linking di contenuti nel Semantic Web si sottolinea il progetto “RISM” ritenuto un catalogo di autorità del dominio musicale e che ha approntato un archivio Linked Open Data (LOD) di una vasta collezione di fonti musicali accessibile tramite uno SPARQL endpoint dedicato¹².

3. OBIETTIVI E METODOLOGIA

Sebbene l’obiettivo principale del progetto sia quello di produrre, preservare e rendere fruibile l’edizione digitale del carteggio belliniano, ci sono alcuni elementi di innovazione che contraddistinguono BellinInRete. L’edizione digitale sarà consultabile, oltre che online, anche all’interno dell’allestimento del museo belliniano, consentendo al visitatore di poter “toccare con mano” i documenti originali per un’esperienza di fruizione maggiormente immersiva.

La sperimentazione del tagset *correspDesc* dell’header di TEI (cfr 3.1) e l’uso del paradigma dei LOD all’interno dell’attività di annotazione delle lettere (cfr 3.2) costituiscono due ulteriori elementi a carattere innovativo.

Infine, come ci si aspetta da un progetto di questo tipo, i testi (inediti in rete) che saranno pubblicati e condivisi, consentiranno di far conoscere Vincenzo Bellini anche a un pubblico di non addetti ai lavori.

Di concerto con gli esperti del dominio, nella prima parte del progetto ci si è concentrati sull’edizione digitale delle lettere di Bellini e sul lavoro di schedatura e di catalogazione. La produzione di tale edizione prevede, allo stato attuale, di lavorare in parallelo su due fronti tecnologici: da un lato, la codifica TEI dei fenomeni testuali e dei metadati e, dall’altro, la descrizione delle entità rilevanti per mezzo di formalismi e vocabolari definiti nell’ambito del Semantic Web [11].

3.1. CODIFICA DELLE LETTERE

L’edizione digitale delle missive di Vincenzo Bellini custodite presso il Museo Civico Belliniano di Catania, inserita nel contesto del progetto “Museo Virtuale della Musica BellinInRete”, costituisce un’iniziativa di grande rilevanza culturale e con un chiaro contributo innovativo in quanto al momento non esistono in Italia progetti analoghi relativi ai compositori d’opera del XIX secolo.

Le attività relative alla conservazione e valorizzazione delle lettere di Bellini consistono principalmente nella codifica e nella pubblicazione in rete di un corpus di 140 missive delle quali quaranta di mano belliniana e le rimanenti redatte dai corrispondenti del musicista. Durante la prima fase del lavoro si è proceduto all’analisi dei documenti per identificare le informazioni testuali e metatestuali oggetto di codifica. In particolare, l’edizione digitale delle lettere di Bellini prevede di riprendere ed estrarre le informazioni d’interesse già edite nello studio filologico condotto in [3].

A completamento dell’edizione cartacea, specifico interesse dell’edizione digitale è stato rivolto agli aspetti metatestuali relativi alla conformazione fisica delle lettere, come ad esempio le modalità di piegatura e le condizioni di conservazione, i timbri postali, la presenza o meno del sigillo in ceralacca. Inoltre, hanno costituito altri elementi di particolare rilevanza i destinatari delle lettere, i nomi citati, nonché gli eventi riferiti nelle missive.

La fase di analisi ha prodotto una schema di codifica¹³ che consente di registrare e di elaborare i dati relativi al corpus belliniano al fine di mettere a disposizione dell’utente funzionalità di navigazione avanzate in grado di descrivere informazioni di natura storico-critica, mostrando accanto al testo trascritto, anche l’apparato critico redatto opportunamente dall’editore.

Lo schema XML definito prevede, tra l’altro, l’adozione di due importanti estensioni dell’instestazione TEI (*Header*): l’elemento *correspDesc*¹⁴ e l’elemento *msDesc*¹⁵, il primo adatto alla rappresentazione della corrispondenza, presente nel descrittore *profileDesc*, e il secondo utile alla descrizione fisica delle lettere, presente nel modulo TEI “*manuscript description*”.

Oltre alla tradizionale codifica delle gerarchie strutturali del documento e delle gerarchie logiche del testo, lo schema consente di rappresentare anche informazioni relative al contenuto delle lettere (attraverso gli elementi *front*, *body* e *back* di *text*) e alla struttura materiale del carteggio (tramite l’elemento *facsimile*). Inoltre, lo schema di codifica consente di

⁸<http://vangoghletters.org/>

⁹<http://letters.mozartways.com/>

¹⁰<http://claviusontheweb.it/>

¹¹http://research.cch.kcl.ac.uk/proust_prototype/

¹²<https://opac.rism.info/index.php?id=8&L=1#c116/>

¹³Una prima versione della codifica è disponibile al seguente link, <http://bit.ly/2fY6I84/>

¹⁴<http://www.tei-c.org/release/doc/tei-p5-doc/it/html/ref-correspDesc.html>

¹⁵<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-msDesc.html>

registrare a diversi livelli di granularità collegamenti testo-immagine mediante tecnologie utili all'identificazione univoca e globale delle regioni d'interesse. A tale scopo sono stati scelti due noti formalismi e protocolli: il framework del Canonical Text Service (CITE-CTS) [4] per gli aspetti citazionali del testo e il protocollo dell'International Image Interoperability Framework (IIIF) [12] per gli aspetti di recupero delle immagini.

Attraverso le diverse modalità di accesso e di ricerca, sarà possibile navigare il carteggio in modo puntuale ed esaustivo, collegandone le informazioni più rilevanti (per esempio le entità nominate, i cataloghi, la bibliografia) a dataset di interesse presenti in rete. Nello specifico, sono state preferite iniziative di descrizione delle entità di dominio come RISM a dataset più *general purpose* come quello di DBpedia¹⁶.

3.2. CATALOGAZIONE

La Direzione del Museo ha reso disponibili un inventario dei manoscritti (redatto negli anni 2000), alcuni elenchi di materiali documentari presenti nella biblioteca e degli oggetti custoditi nel Museo. Gli elenchi sono redatti senza una specifica struttura di riferimento e senza aderire a particolari convenzioni. Tale patrimonio è solo in parte digitalizzato, non è catalogato, né è presente in cataloghi nazionali e internazionali (SBN, SAN, Europeana Sound, ecc.). Nello specifico sono state prese in considerazione le regole di catalogazione per il materiale musicale alla base della piattaforma SBN: titolo uniforme, manuale di catalogazione dei materiali musicali, descrizione dei manoscritti musicali.

Gli altri elementi del patrimonio belliniano saranno metadati secondo le schede di catalogo dell'Istituto per il Catalogo e la Documentazione (ICCD). In tal modo sarà possibile sfruttare i protocolli di condivisione messi a disposizione dall'Online Public Access Catalog (OPAC - Z39.50), utili per codificare opportunamente le informazioni di accesso e di indicizzazione del documento da catalogare.

In aggiunta alla pubblicazione dei metadati nei cataloghi nazionali, si prevede di creare un'ontologia dei contenuti del patrimonio belliniano, estraendo i concetti di particolare interesse presenti nelle lettere (nomi di persona, luoghi, ecc.) e negli altri documenti facendo riferimento a schemi di metadati standard, come RDA¹⁷, FRBRoo¹⁸ e FOAF¹⁹. Inoltre, nell'ottica del riuso e dell'interoperabilità delle risorse del Semantic Web, è stata condotta un'accurata analisi delle ontologie, vocabolari e thesauri esistenti, con particolare attenzione al dominio musicale; quindi sono stati presi come riferimento il sopracitato progetto RISM per il dominio musicale e le risorse VIAF, CIDOC-CRM e GeoNames per gli altri concetti.

3.3. PROTOTIPO DI VISUALIZZAZIONE

Uno degli aspetti più innovativi del progetto BellinInRete è rappresentato dal lavoro di codifica e di implementazione software per la realizzazione, l'indicizzazione e la visualizzazione dell'edizione digitale. A tale scopo è stato realizzato un prototipo (Figura 1) per l'elaborazione del testo basato sulle specifiche discusse precedentemente. Il prototipo²⁰, basato su tecnologie Web open source, verrà adattato per essere conforme ad EVT (Edition Visualization Technology) [9], un software, anch'esso open source, nato per la creazione e la navigazione Web di edizioni digitali di manoscritti. La gestione delle immagini è stata realizzata con OpenSeadragon²¹, una libreria conforme alle specifiche IIIF largamente utilizzata per la visualizzazione di immagini in alta definizione.

Per favorire la manutenzione sia dell'edizione digitale sia delle risorse di progetto sono state adottate tecnologie open source e formati standard in modo da ridurre drasticamente il rischio di obsolescenza e garantire l'accesso e la disponibilità a lungo termine dei dati prodotti. In ultima istanza la Direzione del Museo avrà l'onere di affrontare le problematiche relative ad una adeguata gestione dell'edizione digitale e delle tecnologie ad essa collegate.

4. CONCLUSIONI E LAVORI FUTURI

Il lavoro relativo alla edizione digitale delle lettere di Vincenzo Bellini, qui illustrato, è parte del progetto Museo Virtuale della Musica BellinInRete. Nell'ambito delle attività che interessano il corpus epistolare sono state discusse le scelte adottate per la codifica del testo, con particolare riguardo alla grammatica epistolare, alla codifica delle entità nominate e alla visualizzazione parallela di testo e immagine. Dal punto di vista catalogografico sono stati analizzati i vocabolari ad oggi più utili alla creazione di dataset autorevoli e facilmente integrabili in una prospettiva Linked Open Data.

Per quanto riguarda l'indicizzazione del testo, l'applicazione si avvarrà della piattaforma Omega²² [6], in corso di sviluppo presso l'Istituto di Linguistica Computazionale del CNR di Pisa. A sviluppo concluso, EVT e Omega andranno a completarsi in un ambiente integrato di *digital scholarly editing* funzionale alla consultazione delle lettere di Bellini.

¹⁶<http://wiki.dbpedia.org/>

¹⁷Resource Description and Access, <http://www.rda-rsc.org/>

¹⁸Functional Requirements for Bibliographic Records object oriented, <http://www.cidoc-crm.org/frbroo/home-0>

¹⁹Friend Of A Friend, <http://xmlns.com/foaf/spec/>

²⁰<http://licodemo.ilc.cnr.it/bellini/>

²¹<https://openseadragon.github.io/>

²²<https://github.com/literarycomputinglab/>

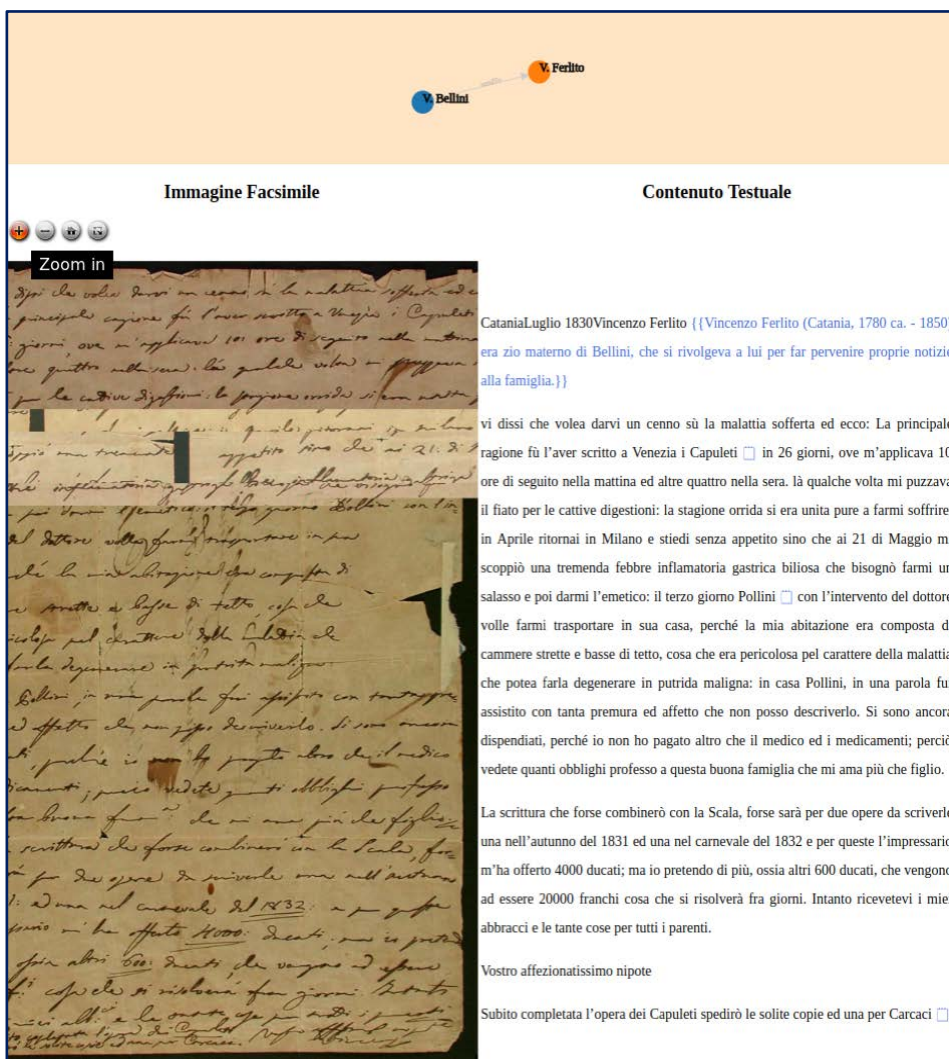


Figura 1 – Prototipo di elaborazione e visualizzazione testo-immagine del corpus epistolare.

Per quanto riguarda l’analisi del corpus saranno prese in considerazione anche le “minute”, che Bellini redigeva prima di scrivere le lettere “formali”, e che sono di grande interesse per gli interventi di correzione morfologica e stilistica praticati dal musicista.

L’edizione digitale, basata sulla codifica della grammatica epistolare (apertura, corpo, chiusura e imbustamento della lettera), verrà corredata da indici di persone, luoghi e termini del dominio musicale ripresi dal lessico LESMU [8] e ricodificati attraverso l’adozione del modello lessicale *lemon*²³ [2].

5. RINGRAZIAMENTI

Questa ricerca è parzialmente finanziata dal Patto per Catania a valere sul Fondo Sviluppo e Coesione 2014-2020: Piano per il Mezzogiorno.

6. BIBLIOGRAFIA

- [1] Abrate, Matteo, Angelo Mario Del Grosso, Emiliano Giovannetti, Angelica Lo Duca, Damiana Luzzi, Lorenzo Mancini, Andrea Marchetti, Irene Pedretti, and Silvia Piccini. 2014. “Sharing Cultural Heritage: The Clavius on the Web Project.” In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC), Reykjavik*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 627–634. European Language Resources Association (ELRA).
- [2] Bellandi, Andrea, Emiliano Giovannetti, Silvia Piccini, and Anja Weingart. 2017. “Developing LexO: A Collaborative Editor of Multilingual Lexica and Terminological Resources in the Humanities. 19 September 2017 Montpellier (France).” In *Proceedings of LOTKS In Conjunction with the 12th International Conference on Computational Semantics (IWCS)*.
- [3] Bellini, Vincenzo. 2017. *Carteggio*. Edited by Graziella Seminara. Leo S. Olschki.
- [4] Blackwell, Christopher, and Neel Smith. 2014. “The Homer Multitext: Technically Speaking ...” *The Homer Multitext*.
- [5] Burnard, Lou. 2015. “TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 2.9.1.” <http://www.tei-c.org/Guidelines/P5/index.xml>.

²³<http://lemon-model.net/>

- [6] Del Grosso, Angelo Mario, Emiliano Giovannetti, and Simone Marchi. 2017. "Il Modello a Microkernel Di Omega Nello Sviluppo Di Strumenti per Lo Studio Dei Testi: Dagli ADT Alle API." In *Associazione per l'Informatica Umanistica E per La Cultura Digitale*, edited by Fabio Ciotti. Roma: Sapienza Università Editrice.
- [7] Neuber, Frederike. 2016. "correspSearch." In *European Society for Textual Scholarship*.
- [8] Nicolodi, Fiamma, and Paolo Trovato. 2002. "La terminologia musicale italiana, il Lesmu e l'Europa. Con qualche appunto su 'barbaro' e 'gotico.'" *Musica e storia*, no. 1/2002: 51–70. doi:10.1420/7600.
- [9] Rosselli Del Turco, Roberto, and Chiara Di Pietro. 2016. "Between Innovation and Conservation: The Narrow Path of UI Design for the DSE." In *Digital Scholarly Editions as Interfaces*.
- [10] Stadler, Peter, Marcel Illetschko, and Sabine Seifert. 2016. "Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing correspDesc." *Journal of the Text Encoding Initiative*, no. 9.
- [11] Vavliakis, Konstantinos N, Georgios Th Karagiannis, and Pericles A Mitkas. 2012. "Semantic Web in Cultural Heritage after 2020." In *11th International Semantic Web Conference · ISWC 2012*. Boston, MA, USA.
- [12] Ying, William, and James Shulman. 2015. "'Bottled or Tap?' A Map for Integrating International Image Interoperability Framework (IIIF) into Shared Shelf and Artstor." *D-Lib Magazine* 21 (7/8).

Linked data ed edizioni scientifiche digitali. Esperimenti di trasformazione di un *Quaderno di appunti*

Marilena Daquino¹, Francesca Giovannetti², Francesca Tomasi³

¹ Università di Bologna, Italia - marilena.daquino2@unibo.it

² Università di Bologna, Italia - francesc.giovannetti5@studio.unibo.it

³ Università di Bologna, Italia - francesca.tomasi@unibo.it

1. INTRODUZIONE

Tra le varie lezioni che l'universo dei *Linked Open Data (LOD)* sta contribuendo ad impartire, la necessità di un approccio data-centrico alla rappresentazione del sapere è l'insegnamento principe. Il passaggio al Web 3.0, ovvero all'altrimenti detto Web semantico, ha determinato una ridefinizione del concetto di risorsa informativa, che costringe a pensare alla rete globale come ad una collezione di entità, identificate in modo univoco e interconnesse attraverso link tipizzati, e non più solo rete di documenti. Partendo da questo assunto, il presente contributo intende muovere in una triplice direzione: riflettere sul passaggio delle edizioni scientifiche digitali al modello Linked Open Data; descrivere le fasi del processo di transizione; analizzare l'edizione digitale del *Quaderno di appunti* di Paolo Bufalini quale esempio di questo processo e delle complessità insite nell'attività di trasformazione.

2. DALLE EDIZIONI SCIENTIFICHE DIGITALI AI LINKED OPEN DATA

Le edizioni scientifiche digitali tradizionalmente lavorano su fonti che individuano gli oggetti della ricerca nei documenti. Tali documenti assumono i connotati di risorse semi-strutturate, ottenute attraverso l'adozione di un modello di markup, tipicamente a base XML¹, che trova nello schema TEI un vocabolario condiviso.

Una collezione di testi digitali marcati in XML manca però della nozione di struttura che è insita, ad esempio, nei database, modello *naturaliter* di rappresentazione data-centrico. Il passaggio dal database al dataset LOD risulta più lineare, seppure non scevro di insidie, per la natura stessa dei dati e per la presenza di un modello concettuale di riferimento.

Il processo di transizione da una collezione di documenti XML/TEI ad un dataset LOD² è invece più complesso e merita una riflessione importante, in quanto significa tradurre il concetto di documento da entità non strutturata, o semi-strutturata, in dato strutturato. Tale assunto costringe ad abbracciare il principio di decostruzione del concetto di documento, adottando quello di dato: dalla dimensione di risorsa informativa a testo pieno, il documento diventa tanto un oggetto (soggetto), quanto una potenziale collezione di oggetti (soggetti). Questi soggetti sono caratterizzati da una serie di proprietà che consentono la creazione di relazioni con altri oggetti, siano essi altri documenti o entità reali (e.g. opere, persone, luoghi, eventi e periodi storici). Acquisendo questo processo come un modello di riferimento, le numerose edizioni digitali su Web, che adottano XML/TEI³, ad oggi standard *de facto* nella comunità filologica⁴, si vedono incentivate ad una riorganizzazione del proprio sapere in una forma tale da garantire l'ingresso nella LOD cloud delle (*semantic*) *scholarly editions*⁵, ovvero nel sistema di interconnessioni del *cultural heritage (texts)*.

3. IL WORKFLOW: GLI STEP DEL PROCESSO DI TRANSIZIONE

Il processo di transizione al modello LOD, prima di essere un'operazione tecnica, è innanzitutto attività concettuale. Fondamentale è l'analisi del modello di markup adottato nell'edizione, attività necessaria per riconoscere e qualificare le scelte di descrizione delle caratteristiche della fonte da parte dell'interprete. Potenzialmente ogni elemento del markup diventa un soggetto, un oggetto o un asserto relazionato all'entità documento. Pertanto lo studio del modello di markup è rilevante ai fini delle possibilità descrittive finali secondo il nuovo modello di rappresentazione.

Da ciò consegue la definizione delle *suite* delle relazioni che è possibile costruire. La scelta dei predicati ontologici deve sostare ad una consapevolezza: aumentare l'espressività della base di conoscenza fornita dal testo, senza però disperdere il potenziale informativo in nome della, pur necessaria, semplificazione. La scelta dei più opportuni modelli ontologici è dunque il momento centrale di questa fase. Infine, la costruzione del grafo della conoscenza è dirimente. I dati sono *linked* quando sono capaci di dialogare con quel sapere disseminato nel web, in grado di arricchire il testo digitale e capace di espandere l'esperienza informativa dell'utente finale oltre i confini del testo stesso.

Calare questo processo nel contesto delle edizioni scientifiche digitali, significa porsi una serie di interrogativi:

¹ Tipicamente, ma non sempre. Varie soluzioni alternative al markup embedded sono state proposte. Si veda una rassegna in [12] in particolare pp. 138-142.

² Si vuole qui proporre qualcosa di diverso dunque da soluzioni che hanno riflettuto sull'uso del markup stand-off, come [11], in particolare a base RDF, come in [13].

³ Corposo il numero di edizioni menzionate nella sezione *Projects Using the TEI* del sito ufficiale del progetto, <http://www.teic.org/Activities/Projects/>

⁴ Una rapida disamina può essere effettuata analizzando il catalogo delle edizioni digitali in [9]. Non sarà irrilevante notare come le Linee Guida di MLA [10] individuino nell'uso di XML/TEI un criterio per la valutazione della qualità delle edizioni digitali.

⁵ Per una prima valutazione dell'uso dei LOD nelle edizioni scientifiche digitali si veda il progetto "ASED - Annotazione Semantica per Edizioni Digitali", <https://ased.celi.it/>

1. il passaggio dal markup embedded al dataset. Ovvero: come trasformare XML/TEI in RDF? Non si tratta di un mero esercizio meccanico, ma di una presa di coscienza della funzione dei sistemi di marcatura embedded o inline. Sarà allora necessario distinguere le entità reali dalle loro occorrenze all'interno del testo (i.e., le stringhe di caratteri) e preservare i fenomeni registrati nel testo per la sua visualizzazione e fruizione (e.g., aspetti grafici, impaginazione).
2. il passaggio dagli elementi TEI all'ontologia (per cui cfr. [8]). Ovvero: quali modelli ontologici riusare? È necessario creare nuove classi e proprietà? Si deve invece 'ontologizzare' lo Schema (per cui cfr. [1], [2])? I modelli concettuali scelti per la *mapping* su TEI sono determinanti per stabilire la capacità comunicative della base di conoscenza e svelare concetti latenti.
3. la creazione dei collegamenti con i sistemi di controllo delle autorità. Ovvero: quali collegamenti istituire? Come dare consistenza alla rete del sapere? Alla base della scelta dei dataset relazionati, così come per la scelta dei modelli, è di nuovo la necessità di aprire i dati al dialogo con altre risorse in grado di arricchire la base di conoscenza così da soddisfare i bisogni informativi di varie tipologie di utenti, anche specializzati.

Questi interrogativi sollevano una riflessione sul senso del processo stesso. Una strategia di valorizzazione delle edizioni digitali quale patrimonio culturale che si avvale delle tecnologie del Semantic Web richiede:

- pianificazione: la definizione del modello di markup in previsione della trasformazione dei dati è la fase più dispendiosa in termini di tempo;
- *know how* interdisciplinare, per la ridefinizione del modello del testo e l'estrazione dei dati;
- maggior numero di risorse tecnologiche, per la pubblicazione e la manutenzione.

Ciò significa, in termini di costi, una maggiore domanda di *expertise*, intesa come consapevolezza e capacità di gestione dei vari momenti del processo – che una solida formazione in *Digital Humanities* dovrebbe assolvere. I costi legati all'infrastruttura tecnologica non variano significativamente: ci si avvale di software *open source* e la dimensione dei dati creati permane di medio-bassa scala – ben diverso è il ragionamento in caso di *Digital Libraries*. In ultima analisi, ai tempi e le risorse impiegate per la realizzazione di una edizione scientifica digitale “tradizionale” si aggiunge qui il costo della “destrutturazione dell'edizione” in vista della modellazione della conoscenza.

I benefici immediati sono evidenti quando le tematiche sollevate dal testo (e.g., la rete di personalità, testi, eventi o altri oggetti culturali veicolati dal testo pieno) trovano riscontro e ampliamento in fonti di dati esterne, che possono essere integrate, arricchire la base di conoscenza e ridurre i tempi di ricerca e approfondimento scientifico – a cui però si sostituiscono tempi e risorse impiegate per la riconciliazione dei dati. D'altro canto, la condivisione della conoscenza estratta dal testo stesso, consente una maggiore condivisione del proprio lavoro scientifico in ambiti diversi da quello accademico, quindi una maggiore possibilità di riutilizzo e aumento di impatto della cultura nella società dell'informazione. Il lavoro condotto sul *Quaderno di appunti* di Paolo Bufalini aiuterà ad affrontare la dimensione epistemica del processo qui illustrato.

4. UN ESEMPIO: IL QUADERNO DI APPUNTI DI PAOLO BUFALINI

Il lavoro condotto sul *Quaderno di appunti* manoscritti di Paolo Bufalini, membro del PCI e Senatore della Repubblica dal 1963 al 1992, latinista e raffinato traduttore di Orazio, offre un esempio di processo di transizione da documenti XML/TEI a LOD. Fra il 1981 e il 1991 Bufalini tenne un quaderno di appunti privati composto da 145 pagine rilegate e 2 carte sciolte. Il *Quaderno* contiene frammenti testuali relativi a momenti significativi della vita intellettuale e sociale dell'autore: citazioni tratte da testi letterari, note personali e commenti, narrazioni di eventi, prove di traduzione dal latino all'italiano. In particolare, data la natura privata del *Quaderno*, le relazioni fra i frammenti testuali (testo-testo, testo-nota, testo-traduzione e traduzione-traduzione) non sono rese esplicite da Bufalini, così come non sempre viene dichiarata la fonte dei testi citati.

Il *Quaderno* veicola dunque una complessità significativa in termini di concettualizzazione. Quali frammenti testuali sono presenti nel *Quaderno*? Le traduzioni sono inedite o si basano su edizioni esistenti? Quale edizione (se una ne aveva) Bufalini teneva sotto mano mentre scriveva un appunto? A quale frammento testuale si riferisce una particolare annotazione? Qual è la paternità di un frammento testuale non attribuito, sia esso testo primo o traduzione? Una serie di interrogativi importanti per rispondere ai quali è stato necessario uno studio del testo del manoscritto, finalizzato in prima battuta a garantire la qualificazione, la tipizzazione e l'identificazione dei frammenti.

Alla morte di Paolo Bufalini, il *Quaderno* fu donato dagli eredi al Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna, che si è occupato di trascrivere gli appunti e di identificare le fonti delle 196 citazioni presenti (cf. [3] e [4]). I testi, accompagnati da traduzioni e annotazioni, provengono dai classici della letteratura italiana e latina (Dante, Manzoni, Petrarca, Carducci, Virgilio, Orazio, Tacito, Cicerone, Ennio, Marziale, etc.) e delle letterature europee (Flaubert, Hegel, Shakespeare, Mann, Yourcenar, etc.).

I frammenti che compongono il *Quaderno* formano dunque un complesso network di relazioni che suggerisce la necessità di una lettura non lineare del testo rispetto al supporto documentario.

Il modello di codifica XML/TEI, elaborato per l'annotazione, ha consentito la marcatura, fra gli altri, degli elementi portatori di relazioni (in particolare testi citati, riferimenti bibliografici, persone, rispettivamente marcati tramite gli elementi `tei:cit`, `tei:bibl`, `tei:persName`), mentre nell'intestazione del documento XML/TEI, tali frammenti, identificati univocamente, sono stati collegati agli *authority file* di riferimento ed è stata fornita una prima descrizione della natura delle relazioni (tramite l'elemento `tei:relation`).

Ma la complessità del sistema di interconnessioni rende difficile il solo utilizzo di XML/TEI per esprimere compiutamente le relazioni⁶.

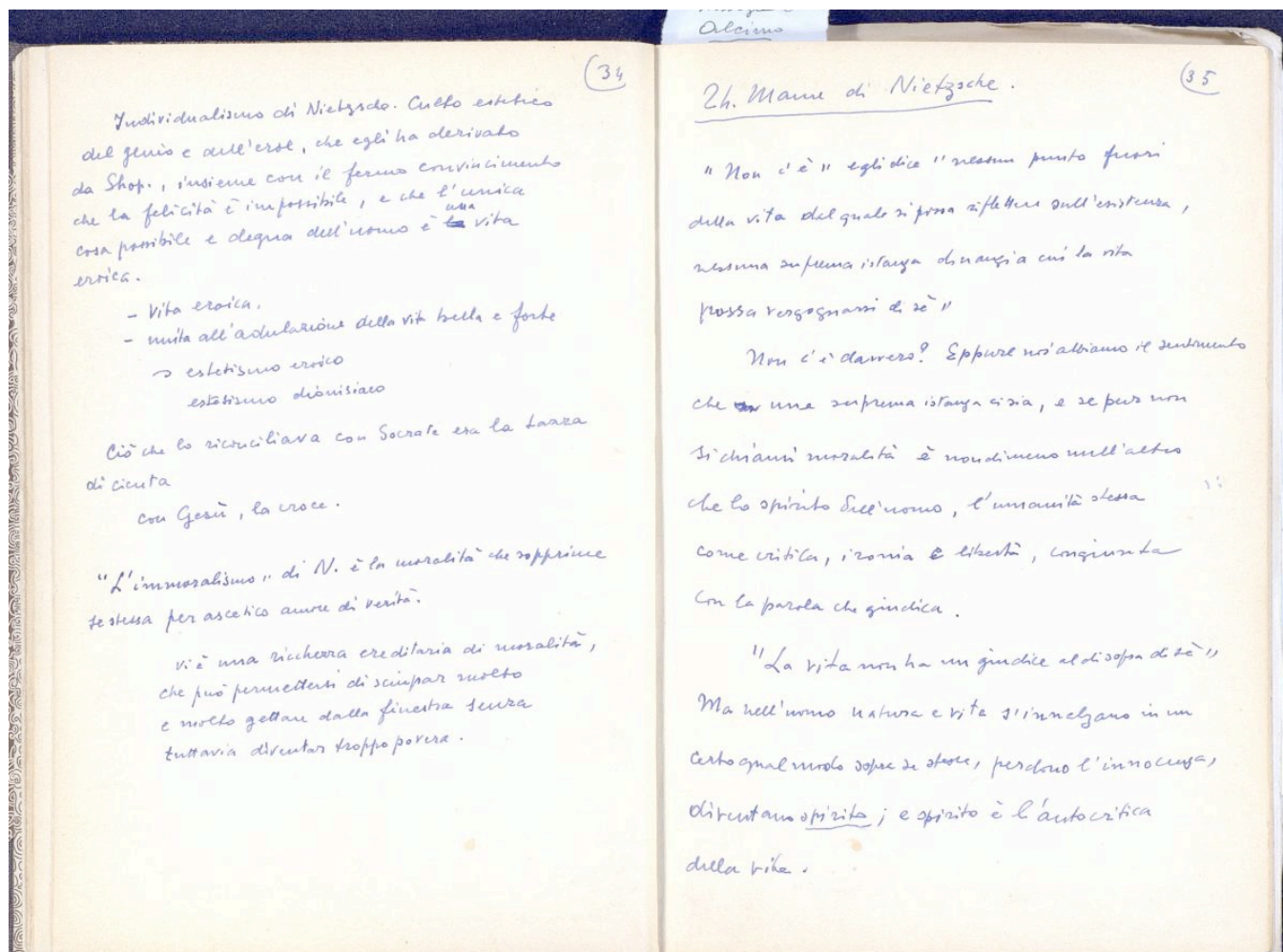


Fig. 1 Facsimile del Quaderno di appunti di P. Bufalini, pp. 34-35

Per esemplificare un caso di annotazione complessa, può essere d'ausilio analizzare una circostanza ricorrente nel *Quaderno*: le relazioni fra i testi e gli autori citati. Le pagine 34-35 (Fig. 1) contengono riflessioni complesse su Nietzsche: secondo Bufalini (o meglio secondo la nostra ipotesi di interpretazione del pensiero di Bufalini – e ciò aggiunge un'ulteriore dimensione al modello di rappresentazione dei dati), il pensiero di Schopenhauer influenza l'idea di individualismo e di impossibilità della felicità sviluppata da Nietzsche. Bufalini riporta un commento di Thomas Mann su Nietzsche e lo espande con considerazioni personali.

La Fig. 2 mostra la codifica XML/TEI di elementi e relazioni coinvolti in questo esempio. Il commento di Thomas Mann su Nietzsche è identificato attraverso il riferimento #bibl074f. Le persone coinvolte (Friedrich Nietzsche e Arthur Schopenhauer) sono identificate univocamente, collegate a fonti esterne riconosciute e ampiamente utilizzate quali VIAF⁷ e DBpedia⁸ (nell'intestazione del file XML/TEI), mentre l'occorrenza dei loro nomi è registrata nel frammento interessato nel corpo del documento (key="FN" e key="AS").

⁶ E la letteratura in materia concorda su questa complessità. Si veda in particolare [5] per la funzione delle ontologie (RDF/OWL) come strumento per l'arricchimento dell'espressività di TEI o ancora [7] come esempio di valorizzazione di TEI attraverso CIDOC-CRM.

⁷ The Virtual International Authority File, <http://viaf.org/>, authority file per l'identificazione di personalità storiche, i cui dati sono pubblicati come Linked Open Data.

⁸ DBpedia, <http://wiki.dbpedia.org/>; <http://it.dbpedia.org/>, Linked Open Data estratti dalle voci di Wikipedia

```

<bibl xml:id="saggi-sch-ni-freud" resp="#CSPC">
  <author key="TM"/>
  <author key="BA" role="translator"/>
  <title level="m" type="WorkCollection" subtype="Book"
    sameAs="http://www.worldcat.org/oclc/83512133">Saggi. Schopenhauer, Nietzsche, Freud</title>
  [...]
  <citedRange xml:id="bibl074f" unit="Excerpt" n="sample-7">sample 7</citedRange>
</bibl>
[...]
<person xml:id="FN">
  <persName sameAs="http://viaf.org/viaf/89798474 http://it.dbpedia.org/resource/Friedrich_Nietzsche">
    Nietzsche, Friedrich Wilhelm (1844-1900)</persName>
</person>
<person xml:id="AS">
  <persName sameAs="http://viaf.org/viaf/17229367 http://it.dbpedia.org/resource/Arthur_Schopenhauer">
    Schopenhauer, Arthur (1788-1860)</persName>
</person>
[...]
<listRelation>
  <relation type="Influence" name="agreesWith" active="#FN" passive="#AS" corresp="#cit-059" source="#TM"/>
</listRelation>
[...]
<quote xml:id="cit-059" source="#bibl074f">
  <p>Individualismo di <persName key="FN">Nietzsche</persName>. Culto estetico<lb/> del genio e
  dell'eroe, che egli ha derivato<lb/> da <persName key="AS">Schop.</persName>, insieme con il fermo
  convincimento<lb/> che la felicità è impossibile, e che l'unica<lb/> cosa possibile e degna
  dell'uomo è <del hand="#PB" rend="overstriked">la </del>una vita<lb/> eroica.</p>
</quote>

```

Fig. 2 Esempio di markup XML di: riferimenti bibliografici (tei:quote + @source), persone (tei:person per la qualificazione + tei:persName per l'annotazione inline) e influenze tra persone citate (tei:relation + attributes)

Da questo sistema di rappresentazione delle citazioni è possibile estrarre le entità (qualificandole attraverso URI come soggetti/oggetti) e relazioni RDF (predicati sulla base delle ontologie scelte). Per rappresentare in RDF queste situazioni articolate, si è optato per il maggior riuso possibile di modelli esistenti, in quanto buona pratica nella comunità del Semantic Web. La scelta delle ontologie ha dovuto sottostare a tre requisiti descrittivi fondamentali, corrispondenti a tre differenti approcci nella rappresentazione della conoscenza e, nello specifico, a diverse ontologie:

- l'articolazione dei diversi livelli che compongono l'oggetto testuale (il contenuto del testo, l'edizione specifica citata, l'esemplare posseduto dall'autore - quando noto). FRBR è il modello scelto per la destrutturazione del testo, un approccio *data-driven* di scomposizione dell'entità documento in quattro livelli descrittivi. Nello specifico è stata utilizzata l'ontologia FaBiO⁹, la formalizzazione OWL di FRBR offerta dalle SPAR Ontologies.
- la specificità delle relazioni: i ruoli (persone citate, autori), tra opere (opere citate da Bufalini, citazioni tra opere terze registrate da Bufalini), influenze tra autori (accordo, disaccordo, disputa, citazione generica). La descrizione di tali situazioni ha richiesto un approccio *situation/event-driven*, così da poter ulteriormente annotare l'eventocitazione con tutti gli elementi di contesto (persone coinvolte, ruoli, testi citati). A tal fine sono state utilizzate le ontologie CiTO¹⁰ e PROV Ontology¹¹.
- la *provenance* delle asserzioni fatte dall'autore e dagli autori citati. Un approccio *interpretation-driven* è indispensabile per preservare il processo di analisi sia dell'autore, che rileva i fenomeni descritti, sia dell'editore, il quale esplicita le relazioni sulla base di una propria interpretazione soggettiva della volontà di Bufalini. Per la rappresentazione dell'atto interpretativo si è scelto di utilizzare il modello delle nanoPublication¹², che prevede la formalizzazione di due livelli di *provenance* (inerente gli asserti estratti dal testo pieno e inerenti la pubblicazione dell'edizione) mediante l'utilizzo di *named graphs*.

Infine, una volta acquisito il modello di markup e stabilite le relazioni da estrarre dal testo pieno, la scelta degli *authority file* e dei dataset da collegare è fondamentale per completare la base di conoscenza e fornire una migliore esperienza informativa all'utente. Sia il riferimento all'identificativo VIAF, sia il link alla rispettiva entità in DBpedia sono stati creati per estendere le informazioni biografiche delle personalità citate. Quando è stato possibile decifrare la specifica edizione di un testo citato è stato creato un link a Worldcat e VIAF; nei casi in cui è nota l'opera, ma non l'edizione citata, è stato creato il solo link a VIAF.

La base di conoscenza, così arricchita grazie all'integrazione di risorse esterne, verrà ulteriormente arricchita con la versione integrale dei frammenti testuali riportati nel *Quaderno*, il che consentirà di ricostruire, anche virtualmente, una parte della biblioteca di Paolo Bufalini.

⁹ FRBR-Aligned Bibliographic Ontology (FaBiO), <http://www.sparontologies.net/ontologies/fabio>

¹⁰ Citation Typing Ontology (CiTO), <http://www.sparontologies.net/ontologies/cito>

¹¹ Provenance Ontology (PROV-O), <https://www.w3.org/TR/prov-o/>

¹² NanoPublication, http://nanopub.org/guidelines/working_draft/

5. CONCLUSIONI

In questo contributo abbiamo sollevato l'attenzione sulle problematiche veicolate dall'adozione dei Linked Open Data nel mondo delle edizioni scientifiche digitali. Per sfruttare a pieno le possibilità offerte dal medium tecnologico (i.e., la possibilità di rappresentare in modo più espressivo i fenomeni estratti dal testo e l'integrazione di dati con fonti esterne) è necessario ripensare l'oggetto testo e le sue forme sulla base del nuovo modello di rappresentazione. Il caso d'uso del *Quaderno di appunti* di Paolo Bufalini è un esempio di come tali problematiche vadano risolte avendo piena coscienza delle fasi di tale processo. Il modello LOD ha consentito di esplicitare le relazioni intra-testuali e inter-testuali tra frammenti di appunti, persone, luoghi, opere citate e il quaderno inteso come documento. Inoltre, ha permesso di esplicitare interpretazioni divergenti su di uno stesso fenomeno testuale, offrendo al lettore-utente la possibilità di accedere al *Quaderno* attraverso molteplici *entry-points*, permettendo di 'svincolare' le informazioni dall'edizione-documento ed integrare il testo del *Quaderno* con altri ambienti di valorizzazione del patrimonio culturale presenti nel web.

Con la definizione di questo workflow, e le relative specifiche di processo, si sono dunque poste la basi per una modellazione sufficiente scalabile in ogni contesto di trasformazione di edizioni digitali in ambienti di conoscenza a base semantica.

6. BIBLIOGRAFIA

- [1] Ciotti, F. and Tomasi, F. 2016. *Formal ontologies, Linked Data and TEI Semantics*. «Journal of the Text Encoding Initiative» 9. <http://jtei.revues.org/1480>.
- [2] Ciotti, F., Daquino, M., and Tomasi, F. 2016. *Text Encoding Initiative semantic modeling. A conceptual workflow proposal*. In: *Digital Libraries on the Move*. «Communications in Computer and Information Science» 612. Berlin, Springer Verlag, pp 48-60.
- [3] Citti, F. 2008. *Paolo Bufalini and the Classics: towards a digital edition of his "Note-book"*. «Conservation Science in Cultural Heritage» 8, pp. 65-87.
- [4] Citti, F. 2010. *I classici nelle carte di un politico. Traduzioni e appunti di Paolo Bufalini*. «Aufidus» 70, pp. 7-32.
- [5] Ciula, A., Spence, P. and Vieira, J.M. 2008. *Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project*. «Literary and Linguistic Computing» 23, 3, pp. 311-325.
- [6] Daquino, M., and Tomasi, F. 2015. *Historical Context Ontology (HiCO): a conceptual model for describing context information of cultural heritage objects*. In: *Research Conference on Metadata and Semantics Research*. Springer, Cham, pp. 424-436.
- [7] Eide, Ø. and Ore, C.-E. 2007. *From TEI to a CIDOC-CRM Conforming Model: Towards a Better Integration between Text Collections and Other Sources of Cultural Historical Documentation*. Paper presented at the DH conference 2007.
- [8] Eide, Ø. 2015. *Ontologies, Data Modeling, and TEI*. «Journal of the Text Encoding Initiative» 8. <http://jtei.revues.org/1191>.
- [9] Franzini, G. 2012-. *Catalogue of Digital Editions*. <https://dig-ed-cat.acdh.oeaw.ac.at/>
- [10] MLA - Modern Language Association. 2011. *Guidelines for Editors of Scholarly Editions*. http://www.mla.org/resources/documents/rep_scholarly/cse_guidelines
- [11] Schmidt, D. 2010. *The inadequacy of embedded markup for cultural heritage texts*. «Literary and Linguistic Computing» April 16, 2010.
- [12] Schmidt, D. 2012. *The role of markup in the digital humanities*. «Historical Social Research» 37, 3, pp. 125-146.
- [13] Tummarello G., Morbidoni C., Pierazzo E. 2005. *Toward textual encoding based on RDF*. In: *Proceedings ELPUB2005 Conference on Electronic Publishing*. Kath. Univ. Leuven, June 2005.

L'habitat come patrimonio e le sue rappresentazioni. Per un atlante digitale dell'insediamento meridionale (XIII-XXI secolo)

Rosanna Rizzi¹, Biagio Salvemini², Lorena Maria Calculi³, Giovanna Patruno⁴

¹ CRIAT, Università degli Studi di Bari, Italia - rosanna.rizzi@gmail.com

² CRIAT, Università degli Studi di Bari, Italia - biagio.salvemini@uniba.it

³ CRIAT, Università degli Studi di Bari, Italia - lcalculi1982@libero.it

⁴ CRIAT, Università degli Studi di Bari, Italia - giannapatruno@gmail.com

ABSTRACT

Il paper qui proposto intende costituire un'occasione di verifica in pubblico degli esiti parziali e dei metodi sperimentati in una ricerca in corso da circa 20 anni presso il Laboratorio di Storia Moderna dell'Università di Bari ed il CRIAT (Centro di Ricerca Interuniversitario per l'Analisi del Territorio).

Il progetto *Per un atlante dell'insediamento meridionale (XIII-XXI secolo)* produce banche dati, cartografia digitale e, in prospettiva, un web-gis interrogabile, sui luoghi abitati come costruzione sociale e culturale – la forma più 'concreta', vistosa ed ingombrante di patrimonio con cui le società di ogni tempo devono fare i conti. Gli strumenti di analisi, costruiti in un processo segnato da ripensamenti ed incertezze, cercano di essere congruenti con una concezione dello spazio umanizzato inteso come sistema polivalente e polifunzionale, dinamico e relazionale, esito di rapporti conflittuali o negoziali fra gruppi sociali localizzati e poteri sovralocali. Dal pluralismo dei poteri, dei corpi territoriali, delle giurisdizioni e delle fonti del diritto, che caratterizzava gli Stati di antico regime e che va riemergendo con forza nella crisi odierna dei territori come spazi inquadrati dentro una ordinata cascata di enti governativi, derivano usi molteplici, al tempo stesso simbolici e strumentali, dello spazio, percezioni e classificazioni diversificate dell'insediamento. Esclusa la possibilità di una realtà insediativa conoscibile in forme univoche ed esaustive, la strada che viene percorsa diventa l'esplorazione di scale analitiche e punti di osservazione molteplici, che presuppongono un livello consistente di indeterminazione negli esiti della ricerca.

PAROLE CHIAVE

History, Settlements, Manuscript, Population, Maps

1. INTRODUZIONE

La ricerca *Per un atlante dell'insediamento meridionale (XIII-XXI secolo)* analizza forme, logiche e funzionamento degli insediamenti umani nel Mezzogiorno d'Italia sul lungo periodo, a partire da fonti archivistiche e cartografiche, di natura civile e religiosa. La trama abitativa meridionale è presentata come esempio di insediamento mediterraneo, e la ricerca individua una metodologia di analisi che, applicata all'intera area del Mediterraneo, si rivela in grado di produrre utili comparazioni. Alla base è una concezione dell'insediamento come fenomeno relazionale e dinamico, prodotto dall'interazione di più soggetti, collettivi o singoli, dotati di molteplici risorse spaziali e svariati strumenti simbolici, materiali e politici.

In opposizione alla titolarità del legittimo potere sugli uomini e sui suoli, gli insediamenti si riconoscono l'un l'altro e costruiscono repertori di località che mutano nella forma e nella dimensione a seconda dei criteri di selezione adottati e delle agenzie di classificazione coinvolte: Stato, Chiesa, feudatari, città dominanti.

Di questi repertori l'Atlante studia la genesi, la logica, la trasformazione, l'efficacia delle identità locali e politiche: ovvero quelle costruzioni sociali e spaziali che oggi rappresentano un patrimonio storico da preservare, e una forma di strutturazione del territorio da gestire nel contesto di società sempre meno insediate spazio.

Utilizzando censimenti di fuochi, tassazioni del Regno di Napoli, registri dello stato delle anime e descrizioni del territorio realizzate per il Sovrano, è stato possibile ottenere database ricchi di dati numerici e statistici collegati a singoli toponimi. Attraverso l'uso del Geographic Information System (GIS) e lo studio dei toponimi storici e delle loro derivazioni attuali si passa alla georeferenziazione degli insediamenti per fornire cartografie dinamiche, ossia capaci di mostrare variazioni dei fenomeni insediativi nel lungo periodo.

2. LE FONTI

Per rendere gestibile e delimitabile la mole delle fonti disponibili attraverso database informatici, si è concentrata l'attenzione sui soli repertori dei centri abitati che coprono l'intero spazio indagato, ossia l'intero Mezzogiorno continentale. Si tratta in generale di liste, corredate di informazioni di vario genere collocabili in tavole a doppia entrata, prodotte dai poteri che hanno o pretendono di avere titolarità su quei luoghi e da saperi che in vario modo rispondono alla domanda di conoscenza che i poteri stessi producono.

In antico regime non esiste alcun repertorio ufficiale dei centri abitati: i poteri insediati producono immagini non coincidenti della griglia dell'insediamento, e mettono in lista luoghi che 'vedono' a partire dalle relazioni che con essi intrecciano. I poteri signorili, ed in particolare quelli ecclesiastici, hanno pretese di inquadramento e regolazione delle società insediate, che producono domande vigorose di conoscenza.

A tali esigenze rispondono, ad esempio, le *Relationes ad limina*, con le quali i vescovi, ogni tre anni a partire dal 1585, riferiscono dello stato morale e materiale delle loro diocesi. Ne derivano insiemi di luoghi ripartiti per vescovadi, con

istituti divisi tra secolari, regolari e caritativo-assistenziali, sistematicamente digitalizzati in database. Un inventario di località che, oltre a non rispecchiare la realtà materiale degli addensamenti edilizi, non si conforma con quello prodotto dal regio erario. Se le chiese curate (e solitamente gli istituti regolari) si collocano il più delle volte in centri di rango elevato classificati dallo Stato come *universitates*, le presenze sacre non curate e spesso gli enti caritativo-assistenziali conferiscono significato, e dunque visibilità, a luoghi minori, celati allo sguardo statale da un infimo status giuridico.

Esiste, infatti, una gerarchia evidente di capacità classificatoria. Il più rilevante produttore di repertori di centri è il fisco regio. In assenza di una autonoma macchina amministrativa in grado di realizzare la raccolta delle imposte, lo Stato premoderno distribuisce la delega fiscale ai centri abitati dotati di frammenti di giurisdizione ed autogoverno – le *universitates* – i cui rappresentanti legittimi si impegnano a raccogliere e corrispondere quote fiscali calcolate in rapporto al rispettivo numero di ‘fuochi’, cioè al numero dei gruppi domestici attivi. Poiché le ‘numerazioni dei fuochi’ sono andate in parte perdute, la ricerca utilizza, accanto ai documenti fiscali superstiti, i dati delle Descrizioni o Corografie del Regno di Napoli (in particolare quelle di Mazzella, Bacco, Caracciolo-Beltrano) – un genere letterario diffusosi fra Cinque e Seicento, che rappresenta il territorio come un insieme di *universitates* raggruppate per province. A metà Settecento lo Stato napoletano, adottando l’obiettivo della equità fiscale diffusosi nell’Europa delle riforme, impone a ciascuna *universitas* di redigere un catasto (detto onciario) che sottragga all’arbitrio il prelievo delle imposte. L’elenco dei centri che compilano il catasto offre ulteriori dati che permettono di prolungare le immagini dell’insediamento prodotte dal fisco regio.

Ma intanto, anche nel Mezzogiorno italiano, queste immagini dell’insediamento come insieme di luoghi dotati di frammenti di potere vanno cedendo il passo ad altre immagini fondate su uno sguardo geografico, geometrico, materializzato dei fenomeni insediativi. La volontà dei sovrani di conoscere il territorio nella sua complessità, di entrare direttamente in rapporto con i gruppi insediati, e non tramite i titolari di immunità e giurisdizioni, produce una domanda di conoscenza a cui risponde una produzione intellettuale del tutto diversa da quella delle corografie cinque-seicentesche. Cartografi geometrici, analisti dello spazio umanizzato per il tramite della peregrinazione e della osservazione diretta, produttori di statistiche, enciclopedisti di luoghi emancipati dalla griglia istituzionale, elaborano repertori del tutto nuovi, che vanno scoprendo e mettendo in lista migliaia di gruppi di case e di uomini rimasti invisibili nei repertori precedenti, vissuti per secoli all’ombra delle *universitates*. In questo ambito sono state utilizzate le opere a stampa o manoscritte di Galanti, Alfano, Di Simone, Sacco e la produzione cartografica di Rizzi Zannoni.

Questi nuovi saperi dello spazio umanizzato vengono assunti come saperi di Stato, fondamento di uno sguardo classificatorio che esclude ogni altro, in particolare quello secolare e potente della Chiesa, formalizza a livello normativo e irrigidisce la fluidità e la pluralità del concetto di insediamento e di centro abitato. Nel passaggio dal Settecento all’Ottocento, la statalizzazione del territorio tende ad espandersi in uno spazio fisico e politico-amministrativo considerato giuridicamente uniforme e lo sguardo classificatorio dello Stato amministrativo assume una fisionomia nuova rispetto al recente passato, in cui la legge veicola un linguaggio univoco di identificazione dei centri abitati (la stessa toponomastica si statalizza) all’interno di una gerarchia apparentemente neutra. Il nuovo sguardo geografico-amministrativo, chiaramente costruito, punta all’uniformità e all’obiettività ma rischia di rimanere parziale, di occultare i fenomeni sociali di lungo periodo riferibili alle questioni dell’insediamento che conserva il suo carattere complesso, processuale e polivalente.

Collocandosi al livello minimo della cascata di enti che coprono il territorio dello Stato, il comune come ente amministrativo territoriale si priva di autonomia politica, si immerge in uno spazio giuridicamente liscio; d’altronde il comune non nega l’esistenza, al di sotto di se stesso, di frammenti insediativi, ma, monopolizzando le funzioni amministrative decentrate, a partire da quelle di inquadramento della popolazione (prima affidate ai parroci ed ora allo stato civile) ne occulta la rilevanza. L’insediamento si traduce in un repertorio di comuni che lo Stato, monopolista della classificazione dei centri abitati, riconosce per legge e censisce sistematicamente.

Utilizzando sistematicamente i censimenti, l’indagine viene prolungata fino ad oggi.

3. METODI E STRUMENTI

Le informazioni contenute nei repertori vengono interpretate, raccolte e schedate in database e a ciascuno dei luoghi abitati documentati, circa 5.000, viene attribuito un codice univoco (ID), che permette di inserirlo in insiemi spaziali e di verificarne la continuità o meno sul lungo periodo. La massa di informazioni così ottenuta è georiferita tramite sistemi di informazione geografica che consentono la produzione di una cartografia dinamica, in grado di restituire il carattere processuale e plurale del fenomeno insediativo.

Si tratta di procedure non prive di insidie: il linguaggio degli strumenti informatici e digitali impone scelte drastiche che sembrano oggettivare, semplificare, irrigidire la straordinaria complessità dei fenomeni insediativi. Per ridurre i rischi di questa natura, sono state adottate soluzioni compromissorie, tali da permettere di rappresentare gli elementi della ricerca. Per l’*Atlante dell’insediamento meridionale*, l’unità minima considerata è stata l’insediamento, nonostante sia un fenomeno dinamico e sfaccettato, squisitamente relazionale e non solo spaziale. Non è un luogo fisico associabile ad un toponimo che permane nel tempo, individuabile geograficamente attraverso un areale ben delimitato da confini noti.

Il problema della rappresentazione sfruttando le potenzialità del GIS ha imposto quindi di individuare una entità tra punto, linea o poligono, nella georeferenziazione dei singoli record derivanti dalle fonti storico-archivistiche utilizzate. Da questo deriva il limite dello strumento, poiché costringe a scelte obbligate. Difatti, dal punto di vista geometrico, non è possibile ricavare l’area fisica, ossia l’estensione effettiva dell’insediamento per la mancanza di fonti grafiche e cartografiche storiche,

non per carenze archivistiche, ma per l'assenza del concetto spaziale tra il XIII e il XIX secolo. I flussi variano nel tempo e in base alla variabile considerata, e il toponimo è una tra le tante.

Per l'impossibilità di pensare realtà spaziali come contenitori inerti di gruppi umani e dunque come realtà dai confini certi, si è deciso, attraverso l'astrazione concettuale dello spazio geografico e l'individuazione di una sorta di "baricentro" dell'insediamento, di ridurlo a punto adimensionale, più neutro e oggettivo di artificiose ed ipotetiche forme spaziali che confliggono oggi con i limiti amministrativi e politici, e, paradossalmente, proprio per questo capace di dialogare con le spazialità multiple, non euclidee, nelle quali si esprime la vita delle società insediate.

Il punto, inteso come entità legata ad un codice numerico (ID), permette un confronto su ampia scala e su un lungo periodo tra fonti di natura diversa, integrandosi in un sistema interrogabile in base alle necessità dell'utente finale e che possa costruire una cartografia versatile e dinamica, in grado di mostrare realtà insediative molteplici e parallele.

In ogni caso, la cartografia digitale costruita in questa maniera evita le trappole tipiche degli atlanti cartacei: da un lato quella del rifiuto del disordine spaziale, la tendenza ad occultare lacune, incertezze, la complessità ed il carattere relazionale dei nessi fra gruppi umani e spaziali; e dall'altra, quella della trasmissione autoritaria, in vario modo occultata dalla 'neutralità' dello strumento, delle letture dello spazio umanizzato prodotte dagli autori. La costruzione di un WebGis permetterà agli utilizzatori di rivolgere alle banche dati domande di varia natura relative ad ogni taglio spaziale, temporale, tematico, ricavandone risposte numeriche, narrative o cartografiche ritagliate sulle domande stesse.

4. OBIETTIVI

L'atlante in costruzione vuole illustrare genesi, logiche, mutamenti, dinamiche relazionali dei fenomeni insediativi che, strutturatisi in processi secolari, si consegnano al nostro presente con il loro ingombro e le loro inerzie, e pongono problemi drammatici di conservazione, gestione, governo. Esso intende essere uno strumento di lavoro anche per chi pensa e progetta le forme dell'abitare nelle società contemporanee, investite da processi di composizione e scomposizione impetuosi dei rapporti fra società e spazi.

Su questo piano l'ambizione dei ricercatori impegnati nel progetto è, nella sostanza, di inserire dubbi ed esitazioni nella vasta schiera di quanti pensano la globalizzazione come passaggio catastrofico da un mondo 'tradizionale' di comunità abbarbicate al suolo, radicate in ambienti puntuali, ad un mondo di individui delocalizzati. Le società che la storia ci consegna realizzano con i loro spazi rapporti complicati, dotati di una straordinaria ricchezza che occorre imparare a decifrare. Ovviamente il mutamento in cui siamo immersi è impetuoso. Il punto è che pensarlo come passaggio dal semplice al complesso, dallo statico al dinamico, dal locale al globale significa semplicemente precludersi la possibilità di comprenderlo.

5. BIBLIOGRAFIA

- [1] Baldassarri M., Mogorovich P., Salvatori E. 2009, *Database, WebGIS, storia ed archeologia: riflessioni metodologiche dietro un progetto sulla Lunigiana medievale*, Siena.
- [2] Barbanente A., Salvemini B. (a cura di) 2004. *Rileggere il territorio*, numero monografico di "Meridiana", n. 49. ISBN: 9788883341649.
- [3] Borri D., Salvemini B. Tedesco C. (a cura di) 2012. *Looking for Multi-dimensional Territories*, numero monografico di "Plurimondi. An International Forum for Research and Debates on Human Settlements", n. 10. ISSN: 1129-4469.
- [4] Caprioli M., Salvemini B., Tarantino E. 2005. *L'insediamento meridionale e la sua rappresentazione cartografica. Temi e prospettive di un atlante storico in costruzione*. In "Bollettino dell'Associazione Italiana di Cartografia", n. 123-4-5, Trieste, pp. 355-367. ISSN 0044-9733.
- [5] Cengarle F., Somaini F. 2008. *Riflessioni e ipotesi di lavoro su storia e cartografia storica*. In "Società e Storia", n. 122, pp. 809-826. ISSN 0391-6987. DOI: 10.3280/SS2008-122007
- [6] Cengarle F., Somaini F. 2009. *La pluralità delle geografie (e delle cartografie) possibili*. In "Reti medievali", pp. 1-17. ISSN 1593-2214. DOI: <http://dx.doi.org/10.6092/1593-2214/66>
- [7] Comba R. 1981. *Il territorio come spazio vissuto. Ricerche geografiche e storiche nella genesi di un tema di storia sociale*. In "Società e Storia", n. 11, pp. 1-27. ISSN 0391-6987.
- [8] De Pinto F., Polignano G., Salvemini B. 2010. *Carte dei moderni, repertori degli antichi. Per una cartografia dell'insediamento pugliese fra antico regime e monarchia amministrativa*, in Denitto A.L. (a cura di), *Atlas: atlante storico della Puglia moderna e contemporanea: materiali su amministrazione, politica, industria*, Bari, pp. 7-28. ISBN: 8872286026.
- [9] Fasano E., Massafra A. 1998. *Un Atlante storico che non si fece, ma...*, in Iachello E., Salvemini B. (a cura di), *Per un Atlante storico del Mezzogiorno e della Sicilia in età moderna*, Napoli, pp. 123-139. ISBN-13: 978-8820727505
- [10] Lepetit B., Ozouf M., Salvemini B. 1993. *Pratiche dello spazio e identità sociali, temi e problemi di una riflessione in corso*, in "Meridiana", n. 18, pp. 141-150. ISSN: 1973-2244.
- [11] Lepetit B., Salvemini B. (a cura di) 1995. *Percezioni dello spazio*, numero monografico di "Quaderni storici", Anno XXX, n. 90. ISBN: 978-88-15-04833-2.
- [12] Massafra A. 2003. *Il 'laboratorio' dell'Atlante storico italiano: un bilancio ancora aperto*. In Ossola C., Verga M., Visceglia M.A. (a cura di), *Religione, cultura e politica nell'Europa dell'età moderna. Studi offerti a Mario Rosa dagli amici*, Firenze, pp. 41-73. ISBN: 9788822252685.
- [13] Patruno G. 2017. *Territorio, insediamento e istituzioni ecclesiastiche nel Mezzogiorno moderno*, Bari. ISBN: 978-88-7228-830-6.
- [14] Romano M., Salvemini B. (a cura di) 2013. *The Italian Territories 150 years after the Unification: between past and future*, numero monografico di "Plurimondi. An International Forum for Research and Debates on Human Settlements", n. 11. ISSN: 1129-4469.

- [15] Salvemini B. 2006. *Il territorio sghembo. Forme e dinamiche degli spazi umani in età moderna*, Bari. ISBN: 978-88-7228-360-8.
- [16] Salvemini B. 2007. *Atlante Storico dell'Insediamento Meridionale (secoli XV-XX)*, in Bordone R., Guglielmotti P., Lombardini S., Torre A. (a cura di), *Lo spazio politico locale in età medievale, moderna e contemporanea, Atti del convegno internazionale di studi: Alessandria 26-7 novembre 2004*, Alessandria. ISBN 978-88-7694-970-8.
- [17] Salvemini B., Spagnoletti A. (a cura di) 2012. *Territori, poteri, rappresentazioni. Studi in onore di Angelo Massafra*, Bari. ISBN: 8872286794.
- [18] Salvemini B. 2014. *Alla ricerca di una 'tara' di lungo periodo: esercizi cartografici sull'insediamento meridionale*. In Galasso G. (a cura di), *Mezzogiorno, Risorgimento e Unità d'Italia. Atti del convegno, 18, 19 e 20 maggio 2011, Roma*, Roma, pp. 107-146.

Presentazione di Czech-IT!: Un corpus linguistico di parlanti cechi apprendenti la lingua italiana

Marco Petolicchio¹, Marcello Bolpagni²

¹ Palacky University Olomouc (CZ) - marco.petolicchio@gmail.com

² Sleska University Opave (CZ) - marcello.bolpagni@gmail.com

ABSTRACT

Obiettivo dell'intervento è quello di presentare il progetto Czech-IT!, un corpus in fase di sviluppo basato sull'analisi quantitativa di parlanti madrelingua ceca in fase di acquisizione dell'italiano. Differenti tipologie di testi sono inserite nel corpus: comunicazioni via mail per analisi di linguaggio burocratico, ricerche sul linguaggio informale (messaggi di testo e conversazioni) e linguaggio parlato e spontaneo. Si utilizzano strumenti di NLP ed automazione per ottenere informazioni morfologiche, sintattiche e POS dei testi.

1. INTRODUZIONE

In questo intervento vorremmo presentare il progetto "Czech-IT! A Linguistic corpus of native Czech learners acquiring Italian language", un corpus linguistico online, reso disponibile alla massa di studiosi e interessati, dal Luglio 2017 e tuttora in fase di sviluppo all'indirizzo web <https://czech-it.github.io/>.

Si tratta di un corpus composto dall'analisi di differenti situazioni comunicative e diversi livelli di esperienza linguistica con la lingua target, che mira a descrivere un ampio spettro di variazione nel processo acquisizionale. La varia modalità delle entrate permette in primis la possibilità di individuare diversi fenomeni linguistici nelle strategie di acquisizione linguistica degli apprendenti, e in seconda analisi di approfondire altri campi di ricerca, come la linguistica contrastiva, la sociolinguistica e la linguistica computazionale.

Attualmente, il corpus Czech-It! conta circa 175 entrate da parte di più di 50 apprendenti, per un totale di circa 8000 parole, a cui si aggiungono le risposte ai questionari a risposta chiusa.

1.1 Motivazioni

Il Ceco (CZ) è una lingua del gruppo Slavo-Occidentale della famiglia indoeuropea (IE) [2, 11], con una spiccata complessità morfologica derivazionale e flessionale, riscontrabile nella formazione delle parole. L'Italiano (IT) è una lingua romanza della famiglia IE, strettamente collegata al latino¹, che esibisce un ampio spettro di variazione distribuito tra dialetti, lingue regionali e stili specialistici. I rapporti commerciali e culturali tra la Repubblica Ceca e l'Italia sono effettivi e profondi, e lo studio della lingua italiana tra parlanti madrelingua ceca è un fatto certamente degno di nota.

Dal punto di vista linguistico, CZ e IT conoscono un insieme di fenomeni divergenti, che permette un'investigazione in senso comparativo, incentrata sugli errori mostrati durante il percorso di acquisizione: ad esempio, l'assenza del sintagma del Determinatore (DP) e la ricca flessione nominale nella sintassi del CZ, laddove l'IT non mostra tale livello di complessità morfologica e non permette un simile uso dell'articolo [4, 15], che risulta in esempi di omissione (ex.1) o forme ipercorrette (ex.2):

(1) Sono vestiti vecchi 10 anni, quindi le uso per ø barca [I001, L001]

(2) Mia insegnante Danka ha il coniuge e lui è l'italiano [I023, L004]

1.2 Metodi

Data la sua natura libera da framework e indipendente da questioni teoriche, Czech-IT! vuole essere una risorsa sia per studi speculativi data-based, che per processi di insegnamento della L2 basati sulle acquisizioni empiriche. Il progetto e l'insieme dei dati raccolti sono rilasciati con licenza Creative Commons Attribution 4.0 International License, grazie alla quale rappresenta una risorsa inseribile nell'universo open knowledge [16].

2. IL CORPUS

Gli studi sul Second Language Acquisition (SLA) rappresentano un terreno fertile di ricerca, sia da una prospettiva applicativa [9] che da una teorica [10, 17]. L'uso di tecnologie computazionali e l'architettura digitale [6, 14, 19] per analisi comparative e quantitative rappresenta un passaggio fondante nell'attuale panorama di sviluppo degli studi linguistici, che apre ad un modello interdisciplinare di ricerca.

2.1 Un corpus basato sugli apprendenti

Molteplici sono le esperienze di corpora basati su apprendenti e notevoli per estensione sono quelli centrati sull'italiano come L2 - citando a titolo di esempio il VALICO (www.valico.org/), il LIPS (www.parlaritaliano.it/index.php/it/dati/653-corpora-lips) ed il corpus multilingue MERLIN (www.merlin-platform.eu/).

Tuttavia, tali strumenti purtroppo non sempre sono sensibili dell'alta variazione sociolinguistica in fase di acquisizione, preferendo l'analisi di un singolo asse linguistico (lo scritto nel caso del Merlin) oppure focalizzandosi su esercizi guidati.

¹ Per il bene di chiarezza: ci riferiamo all'IT nella sua varietà standard. Cfr. [3, 8] per una visione generale dell'IT contemporaneo.

Per tentare di ovviare ad una possibile monodimensionalità e raggiungere l'obiettivo di offrire un ampio spettro di situazioni linguistiche, tale da poter mettere in luce le diverse strategie dell'apprendente, si è scelto di schedare differenti tipi di situazioni comunicative:

- Un subcorpus di email per il linguaggio (quasi-)burocratico e accademico;
- Messaggi di testo come SMS, chat ed altre piattaforme di messaggistica informale;
- Analisi di conversazioni in modalità spontanea;
- Questionari online creati con l'intento di ottenere auto-valutazioni da parte degli apprendenti rispetto al loro processo di acquisizione: i test prevedevano domande a risposta chiusa ed esercizi di scrittura.

I dati sono stati inseriti in una forma base di tipo testuale, in cui sono state archiviate le informazioni sull'apprendente, la data di inserimento nel corpus e le note del revisore, mentre il contenuto testuale di ogni esempio è stato processato attraverso tool automatici che hanno portato ad annotazioni sintattiche, morfologiche e categorizzazioni delle parti del discorso, rilevanti per fini quantitativi e statistici. Allo stato attuale, un dataset primario che contiene le entrate del corpus è collegato ad altri due fogli di calcolo: l'uno relativo agli apprendenti e l'altro alle categorizzazioni manuali dei fenomeni linguistici e trattamento automatico dei testi (tokenizzazione, lemmatizzazione e tagging POS -Part of Speech).

La separazione dei dati non marcati dallo schema di annotazione [13, 12] agevola la possibilità di trattare gli stessi in un più ampio spettro di output: per fini statistici e di data-visualization e si inserisce all'interno dei più recenti studi riguardo lo sviluppo delle iniziative in termini di progettazione digitale e digital scholarship.

Inoltre, l'architettura alla base può essere effettivamente implementata senza la necessità di ripensare la piattaforma nella sua interezza. Ciò permette ai dati di essere indipendenti dagli obiettivi contingenti e facilmente accessibili all'intera comunità di studenti, ricercatori ed utenti.

Tale progetto può essere utile per approcci data-based al SLA e per ricerche teoriche su interlingua, variazione sintattica e linguistica computazionale. Rispetto al corpus orientato sull'asse diamesico, le attività controllate permettono di ottenere informazioni da una prospettiva multifocale sullo SLA e sviluppare corpora orientati in senso diamesico e diafasico, testimoniando un ampio range di situazioni comunicative.

Oltre al modo automatico di analizzare i dati, si segnala anche un set manuale di annotazioni ai testi, che fornisce una tassonomia dei fenomeni linguistici riscontrati. Le annotazioni automatiche e le categorizzazioni manuali permettono al dataset di essere interpretabile tanto dall'utente quanto dalla macchina, il che porta ad un corpus interrogabile in maniera ipertestuale basato su tecnologie web.

Un altro aspetto fondamentale del data management è quello del cosiddetto data curation, ovvero la salvaguardia dei dati stessi durante il processo di acquisizione e di trasformazione, e la delicata questione del loro trattamento [7, 18]. A questo proposito, si è reso necessario anche un sistema coerente di acquisizione dei dati che rispettasse i dati personali degli apprendenti. Così, a ognuno è stato, preventivamente alla collazione dei testi prodotti, presentato un modulo di consenso alla pubblicazione, l'interpolazione e interrogazione a fini scientifici della totalità delle produzioni da egli devolute al progetto Czech-IT. Tutto questo naturalmente garantendo all'autore dei testi, cioè all'apprendente, il diritto delle produzioni linguistiche indicate. Infine, l'apprendente non autorizza invece in nessun caso la pubblicazione di informazioni personali quali il nome, il cognome e l'indirizzo email. In questo modo i dati sono preservati e resi in forma anonimi già in fase di raccolta tramite l'utilizzo di un codice alfanumerico quale identificativo delle entrate del corpus: learners (attraverso la sigla Lxxx) e produzioni linguistiche (Ixxx) (es. L001, I001).

2.2 L'architettura informatica

L'architettura del progetto è distribuita e decentrata: ciò è reso possibile attraverso l'adozione di Github quale piattaforma di interscambio e pubblicazione di dati, e da software specifici per l'ottimizzazione del flusso di lavoro distribuito quali Trello e Slack. In questo modo i curatori del progetto hanno la possibilità di lavorare sugli stessi file nonostante una certa distanza fisica tra essi e comunicare tempestivamente gli aggiornamenti necessari.

Gli strumenti di Natural Language Processing (NLP) sono usati per le attività automatiche: la tokenizzazione ed il conteggio delle parole per ogni frase sono realizzati attraverso l'impiego di librerie dell'ambiente Python specificatamente designate per scopi linguistici, come NLTK[5], mentre l'attività di tagging POS [1] è attualmente in fase di rifinitura e verrà effettuata con Treetagger [20] in base ai parametri sviluppati da M. Baroni.

Similmente, è in programma lo sviluppo di un sistema di annotazione coerente dei dati linguistici così da poter essere fruiti al di là delle necessità contingenti del progetto: metadati ed annotazioni linguistiche utilizzabili sia per processazioni automatiche che per essere interrogati dall'utente.

3. CONCLUSIONI

In questo intervento abbiamo mostrato i principali paradigmi coinvolti nel progetto: una divisione dei subcorpora di tipo diamesico che permette di ottenere informazioni su diverse situazioni linguistiche, un certo numero di parlanti con differenti background, un sistema orientato in diafasia per le forme di autovalutazione dei parlanti.

L'uso delle procedure automatiche per analizzare il linguaggio naturale è coinvolto nelle strategie digitali e computazionali: gli strumenti di NLP permettono di ottenere una divisione in tokens dell'intero corpus, il conteggio delle parole, POS-tagging e lemmatizzazione. Ciò favorisce la possibilità di lavorare con un grande numero di dati a partire da un singolo testo. Gli apprendenti sono inseriti in un dataset con un insieme di informazioni riguardo la loro formazione: grado di istruzione, età, altre lingue conosciute e livello acquisito di Italiano scritto e orale. Quest'ultima informazione è fornita

dagli apprendenti sottoforma di autovalutazione, e, in caso di studenti di corsi di lingua universitari o privati, in sintonia con i requisiti previsti dal corso stesso. Dal punto di vista del data curation, i contributi degli apprendenti sono elaborati con il consenso informato di questi ultimi, e i loro dati personali resi immediatamente anonimi.

Infine, tale procedimento rappresenta un percorso economicamente affrontabile per sviluppare corpora linguistici: attualmente è ospitato su Github ed usa molti software open source per processare i dati. Si è optato, in continuità con le facilitazioni sperimentate dall'uso di materiale aperto, di rendere completamente accessibile alla comunità l'intera collezione di dati in forma aperta.

3.1 Scopi

Un simile progetto si propone dunque come una semplice piattaforma per ricerche linguistiche e per analisi collaborative orientate ai dati. Czech-IT! è un corpus incentrato sull'analisi di alcuni pattern linguistici esibiti durante il processo di acquisizione della seconda lingua (SLA), da ceco (CZ) a italiano (IT).

Il vantaggio di un tale modo di procedere è duplice: da un lato permette una chiara separazione tra i dati e il loro studio, dall'altro offre una modalità di collezione di dati indipendenti da framework teorici, che può essere usata in una grande varietà di ricerche linguistiche.

Tali tipologie di corpora possono essere sfruttati sia per ricerche accademiche che per iniziative private ed aziendali, come nel caso di modelli di insegnamenti nel mondo SLA, orientati su una prospettiva basata empiricamente sull'analisi degli errori e dell'interlingua, così da poter fornire ipotesi mirate ed esercizi centrati sull'apprendente.

3.2 Sviluppi futuri

Il primo passo verso cui ci si vuole orientare è costituito dal superamento del limite minimo di 15000 tokens nel dataset, così da poter risultare in una mole di dati dotata di un certo (seppur piccolo) grado di scientificità. A questo si affiancano le operazioni computazionali sopra discusse, l'adozione di un sistema di markup coerente e la possibilità di disporre di annotazioni sintattiche automatiche. Questi processi ci sembra si inseriscano nella visione di un progetto aperto e disponibile per gli studiosi interessati al tema.

4. BIBLIOGRAFIA

- [1] Steven Abney. «Part-of-Speech Tagging and Partial Parsing». In: *Corpus-Based Methods in Language and Speech Processing*. A cura di Steve Young e Gerrit Bloothoof. Dordrecht: Springer Netherlands, 1997, pp. 118–136. ISBN: 978-94-017-1183-8. DOI: 10.1007/978-94-017-1183-8_4.
- [2] Robert S. P. Beekes e Michiel A. Cor de Vaan. *Comparative Indo-European Linguistics*. Vol. 1. John Benjamins Publishing, 2011. ISBN: 978-90-2721-185-9.
- [3] Giuseppe Berruto. *Sociolinguistica dell'italiano contemporaneo*. Manuali universitari (Rome, Italy): Linguistica. Carocci, 2012. ISBN: 9788843063499.
- [4] Valentina Bianchi. «Sulla struttura funzionale del sintagma nominale italiano». In: *Rivista di Grammatica Generativa* 17 (1992), pp. 105–127.
- [5] Steven Bird, Ewan Klein e Edward Loper. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, 2009. ISBN: 978-0-596-51649-9.
- [6] Alexander Clark. *The Handbook of Computational Linguistics and Natural Language Processing*. Vol. 1. 2010. ISBN: 978-14-0515-581-6.
- [7] Melissa H. Cragin e al. *An Educational Program on Data Curation*. 2007.
- [8] Paolo D'Achille. *L'italiano contemporaneo*. Itinerari (Il Mulino): Linguistica. Il Mulino, 2003. ISBN: 9788815088710.
- [9] Rod Ellis. *The Study of Second Language Acquisition*. Vol. 1. 1994. ISBN: 978-01-9442-257-4.
- [10] Maria Teresa Guasti. *Language Acquisition: The Growth of Grammar*. The MIT Press, 2002. ISBN: 9780262072229.
- [11] Harald Hammarström, Robert Forkel e Martin Haspelmath. *clld/glottolog: Glottolog database 3.0*. Mar. 2017. DOI: 10.5281/zenodo.437430.
- [12] Nancy Ide e James Pustejovsky. *Handbook of Linguistic Annotation*. Springer Netherlands, 2017. ISBN: 9789402408812.
- [13] Nancy Ide e Laurent Romary. «International standard for a linguistic annotation framework». In: *Natural Language Engineering* 10.3-4 (2004), pp. 211–225. DOI: 10.1017/S135132490400350X.
- [14] Mohamed Zakaria Kurdi. *Natural Language Processing and Computational Linguistics*. 2016.
- [15] Giuseppe Longobardi. «Reference and proper names: a theory of N-movement in syntax and logical form». In: *Linguistic Inquiry* (1994), pp. 609–665. ISSN: 0024-3892.
- [16] Marco Petolicchio e Marcello Bolpagni. *Czech-IT! - Linguistic corpus of native Czech learners acquiring Italian language*. Lug. 2017. DOI: 10.5281/zenodo.824984.
- [17] Jason Rothman e Roumyana Slabakova. «The generative approach to SLA and its place in modern Second Language studies». In: *Studies in Second Language Acquisition* (2017), pp. 1–26. DOI: 10.1017/S0272263117000134.
- [18] Arjun Sabharwal. *Digital Curation in the Digital Humanities*. Chandos Publishing, 2015. ISBN: 978-0-08-100143-1. DOI: <https://doi.org/10.1016/B978-0-08-100143-1.00004-0>.
- [19] Heike Zinsmeister Sandra Kuebler. *Corpus Linguistics and Linguistically Annotated Corpora*. annotated edition. Bloomsbury Academic, 2015. ISBN: 1441116753,9781441116758.
- [20] Helmut Schmid. «Probabilistic Part-of-Speech Tagging Using Decision Trees». In: *Proceedings of International Conference on New Methods in Language Processing*. 1994.

Temporal Dimension in Alcide De Gasperi: Past, Present and Future in Historical Political Discourse

Rachele Sprugnoli, Giovanni Moretti, Sara Tonelli

Digital Humanities Group, Fondazione Bruno Kessler
Via Sommarive 18, 38123 Trento
{sprugnoli,moretti,satonelli}@fbk.eu

1. INTRODUCTION

Political discourse is grounded in a multi-dimensional deixis encoded in the language used by the speaker/writer. Chilton (2004) proposes a discourse space model, that is a conceptual structure made by three fundamental axes: time, space and modality. These axes intersect in a deictic center representing the *here* and *now* of the speaker/writer (Hart, 2007). From its theorization, this model has become a reference in the Humanities and Social Science fields to study rhetoric, persuasion and social identity in contemporary political communication (De Cock, 2005; Ferrari, 2007; Chilton, 2017). Our work aims at extending this type of analysis to historical political discourse by taking into consideration the complete corpus of Alcide De Gasperi used within the De Gasperi Project (Sprugnoli et al., 2016) and by taking advantage of state-of-the-art Natural Language Processing (NLP) and Semantic Web resources. The application of automatic tools is crucial to cover the vast corpus written by De Gasperi (almost 3,000 documents and 3 millions tokens spanning from 1901 to 1954) and provides a novel, comprehensive insight into his works.

This abstract focuses on the temporal dimension because it has a high political value: in political discourse, past, present and future are put in a conceptual opposition determining the political identity of the speaker/writer and characterizing her/his rhetorical style. Classical rhetoric, as defined by Aristotle, can be distinguished into three classes each corresponding to a different type of temporality: *judicial* rhetoric focuses on the past, *epideictic* rhetoric on the present and *deliberative* rhetoric is linked to the future (Dunmire, 2008). Due to their varied content, contemporary political discourses are characterized by the simultaneous presence of the aforementioned three forms of rhetoric (Wodak and De Cillia, 2007; Minniti, 2015). Using state-of-the-art NLP, it is now possible to study this aspect on a large scale and apply it to corpora of the past, which have been extensively studied only using qualitative approaches.

2. AUTOMATIC ANNOTATION OF TEMPORAL DIMENSION

Chilton (2004) applies his Time-Space-Modality model to a set of political discourses by manually identifying language references to the three dimensions within the texts. Other manual annotations are reported in the literature: Shenhav (2005) codes the paragraphs of texts that point to ten different periods of times in Jewish history, while Cienki et al. (2008) propose a complete tagging of spatial, temporal and modality words to detect the worldview of Dutch parties. In both cases, the authors apply their methods to small text corpora which are manually annotated. In our work, instead, we employ state-of-the-art NLP tools for the automatic identification of temporal information in order to process a bigger set of data and to widen the scope of the annotation by including several types of textual references. Specifically, we focus on the main textual elements conveying temporality information, i.e. temporal expressions, verbal morphology and mentions of persons and events. Each of these elements was recognised in De Gasperi's corpus and categorized in three different classes (*Past*, *Present* and *Future*) by considering the date of publication, provided as a metadata of the document, as the deictic center of each writing. The above mentioned textual elements were detected as follows:

- *precise and fuzzy dates* were extracted by using HeidelTime (Strötgen and Gertz, 2013). This system recognizes temporal expressions linguistically denoted by several lexical triggers (e.g. proper names like “Natale” and nouns like “estate”) and associates each temporal expression to its value, that is a normalized representation of the time it denotes; in this way, also relative information (e.g. “ieri”) could be associated with a precise date starting from a given temporal anchor (e.g. “1 Ottobre 1944”). HeidelTime evaluation on a subset of De Gasperi's writings reports an F1 score of 0.79 (Manfredi et al., 2014). The dates were further labeled as *Present*, *Past* or *Future* given the document date .
- *verb tense* was extracted with the morphological analyzer included in TINT, a pipeline for Italian NLP (Palmero Aprosio and Moretti, 2016); this information was used to associate each verb to a *Present*, *Past* or *Future* label.
- cited *persons* and *events* identified through the Wiki Machine (Palmero Aprosio and Giuliano, 2016), a tool that automatically links the expressions mentioned in a text to their corresponding Wikipedia page and Dbpedia entry (reported linking performance on English: P 0.78, R 0.74, F1 0.76). Through this connection, it is possible to associate each person and event mention to a date (e.g., *dbp:date* for entities of type Event; *dbo : birthDate* and *dbo:deathDate* for entities of type Person) and then to a *Past* or *Present* label, again using the document date as a reference. For persons' mentions, they were considered part of the past if the referent was dead before the document publication time. In a similar way, events are labelled as *Past* if occurred in a year prior to the year of creation of the document.

3. ANALYSIS AND DISCUSSION

The automatic annotation described in the previous section was applied to 2,762 documents divided into four volumes of De Gasperi's writings (De Gasperi, 2006; 2008a; 2008b; 2009): each volume contains documents written in a specific period of De Gasperi's life, in which he was holding different public roles. Furthermore, the documents have been manually classified with a "Type" label according to a taxonomy developed by history scholars. Specifically, such types include electoral discourses (i.e. propaganda speeches), official documents, speeches uttered during an institutional meeting, speeches delivered during party meetings, press reports in both local and national newspapers and monographies. For each volume and each document type, we analyse the impact of *Past*, *Present* and *Future*, trying to assess whether there are relevant differences across the document groups, and if these differences can be motivated by the document type or some specific content. Fig.1 shows the outcome of this comparison across the four volumes.

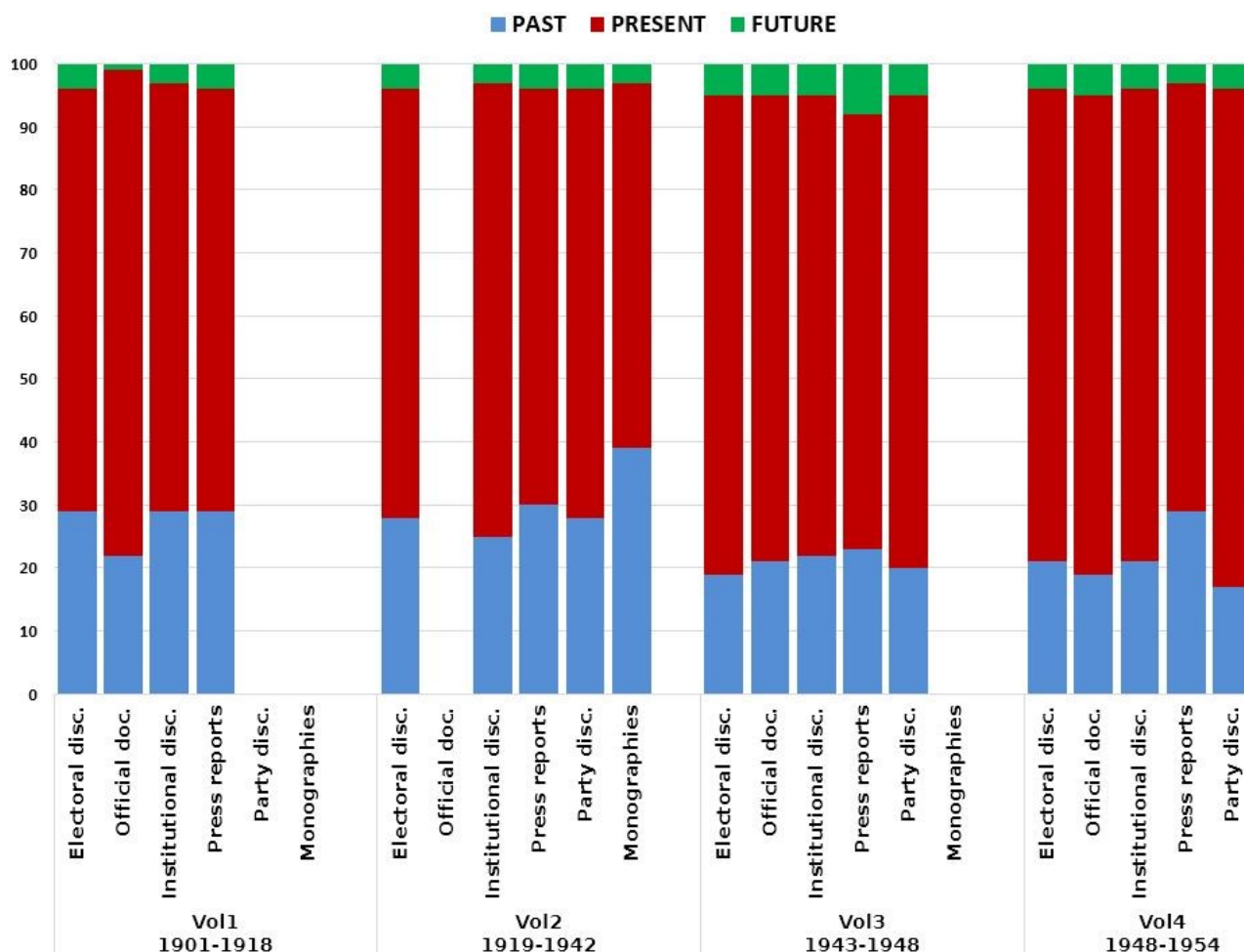


Fig. 1: Percentage of textual elements encoding *Past*, *Present* and *Future* per document type and per volume.

Present prevails in all the classes of documents, while there are peaks related to the *Past* in monographies and press reports. A detailed analysis at document level shows that the *Past* class is relevant also in specific types of institutional discourses such as speeches in commemoration of public figures (e.g. Matteotti) or reporting tragic events of national significance (e.g. the explosion of the Panigaglia ship). As for *Future*, its predominance with respect to *Past* characterizes different documents in different volumes, that is in different periods of Italian history. In the first volume, it marks press reports where the author makes socio-political provisions¹ but also exhortatory articles where De Gasperi, for example, tries to convince the readers not to emigrate from Trentino or to go vote. The period covered by the second volume saw the emergence of political parties in which De Gasperi played a key role, i.e. Partito Popolare Italiano and Democrazia Cristiana: in this context, the *Future* class is relevant in documents presenting the program of these parties. The programmatic commitments of Democrazia Cristiana are also discussed in many press reports published in the third volume: this explains the higher percentage of *Future* in that category and in that volume compared to the others. Moreover, the same volume contains letters and telegrams sent between 1944 and 1946 by De Gasperi as Minister of Foreign Affairs to his international counterparts: in these documents, the *Future* class is particularly present given that the main subject is the development of prospective collaborations. In the fourth volume, *Future* emerges in two main types of documents: those

¹ An interesting example is the article "La guerra universale del 1911..." published in 1908, in which De Gasperi predicts the outbreak of a decisive war in few years based on the analysis of armaments growth worldwide.

discussing the reforms to be adopted for the reconstruction of the newly born Italian Republic and those about the forthcoming creation of a European Community.

Other insights come from the analysis of mentioned people and events. As for the former, Table 1 shows the three most cited persons for each volume and for the *Present* and *Past* classes. Note that, on the basis of the document publication time, they can be labeled as belonging to *Present* or *Past* even within the same volume. Mussolini, Battisti and Mazzini, for example, are the only three persons mentioned in all the four volumes: during De Gasperi's life the first two change from *Present* to *Past*.

	PRESENT	PAST
VOL1	E.Conci C.Battisti R.Murri	L.van Beethoven K.Marx A.Gasparini
VOL2	B.Mussolini G.Giolitti I.Bonomi	Leone XIII O.von Bismarck K.Marx
VOL3	P.Togliatti P.Nenni I.Bonomi	T.W.Wilson C.Battisti B.Mussolini
VOL4	P.Nenni P.Togliatti J.B.Tito	B.Mussolini C.Endrici G.Giolitti

Tabl. Top-three mentioned persons labelled as *Present* or *Past*

Linking each person's name to Wikipedia using the Wiki Machine allowed us to classify mentioned people according to Wikipedia categories and, as a consequence, understand the change of De Gasperi's attitude over time. In the first volume, there is a high variety of categories, in particular regarding the people of the past: for example, De Gasperi refers to many different artists (e.g. Shakespeare), philosophers (e.g. Engels), monarchs (e.g. Federico II di Svevia) and religious figures (e.g. San Francesco). The latter are the most frequent in the second volume, where many Popes are mentioned. In the last two volumes, mentions are more uniform with a strong prevalence of politicians among both contemporary and past people. For what concerns the events, Risorgimento is frequently mentioned in all the volumes. As already seen for people mentions, also events are more varied in the first volume, spanning from the classical era to modern and contemporary times. There, many military conflicts are cited: from the battle of the Raudine Plain to the battle of Sadowa and the Italo-Turkish War. Mentions of a distant past are less recurring in the other volumes: the Second World War and the liberation of Rome in 1944 appear in the third volume and the cold war burst into the fourth.

4. CONCLUSIONS AND FUTURE WORK

In this proposal, we present an innovative method to analyse the temporal dimension of political discourse on a large scale. This method has been applied to a specific corpus of historical writings, showing how the interplay between temporal dimension and document type or content can be automatically detected. The methodology could be easily applied to other corpora, at least in Italian and English, for which the necessary NLP tools are already available.

As for future work, a finer-grained distinction between near and distant past/future could be introduced, following a suggestion provided by history scholars. Then, the other dimensions defined by Chilton - *Space* and *Modality* - could be also taken into account, given the availability of tools to identify them automatically. Finally, in order to make this study accessible to a broad public, we would like to create a set of interactive visualisations, allowing users to navigate the analyses and possibly retrieve specific documents from the corpus by integrating them in the ALCIDE platform (Moretti et al., 2016).

5. REFERENCES

- [1] Chilton, Paul. 2004. *Analysing political discourse: Theory and practice*. Routledge.
- [2] Chilton, Paul. 2017. Toward a neuro-cognitive model of socio-political discourse, and an application to the populist discourse of Donald Trump. *Langage et société* 2(2017): 237-249.
- [3] Cienki, A., Kaal, B. and Maks, I., 2010. Mapping world view in political texts using Discourse Space Theory: Metaphor as an analytical tool. In *RaAM 8 conference, Vrije Universiteit Amsterdam*.
- [4] De Cock, Barbara. 2005. Spain, Portugal and Europe in Spanish international relations discourse: a linguistic approach to group and identity construction. *España en la Europa del siglo XX. Cuadernos de Yuste* 3(2005): 279-300.
- [5] De Gasperi, A. 2006. Scritti e discorsi politici di Alcide De Gasperi. Alcide De Gasperi nel Trentino asburgico. Edited by E. Tonezzer, M. Bigaran, M. Guiotto. Bologna: Il Mulino.
- [6] De Gasperi, A. 2008a. Scritti e discorsi politici di Alcide De Gasperi. Alcide De Gasperi dal Partito popolare italiano all'esilio interno 1919-1942. Edited by M. Bigaran, M. Cau. Bologna: Il Mulino.
- [7] De Gasperi, A. 2008b. Scritti e discorsi politici di Alcide De Gasperi. Alcide De Gasperi e la fondazione della Democrazia cristiana, 1943-1948. Edited by V. Capperucci, S. Lorenzini. Bologna: Il Mulino.
- [8] De Gasperi, A. 2006. Scritti e discorsi politici di Alcide De Gasperi. Alcide de Gasperi e la stabilizzazione della Repubblica 1948-1954. Edited by S. Lorenzini, B. Taverni. Bologna: Il Mulino.

- [9] Dunmire, Patricia L. 2008. The rhetoric of temporality. *Rhetoric in Detail: Discourse Analyses of Rhetorical Talk and Text*, 31, 81.
- [10] Ferrari, Federica. 2007. Metaphor at work in the analysis of political discourse: investigating a 'preventive war' persuasion strategy. *Discourse & Society* 18, 5(2007): 603-625.
- [11] Hart, Christopher. 2007. Critical discourse analysis and conceptualisation: Mental spaces, blended spaces and discourse spaces. In *Cognitive linguistics in critical discourse analysis: Application and theory*. Cambridge Scholars Publishing.
- [12] Manfredi, Giulio, Jannik Strötgen, Julian Zell and Michael Gertz. 2014. HeidelTime at EVENTI: Tuning Italian Resources and Addressing TimeML's Empty Tags. In *Proceedings of the Fourth International Workshop EVALITA* (pp. 39-43).
- [13] Minniti, Mariacarmela. 2015. L'analisi Critica del Discorso Politico: Principali Caratteristiche e Peculiarità. *Illuminazioni*, n. 3 .
- [14] Moretti, Giovanni, Rachele Sprugnoli, Stefano Menini and Sara Tonelli. 2016. ALCIDE: Extracting and visualising content from large document collections to support humanities studies. *Knowledge-Based Systems* 11: 100-112.
- [15] Palmero Aprosio, Alessio, and Giovanni Moretti. 2016. Italy goes to Stanford: a collection of CoreNLP modules for Italian. *arXiv preprint arXiv:1609.06204*.
- [16] Palmero Aprosio, Alessio, and Claudio Giuliano. 2016. The Wiki Machine: an open source software for entity linking and enrichment. *arXiv preprint arXiv*.
- [17] Shenhav, Shaul R. 2005. Concise narratives: a structural analysis of political discourse. *Discourse Studies*, 7(3), 315-335.
- [18] Sprugnoli, Rachele, Giovanni Moretti, Sara Tonelli, and Stefano Menini. 2016. Fifty years of European history through the Lens of Computational Linguistics: the De Gasperi Project. *Italian Journal of Italian Computational Linguistics*, 2(2016):2.
- [19] Strötgen, Jannik, and Gertz. 2013. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation* 47, no. 2 (2013): 269-298.
- [20] Wodak, Ruth, De Cillia, Rudolf. 2007. Commemorating the past: the discursive construction of official narratives about the Rebirth of the Second Austrian Republic. *Discourse & Communication*, 1(3), 337-363.

Thinking to Preservation.

A Proposal for Innovation in Digital Library

Nicola Barbuti¹, Stefano Ferilli², Domenico Redavid³, Tommaso Caldarola⁴

¹ Department of Humanities DISUM - University of Bari Aldo Moro - nicola.barbuti@uniba.it

² Department of Computer Science - University of Bari Aldo Moro - stefano.ferilli@uniba.it

³ Department of Computer Science - University of Bari Aldo Moro - domenico.redavid1@uniba.it

⁴ D.A.BI.MUS. S.r.l., Bari, Italy - t.caldarola@dabimus.it

ABSTRACT

Contemporary libraries have changed their social role and function due to the proliferation and diversification of multimedia digital documents, becoming complex networks able to support communication and collaboration among the various distributed users communities. Technologies have not grown in step with the needs generated by this new approach, except in specific areas and implications. Hence the need to design an integrated digital library architecture that covers by advanced techniques the whole spectrum of functionality, without which the same social and cultural function of a modern digital library is at risk. This paper briefly describes an architecture that aims to bridge this gap, bringing together the experience, expertise and software systems developed by university and companies researchers. A prototype of the system has been developed.

KEYWORDS

Digital Library, Digital Library Management System, Digital Recognition, Layout Analysis

1. INTRODUCTION

Librarianship has evolved and taken new study and research fields in order to conceive and design new systems able to renew the management of information and, at the same time, to promote both cooperation between users and integration of heterogeneous information resources, ushering in the era of *Digital Libraries* (DL) [1].

Over the years, digital libraries (DL) considerably has evolved from simple digital interface of public libraries' physical collections in complex networks. Citizens can access, discuss, evaluate and develop different types of information content by contemporary DL [2].

Facing to the progress in research aimed at creation of DLMS able to support DL containing homogeneous digital collections and metadata, there are still hard difficulties in designing and implementing DL effectively integrating, managing and making accessible multimedia digital objects and related metadata.

This paper proposes an "integrated" DL architecture. The sections are structured as follow: section 2 presents a comparison about the major open source Digital Library Management Systems (DLMS) finalized to choose the best one to our aim. Section 3 details the components that should be integrated or enhanced in the dSpace architecture in order to realize the set of innovative features identified by the component itself. Section 4 gives an overview of some innovative features of the proposed DLMS, in particular the management of different types of metadata and text extraction and indexing from scanned digital objects. Section 5 gives the conclusions.

2. OPEN SOURCE DLMS

The definition of the proposed architecture has required an analysis of available open source DLMS [3] in order to choose the one more suitable to our aims. Among the most known and used we have considered and compared *dSpace*, *EPrints* and *Greenstone* [4]. *Fedora Commons* has not been taken into account because it doesn't adequately support MARC protocol and the bibliographic data exchange protocol Z39.50, a non-negligible characteristics for the purposes of the DL to be realized. By the comparison between the three identified DLMS, the choice has fallen on dSpace essentially for the following three factors:

1. it has a more complete support to the various metadata standards and protocols for interoperability;
2. it has been developed using only one programming language;
3. it has a more complete documentation and there are various communities that provide support.

3. THE ARCHITECTURE

The proposed DLMS extends the architecture of dSpace adding specific modules for the management of innovative features.

The architecture is divided into three levels.

1. The Application Layer. This level includes the following components for access to the system:
 - Web UI is the module for the Web-based access both to the back-office area (via IDPs) and into front-end of the Digital Library through different portals. The back-office area allows the insertion and editing of digital content and its metadata, as well as managing users for access. The front-end allows visualization and rendering of contents through a Web interface that combines all advanced file formats guaranteeing multimedia,

- multichannel and protection. It also presents the front-end for the collaborative tagging (operated by Web 2.0 module).
- Mobile Devices allows viewing and Rendering of contents through tablets and smartphones.
 - Monitoring is the module that allows you to observe the behavior of the system and to produce reports in Excel format, XML, and PDF with the possibility of representation through graphs.
 - The I/O module include interfaces that allow the metadata exchange using OAIS, OAI-PMH, Z39.50 and OAI-ORE. It also enable the rendering of content as Open Data.
2. The *Logic level* includes the following modules suitable for the realization of the system functionalities:
 - Web 2.0 is the module for the management of collaborative tagging, i.e. the module that allows to perform the necessary functions to support the active participation of the portal users to the published contents.
 - Access Management is the module for user authentication and profiling. It allows the access management both in normal mode or through IDP.
 - Text Extractor is the module that enables the extraction of text from documents via ICRPad System (see sec. 4).
 - Content Manager is the module designed to manage the objects, collections and licenses within the DL.
 - Cataloging is the module that allows to manage digital contents that will populate the DL. The MAG/ICCU standard (version 2.0.1) compatible with standard METS is used.
 - Core Tools is the module containing basic entities for system configuration and logging.
 - Search Engine is the module that implements the functionality to support the information finding. Content indexing is done by the Lucene open source tool.
 - Georeferentiation is the module that allows to reference the contents in accordance with their geographic coordinates. This module allows the storage and retrieval of content on the basis of spatial queries.
 3. The *Storage level* manages access to the physical resources of the system deals with the organization of the content, including metadata, information about users and permissions associated with them, the status of the approval flow during the insertion of a content. Specifically:
 - RDBMS Wrapper is the module that allows read/write access to the particular implementation of the DB (in this DLMS is PostgreSQL, but is easily extendable to other types, eg. Oracle).
 - Bitstream Storage Manager is the module that permits the storage on the file system or SRB (backup tool).

The DL operations will involve the different architectural modules. In addition to functionalities of a classical DL in the next section will be described some of the innovative aspects that could be integrated.

4. SOME INNOVATIVE FEATURES

We adopt a standard representation for administrative and descriptive metadata involving the use of different languages and hardly compatible with each other, to enable the efficient management and retrieval of content with different representation formats. To this end, it is necessary to foresee from the beginning the integration of tools that enable a separate management of metadata according to the various representation standards.

In order to facilitate interoperability, it is necessary to implement the dSpace basic architecture and extend its ability to support exchange protocols. The standards supported by the DLMS interface require the integration of tools which ensure the proper functioning and implement approaches that permit the mapping among the different metadata representation usable in the standard interface. One of the most innovative aspects is related to the high interactivity that the DL ensures to diverse user communities in response to their cognitive demands. To this end, into some modules could be implemented features that are not natively supported by dSpace.

- In *Web UI* module could be integrated recognition features, graphic matching and text extraction from digital objects into PDF/A with printed and manuscripts content using the digital platform for the recognition ICRPad [5]. It will support text-based lexical indexing through state of the art technology provided by Lucene and those based on cooccurrences and semantics provided by the system DOMINUS [6] which also includes advanced features for the selective reading of interesting portions of digital format documents. The DLMS will support different digital objects including FITS (Flexible Image Transport System) for the images.
- In *Profiling and Authentication* module will be implemented a user profiling behavior functionality aimed at customization of services: using advanced techniques developed in the Artificial Intelligence field for tracing the interactions of each individual user with services, it will make possible to infer specific information regarding various spheres such as the special interests, the preferences of interaction, the goals, routine activities, and more. Based on this information, customized service for each user will be provided, making its experience with the DL easier and more productive [7].
- In *Cataloging* module will be included a common higher-level metadata schema based on Semantic Web languages. This will allow to take advantages offered by the abstract properties of the metadata during the search operation of the DL contents.

5. CONCLUSIONS

The increase of new technologies, evolving from digital counterpart of public libraries' physical collections in complex networks able to support communication and collaboration among users communities worldwide, created the need for integrated systems to manage DL by advanced functionality. This paper has proposed an innovative DL architecture that aims to bridge this gap by new features, such as the integration of technologies for processing documents covering all phases: acquisition, content extraction, indexing, searching and enjoyment. A beta-system has been developed with very interesting perspective which could eventually lead to realize a DL model that functions as an integrated system for preservation, management and use of complex multimedia digital objects.

6. BIBLIOGRAFIA

- [1] Y. Arms William. The1990s: The Formative Years of Digital Libraries. *Library Hi Tech*, 30(4), 2012, pp. 579 – 591.
- [2] M. Agosti. Digital Libraries. *Mondo Digitale*, n. 43, settembre 2012, pp. 1-13.
- [3] M. Andro, E. Asselin & M. Maisonneuve. Digital libraries: Comparision of 10 software. *Library Collections, Acquisitions, and Technical Services*, Vol. 36, n. 3-4, 2012, pp. 79-83.
- [4] S. H. Trambo, S. M. Shafi, S. Gul. A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone. *International Journal of Computer Applications (0975–8887)*, Volume 59 – No.16, December 2012.
- [5] N. Barbuti & T. Caldarola. An innovative character recognition for ancient book and archival materials: A segmentation and selflearning based approach. In M. Agosti, F. Esposito, S. Ferilli, N. Ferro (Ed.), *Communications in Computer and Information Science*. Vol. 354: Digital Libraries and Archives, IRCDL 2012, Heidelberg: Springer, (pp. 261-270).
- [6] S. Ferilli, F. Esposito, T.M.A. Basile, D. Redavid & I. Villani. DOMINUS^{plus} - DOcument Management INTelligent Universal System (plus). In M. Agosti, F. Esposito, C. Meghini, N. Orio (Eds.), *Digital Libraries and Archives - Post-proceedings of the 7th Italian Research Conference (IRCDL-2011)*, *Communications in Computer and Information Science* 249, 123-126, Springer, 2011.
- [7] G. Semeraro, M. F. Costabile, F. Esposito, N. Fanizzi & S. Ferilli. Machine Learning Techniques for Adaptive User Interfaces in a Corporate Digital Library Service. In *Machine Learning and Applications, Proceedings of the ACAI-99 Workshop W03 on Machine Learning in User Modeling*, 21-29, Chania, Crete, Greece, July 5-16, 1999.

AIUCD 2018
Short papers

An experiment on the development of a digital edition for ancient Greek fragmentary poetry: A case study on Archilochus of Paros

Alberto Stefanini¹, Anika Nicolosi², Monica Monachini³

¹ Novarecon, Novara, Italy - alberto_stefanini@libero.it

² Dip. Discipline Umanistiche, Sociali e delle Imprese Culturali, Parma University, Italy - anika.nicolosi@unipr.it

³ ILC-CNR, Pisa, Italy - monica.monachini@ilc.cnr.it

ABSTRACT

This paper overviews ongoing experiments on a digital edition of Archilochus which is based on the readings, translations and comments by Nicolosi [1] and also integrates feedback and requirements from the Digital Classics community. The experiment encompasses a few fragments of the poet of Paros, so as to provide a mock-up of the prototype for evaluation by its intended end-users, in view of developing a fully fledged digital edition. The mock-up provides the philologist with a set of resources and tools that ease a critical appraisal of the text.

KEYWORDS

Digital methods in the humanities, Interfaces and user-friendly data presentation

1. INTRODUCTION

When adapting an existing text to a digital format, the key concern is how to exploit to the best user profit the ample repertory of solutions and resources digitalization may offer. These include all the features a hypertext provides, through linking the target text to other texts that, according to Genette [1]: intertextual (quotation, plagiarism, allusion), metatextual (i.e. through a critic, reflexive relationship), architextual (as belonging to a literary genre) paratextual (i.e. with its textual periphery) or hypertextual (the relation between a text and a preceding hypotext; examples are parody, spoof, sequel, and translation). These relationships may also include reference to existing databases, where witness are digitized, to imagery (of objects, paysages and tools referred by the poet) and geographical databases, where places referred by the text are described. They may refer to linguistic resources such as dictionaries, logical and syntactic analysers, automatic translators. Existing off-the-shelf solutions provide an interface that basically mimics the printed page of a commentary book. A far more flexible and user friendly solution may be envisaged: to this aim a mock-up of an e-book was prototyped. The intended audience for the e-book are university students and scholars. The mock-up provides and collects support resources and tools that cope with the outcomes of a survey on the use of digital resources and related tools in Ancient Greek scholarship, performed through a questionnaire to ascertain the current practice and the related needs within a group of practitioners in the field¹.

2. MOTIVATION AND OUTLINE

Until the early 2000s, critical digital editions have replicated the print edition model, both in terms of product (i.e. materials included and how they are accessed) and process (i.e. how they are and how they can be used). Only after 2003 this model began to be forced by exploring the potential of the world wide web [2]. At the same time, the development of digital imaging technology and their dissemination on the create repertoires of images of the original witness. However, ten years of intense activity (since the mid-1990s) in the production of critical digital editions have not led to a shared model: the discipline of Digital Humanities has not elaborated a commonly agreed definition of what a digital edition is.

As production of a digital edition is a large endeavour and requires collaboration by a team of persons, its development must follow an iterative practice, where from user needs analysis, users profiles are defined, prototypes of the digital product are produced and are evaluated in a somewhat formalized way, so that the product is modified to undergo a further evaluation stage. The design of specialized digital editions may be inspired by the life cycle of traditional editions but should also adhere to the principles of web design, which make iterative project design and evaluation a fundamental principle: a graphic object must be beautiful to see and inspire confidence also through its aesthetic qualities [3].² These remarks are consistent with a number of studies concerning the way humanists approach and make use of material available on the web [4] [5] [6], all of them stressing on a set of recommendations that emphasize interface usability, whose key tenet are iterative design and evaluation. In that respect, it is necessary to distinguish between wireframes, mock-ups and prototypes. The wireframe is a black and white, low cost sketch of the main views of an interface, since it is basically a

¹ The survey (<http://hdl.handle.net/20.500.11752/OPEN-86>) targeted the scientific community dealing with Digital Classics; it was performed on a restricted sample of Italian digital humanists with focus of interest on ancient Greek philology. The survey was performed from May to September 2016 and it is now available on-line at <http://www.clarin-it.it/content/sondaggio-current-practice-digital-classics-tools>.

² This book provides principles for the design of interfaces for publishing in the field of visual arts and focuses on how to make browsing more interesting and flexible, assist the cognitive path of the reader and search for recurring graphic motifs. As noted by Pierazzo, the topic has so far received little attention in the Digital Humanities sector, although it is today a key theme linked to the enormous development of the world wide web.

sketch on a drawing board or equivalent graphic tool. A mock-up is a more detailed design in which the detailed color rendering display and the fonts to be used are developed. It is an intermediate step to building a prototype of the interface, which is obviously a rather costly operation, since it is essentially an interface design that preludes to full software implementation. A mock-up and even more a prototype allows a fairly comprehensive evaluation of the interface. This can be done with semi-quantitative criteria, by developing an appropriate metric and a questionnaire to evaluate a user's reaction to mock-ups. An extensive bibliography on the methodologies for evaluating websites can be consulted, see a recent work by [7].

The mock-up we present below³ may be considered as an **augmented** digital edition compared to a simple digitization of a paper text, as it has a set of features that a simple digital edition does not present such as:

- a basic hypertext structure
- multimedia integration of text, images and graphics
- integration of material from multiple sources
- textual search tools
- reading support tools such as online translation, vocabulary, syntactic tree consultation

The view shows the so called general functions, encompassing translation, metrics, and bibliography, the syntactic tree of the fragment, the sources, some linguistic peculiarity and geographic information through invocation of the geographic databases Pleiades. The buttons 'introduzione generale' and 'bibliografia' switch to full pages where a general introduction to the fragment and its bibliography are given, respectively. Worth also noticing that the Greek text is **hyperlinked** to Perseus⁴ in order to access its Greek word study tool and its online dictionaries through it.

3. CURRENT WORK AND CONCLUSIVE REMARKS

The mock-up is currently undergoing evaluation by a small sample of prospective users (university students attending a Greek Literature course) in view of in future system developments, to better focus on product requirements (a complete digital edition of Archilochus⁵). Evaluation follows a proper protocol to collect feedback from a small sample of learners. Students will be required to perform individually two distinct but similar tasks:

- a) using traditional bibliographic tools
- b) using mock-up support.

After adequate time, at the end of the test, each student, which was informed of the purpose of the experimentation and was shown how to use of the mock-up in advance, shall complete a questionnaire. The data collected through the questionnaires shall be analysed with due account of:

- the individual students' basic skills and preparation
- the quality of the results produced by the individual students in the given task (insufficient, satisfactory, good, excellent)

We expect to have evaluation results available by February and give directions to the design team about further product development.

³ The mock-up (<https://dSPACE-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/OPEN-83>) is implemented in PowerPoint and heavily exploits its animation function. The choice of overlapping tiles is to some extent unavoidable in order to concentrate in a few slides all available functions. The 'less' button usually allows to clear overlapping. A concrete action plan for its implementation should offer a new workbench and a data repository in which to insert text in a simple and intuitive way and visualize its encoding with specific TEI transcription; provide apparatus, literature and translation, link together primary sources.

⁴ Perseus (<http://www.perseus.tufts.edu/hopper/research>) is a digital library whose 'larger mission is to make the full record of humanity - linguistic sources, physical artifacts, historical spaces - as intellectually accessible as possible to every human being, regardless of linguistic or cultural background... within this larger mission, we focus on three categories of access:

Human readable information: digitized images of objects, places, inscriptions, and printed pages, geographic information, and other digital representations of objects and spaces.

Machine actionable knowledge: *catalogue records, encyclopedia articles, lexicon entries, and other structured information sources...*

Machine generated knowledge: *by analyzing existing information automated systems can produce new knowledge. Machine actionable knowledge allows, for example, us to look up a dictionary entry ... in a dictionary or to find pre-existing translations for a passage in Latin or Greek...*

⁵ Archilochus (in ancient Greek: Ἀρχίλοχος, Archilochos, Paros, circa 680 BC - about 645 BC) is considered the first great Greek lyric and the most famous giambographer.

Funzioni generali

Fr. 2 W.²
ἐν δορὶ μὲν μοι μᾶζα μεμαγμένη, ἐν δορὶ δ' οἶνος
Ἰσμαρικός· πίνω δ' ἐν δορὶ κεκλιμένος.

TRADUZIONE:
 Nella lancia è per me la focaccia impastata, nella lancia il vino
 d'Ismaro, nella lancia bevo reclinato.

Metrica: distico elegiaco

Testimoni: Athen. I 30f' Ἀρχιλόχος τὸν Νάζιον τῷ νέκταρι παραβάλλει (= φρ. 290 Ω.2): ὃς καὶ ποῦ φησιν· ἐν δορὶ — κεκλιμένος, Syn. Epist. 130 Hercher (= Garzya-Roques), II. 30ss. ἐγὼ δὲ ὑπὸ μεσοπυργίαι τεταγμένος ὑπνομαχῶ. ἐν δορὶ — κεκλιμένος, οὐκ οἶδ' εἰ μᾶλλον Ἀρχιλόχοι προσήκοντα ἦν ταῦτα εἰπεῖν. Vd. anche Suda I 645 A. s.v. Ἰσμαρικός οἶνος· Ἀρχιλόχος· ἐν δορὶ — κεκλιμένος. Ἰσμαρος δὲ πόλις, ἣ νῦν Μαρώνεια λεγομένη e v 441 A. s.v. ὑπνομαχῶ· Ἀριστοφάνης (loc. ignot.). ἐγὼ δ' ὑπὸ μεσοπυργίαι τεταγμένος ὑπνομαχῶ, ἀντὶ τοῦ ἐπαγοῦπνῶ. ἦν μὲν γὰρ καὶ ἄλλος ὁ Μάρκελλος ὑπνομαχεῖν πεφυκός (Dio Cass. LXXII 8,4). ἐν δορὶ — κεκλιμένος, οὐκ οἶδα δ' εἰ μᾶλλον Ἀρχιλόχοι προσήκον ἦν ταῦτα εἰπεῖν.

Grammatica: less
 δορὶ vv. 1s. è forma ionica

Traduzione

Bibliografia

Treebank

Introduzione generale

Metrica

Grammatica

Testimoni

Immagini

Loci similes

Luoghi e fatti

Altro

23/11/2017
homepage Next slide

homepage Previous slide
30: Isolata la proposta di Giangrande 1972, 37-40, che, sulla base del confronto con Anacr. fr. 43 (PMG 338), 7 P. (= 82,7 Gent.), interpreta ἐν δορὶ come ἐν τῷ κύρῳ, relegando il poeta giambico nel solito cliché del trickster emarginato e sovversivo. Ad un inverosimile jeu d'esprit che adombra in generale oggetti riconducibili ad una manifattura lignea pensa invece Pocock 1961, 179s. Less

Fr. 2 W.²
ἐν δορὶ μὲν μοι μᾶζα μεμαγμένη, ἐν δορὶ
δ' οἶνος Ἰσμαρικός· πίνω δ' ἐν δορὶ κεκλιμένος

Apparato, Varianti e Congetture:
 δ' ἐν δορὶ κεκλιμένος; δ' om. Syn. Less

Commento (Nicolosi, pp. 64-65):
 ἐν δορὶ κεκλιμένος: ammesso che la triplice occorrenza debba avere ogni volta lo stesso significato, diviene lecito domandarsi come sciogliere la contestata durezza del presunto nesso ἐν δορὶ κεκλιμένος More Less

62-64): Le principali ipotesi oscillano tra il significato tradizionale di 'lancia' (cf., ad es., Hom. *Il.* II 382, V 40 *al.*, VIII 95 *al.*, X 31 *al.*, XI 96 *al.*, XIII 130, XIV 494, XV 473 *al.*, XVI 114 *al.*, XVII 7 *al.*, XX 423 *al.*, XXI 17 *al.*, XXII 112 *al.*, XXIII 893 *al.* e *Od.* X 162 *al.*, XI 532, XIV 277, XIX 448)³¹ – talora anche attribuendo al nesso il valore metaforico 'in armi', 'sotto le armi'³² – e quello ugualmente attestato, pur se meno frequente, di 'legno della nave' e quindi, per metonimia, 'nave' (cf., ad es., Hom. *Il.* XV 410, XVII 744 e *Od.* IX 384)³³. Quest'esegesi, sostenuta – *praeunte* Davison 1960,3 – da Bruno Gentili (1965, 129-134) e poi ribadita dallo studioso in contributi successivi³⁴, muove da una peculiare lettura di uno dei due principali testimoni, Syn. *Epist.* 130 Hercher (= Garzya; Garzya-Roques)³⁵ ... More Less

fr.	W. ^{1,2}	A. ^{1,4}	Tard.	L.-B.	D. ^{1,3}	Cr. ⁵	B. ¹	B. ²	B. ³	B. ⁴	H.-C.	Schn.	Gaisf.	L.	Br.
2	2	2	2	7	2	3	2	2	3	2	2	2	45	56	-

Edizioni critiche:
 Br. = Brunck; L. = Liebel; Gaisf. = Gaisford; Schn. = Schneidewin; H.-C. = Hiller-Crusius; B.¹⁻⁴ = Bergk¹⁻⁴; Cr.⁵ = Crönert; D.¹⁻³ = Diehl¹⁻³; L.-B. = Lasserre-Bonnard; Tard. = Tarditi; A.¹⁻⁴ = Adrados¹⁻⁴; W.^{1,2} = West¹⁻²

23/11/2017
More Less

The figure above shows a screenshot of the textual analysis where the open windows show the comments about the ἐν δορὶ and the ἐν δορὶ κεκλιμένος sentences. On the latter, note 30 was open by clicking on the relevant button on the text. The former shows the first sentence of the comment only, clicking on the 'more' button switches to a separate slide where the complete text is given.

4. REFERENCES

- [1] A. Nicolosi, *Archiloco, Elegie*, Bologna: Patron, 2013.
- [2] G. Genette, *Palimpsestes. La Littérature au second degré*, Parigi: Edition du Seuil, 1982.
- [3] P. Robinson, «Where We Are With Electronic Scholarly Editions, And Where We Want To Be», Università di Monaco, Monaco D, 2004.
- [4] S. Ruecker, M. Radzikowska e S. Sinclair, *Visual Interface Design for Digital Cultural Heritage: a Guide to Rich-Prospect Browsing*, Oxford: Routledge, 2011.

- [5] E. G. Toms e H. L. O'Brien, «Understanding the Information», *Journal of Documentation*, 64, pp. 102-130, <http://dx.doi.org/10.1108/00220410810844178>, 2008.
- [6] N. Audenaert e R. Furuta, «What Humanists Want: How Scholars Use Source Materials», in *JCDL '10: Proceedings of the 10th Annual Joint Conference on Digital Libraries*, New York, NY, 2010.
- [7] C. Warwick, I. Galina, M. Terras, P. Huntington e N. Pappa, «The Master Builders: LAIRAH research on Good Practice in the Construction of Digital Humanities Projects», *Literary and Linguistic Computing*, 23, pp. 383-396, <http://discovery.ucl.ac.uk/13810/>, 2008.
- [8] D. Fogli e G. Guida, «A practical approach to the assessment of quality in use of corporate web sites», *The Journal of Systems and Software*, pp. 52-65, 2015.

A.R.C.A. Project: uno strumento per la creazione di web applications per la pubblicazione degli scavi archeologici

Irene Carpanese¹, Guido Lucci Baldassari²

¹ Università degli Studi di Padova, PhD Student - irene.carpa@gmail.com

² Software Engineer and Independent Researcher - guido.lucci.baldassari@gmail.com

ABSTRACT

Il progetto A.R.C.A. fa parte di un lavoro di dottorato tutt'ora in corso presso il Dipartimento dei Beni Culturali dell'Università degli Studi di Padova (tutor prof.ssa Maria Stella Busana, co-tutor prof. Jacopo Bonetto e prof. Nicola Orio).

Liberamente ispirato al pacchetto CMS Museo&Web (sviluppato all'interno del progetto Minerva), che funge da strumento per la creazione guidata di siti web e per la pubblicazioni online di collezioni in ambito museale, archivistico, bibliotecario, anche A.R.C.A. ha come obiettivo finale la creazione di un software modulare e flessibile, ma rivolto esclusivamente alla raccolta di dati e metadati specifici provenienti da indagini e scavi archeologici. Più che di un CMS si può parlare in questo caso di una web application *ad hoc*, configurabile e riutilizzabile, basata su framework, ossia librerie software altamente specializzate, che ben si prestano a gestire in modo efficace l'eterogeneità del dato. Basata sul pacchetto open source MEAN, che offre la massima flessibilità nell'impostazione delle strutture, i prodotti che saranno creati a partire da A.R.C.A. potranno fungere da semplici repository online, da portali web contenenti informazioni e dati archeologici strutturati in una banca dati o divenire delle vere e proprie pubblicazioni online parallele (e interconnesse) alle pubblicazioni cartacee di un progetto o di una ricerca archeologica.

Il prodotto sarà testato in fase di sviluppo con l'inserimento dei dati provenienti da alcuni scavi archeologici e sarà infine sottoposto ad una serie di "test d'utente", per verificarne la navigabilità e il corretto funzionamento.

PAROLE CHIAVE

Web applications, Scavi archeologici, Pubblicazioni online

1. IL PROGETTO

A.R.C.A. è l'acronimo di Archiving Research Communication in Archaeology ed è il naturale seguito di un lavoro annuale finanziato dal Fondo Sociale Europeo [1], in cui è stato sviluppato un sito web, a carattere divulgativo, inerente alle ricerche multidisciplinari decennali [2] nella Tenuta di Ca' Tron (TV), dal titolo "Navigare nel passato tra la Marca trevigiana e la Laguna di Venezia" (tutor prof.ssa Maria Stella Busana).

Al termine di questo progetto si è evidenziata la necessità di progettare una web application prototipale, semplice e minimale, specifica per realtà di scavo, studio e ricerca archeologiche, completa dal punto di vista divulgativo, riutilizzabile e gestibile dai fruitori, che permettesse di creare veri e propri portali in cui inserire i dati e renderli direttamente consultabili, per standardizzare ed ottimizzare ricerca e divulgazione. Alcune delle linee guida da seguire per lo sviluppo dell'applicazione sono state estrapolate dalle risposte raccolte da un'iniziale indagine online (dicembre 2015-febbraio 2016), che consisteva in un questionario composto da 50 domande inerenti la diffusione e l'utilizzo del "web culturale", che ha previsto il coinvolgimento di un centinaio di possibili futuri fruitori del prodotto[3].

Parallelamente è stata avviata un'approfondita ricerca inerente le numerose soluzioni informatiche disponibili online che meglio rispondessero a determinate necessità e si potessero adattare alla tipologia di dati da trattare; il risultato finale è stata la scelta, a livello implementativo, del pacchetto MEAN[4], composto dalla banca dati Mongo DB[5] di tipo no-sql / non relazionale basata sulla tecnologia JSON, da due framework uno di back end e uno di front end, Express e Angular2/4 e NodeJS6, che funge da strato di comunicazione tra il sistema operativo e lo strato applicativo sovrastante. L'intero sistema software ha il vantaggio di essere scritto in unico linguaggio ECMA Script 5/6 (Javascript), che facilita la comunicazione tra i vari componenti ed eventuali customizzazioni da parte dei futuri implementatori[6].

La possibile interoperabilità tra i sistemi che saranno creati attraverso il pacchetto A.R.C.A. si deve ad alcune caratteristiche peculiari del software stesso, estensibile e modulare, ma a cui si è cercato di imprimere una certa standardizzazione semantica, oltre che alla scalabilità intrinseca della banca dati. Il prodotto finale che verrà messo a disposizione online potrà essere dunque riutilizzabile per progetti differenti, anche diversi tra loro sul piano cronologico o per localizzazione geografica, in quanto sviluppato con una struttura interpretabile "a nodi" o ad oggetti, in cui sono presenti nodi geografici (di cui fa parte il nodo di partenza o "nodo radice"), "nodi dati" e "nodi applicativi", i quali racchiudono quelle meta-informazioni necessarie al funzionamento del sistema stesso. Questa impostazione è stata data con lo scopo di fornire al futuro/i amministratore/i un prodotto modellabile secondo le proprie esigenze, senza imporre uno schema preimpostato che potrebbe tradire la naturale unicità di uno scavo o le necessità di una pubblicazione online, ma impostando comunque delle terminologie ampiamente accettate.

Le uniche accortezze richieste in fase di utilizzo/sviluppo sono quelle di mantenere inalterati i rapporti tra i nodi: il nodo cosiddetto "padre" dovrà rispondere sempre al rapporto 1:n con il nodo/i nodi figlio/i sottostante/i, ma il numero di nodi e la loro denominazione potrà mutare a seconda delle esigenze del singolo prodotto, così come è modificabile la struttura dei nodi stessi.

Lo studio inerente l'interoperabilità dei sistemi ha previsto anche un allineamento terminologico (e in parte strutturale) con le principali ontologie dei beni culturali; nello specifico sono state riprese e adottate le voci stilate dall'ICCD (Istituto Centrale per il Catalogo e la Documentazione) per i nodi geografici (Ricognizioni Archeologiche, Siti) e per i nodi oggetto (Unità Stratigrafiche, Unità Stratigrafiche Murario, Unità Stratigrafiche di Rivestimento, Reperti Archeologici, Beni Numismatici, Reperti Antropologici), con l'inserimento del relativo codice come riferimento all'interno della banca dati di A.R.C.A. per future implementazioni. Attraverso questo lavoro di allineamento potenzialmente sarà possibile far comunicare, in futuro, i dataset di A.R.C.A. con i principali cataloghi nazionali ed europei (come Europeana, ARIADNE) e predisporre anche un eventuale inserimento delle informazioni.

Il prodotto sarà testato con i dati di almeno due scavi, in modo da poterne verificare le funzionalità, le prestazioni e la sua effettiva flessibilità: il primo caso studio individuato è il sito M di Ca' Tron, uno scavo archeologico rurale di epoca romana situato in prossimità della laguna di Venezia, interessato da una serie di studi a carattere multidisciplinare, che quindi ben si presta, per la tipologia di dati molto eterogenei da rappresentare, a diventare un caso studio particolarmente articolato.

Un altro test significativo sarebbe quello di utilizzare il sistema come canale alternativo per la pubblicazione online di dati di cui già esiste (o è prevista) un'edizione tradizionale; si potrebbe in questo modo osservare le potenzialità di A.R.C.A. come appendice di approfondimento interattiva e facilmente consultabile online, da parte di un pubblico di fruitori più o meno specialistico.

2. BIBLIOGRAFIA

- [1] Busana M.S., Carpanese I., Orio N. 2016, *La ricerca a portata di click. Database relazionali e siti web: "contenitori" e "visualizzatori" per i dati archeologici*, P. Basso, A. Caravale, P. Grossi (a cura di), Firenze, Edizioni all'Insegna del Giglio, in «Archeologia e Calcolatori», Suppl. 8, 246-253
- [2] Busana M.S., Ghedini F. 2004: M.S. Busana, F. Ghedini (a cura di), *La via Annia e le sue infrastrutture*, Atti delle Giornate di Studio, Ca' Tron di Roncade, 6-7 novembre 2003, Treviso
- [3] Carpanese I. 2017, "Let's do it together!" *Indagare, progettare, sviluppare la web archaeology con l'aiuto di un questionario*, Firenze, Edizioni all'Insegna del Giglio, in «Archeologia e Calcolatori» 28, 271-290
- [4] Siti ufficiali: <https://expressjs.com/>; <https://angular.io/>; <https://nodejs.org/it/>
- [5] Chodorow K. 2013, *Second Edition MongoDB: the definitive guide*, published by O'Reilly Media, Inc., Sebastopol, CA. ISBN: 978-1-449-34468-9
- [6] Sito ufficiale: <https://www.ecma-international.org/ecma-262/6.0/>

Bibliotheca Perspectivae

Stefano Casati¹, Andrea De Calisti²

¹ Museo Galileo - s.casati@museogalileo.it

² GAP - a.decalisti@gap-italy.it

KEYWORDS

Convergenza/divergenza tra archivio, biblioteca, museo – Valorizzazione.

Sul sito Web del Museo Galileo¹ è consultabile *Bibliotheca Perspectivae*², una collezione digitale dedicata alla prospettiva, ideata per soddisfare le esigenze di un settore storiografico a lungo considerato come un'appendice della storia dell'arte o un 'corollario' di storia della scienza ma oggi sempre più caratterizzato da una precisa autonomia disciplinare. La storia della prospettiva o, più in generale della rappresentazione, infatti vanta ormai una quantità di titoli bibliografici che documentano chiaramente la specificità del problema nel panorama degli studi critici sulle arti e sulle scienze dell'età moderna.

Raggiunto l'intento originario, fornire agli studiosi una 'biblioteca' sulla rappresentazione che raccogliesse in prima istanza le fonti trattatistiche del periodo compreso tra il XV e il XVIII secolo, *Bibliotheca Perspectivae* è stata successivamente arricchita con risorse digitali non legate a opere a stampa e manoscritti. In particolare sono state aggiunte due sezioni: *Laboratorio Didattico* e *Iconografia*. *Laboratorio Didattico* rappresenta uno spazio didattico virtuale costituito prevalentemente da animazioni, filmati e slide-shows che presentano il contesto storico, i personaggi, gli strumenti, gli esperimenti e i luoghi pertinenti alle argomentazioni trattate. La sezione *Iconografia* permette invece, tramite una lista gerarchica di termini, la consultazione di un nutrito e selezionato numero illustrazioni tratte da testi e manoscritti. Ogni illustrazione è corredata da informazioni descrittive e di approfondimento curate da studiosi.

Così strutturata *Bibliotheca Perspectivae* si limita però a proporre 'isole informative', estremamente ricche di contenuti ma confinate nel loro dominio conoscitivo, in parte dischiuse dalle possibilità offerte dal sistema di ricerche della Biblioteca digitale dell'istituto³. Analogamente a quanto si è fatto e si sta facendo sui trattati di architettura, gli scritti d'arte e le fonti di storia della scienza, con la nuova versione di *Bibliotheca Perspectivae* intendiamo invece fornire agli studiosi una 'biblioteca' sulla rappresentazione che raccolga un patrimonio documentario eterogeneo, consultabile attraverso le molteplici possibilità offerte dalla costituzione di un archivio dati integrato.

Obiettivo dell'intervento è presentare la fase sperimentale della nuova versione della *Bibliotheca Perspectivae* che attraverso la scelta tecnologica di un data base no SQL si propone di migliorare le possibilità di navigazione e di recupero delle informazioni offerte dalla versione precedente.

Lo scopo che ci proponiamo consiste nel costruire una Biblioteca digitale tematica di concezione moderna, impostata essenzialmente nella costituzione di un archivio dati integrato che renda possibile l'esplorazione articolata e profonda di 'universi conoscitivi'.

L'aspetto innovativo della nuova versione della *Bibliotheca Perspectivae* risiede non solo nell'approccio multidisciplinare delle descrizioni e nella facilità di consultazione di risorse eterogenee ma anche nella proposizione di percorsi esplorativi basati sulle relazioni tra entità. Al tradizionale approccio di interrogazione del data base per filtri e query si affianca la possibilità per l'utente web di navigare in modo semplice e intuitivo la fitta rete di relazioni tra risorse.

La portata innovativa della nuova *Bibliotheca Perspectivae* è supportata da scelte tecnologiche innovative, in particolare dalla sperimentazione di un data base no SQL come supporto della struttura dati.

Dopo una fase di analisi preliminare la scelta è ricaduta sulla tecnologia a grafi di neo4j.⁴ Il data base a grafi è per sua natura uno strumento particolarmente adatto alla gestione di dati complessi e collegati tra di loro attraverso un insieme di nodi e relazioni, un sistema di rappresentazione della conoscenza flessibile e modulabile in grado di rispondere efficacemente alle esigenze di gestione delle risorse di *Bibliotheca Perspectivae*. Oltre a questo la logica del grafo è nativamente predisposta verso un'esportazione e comunque una comunanza di intenti con il modello RDF seppur con maggiori possibilità di gestire relazioni più articolate.

L'intervento intende riportare i risultati ottenuti nella prima fase del progetto, che ha condotto alla realizzazione di un prototipo di portale ed ha confermato l'efficacia della scelta del data base a grafi, grazie alla maggiore flessibilità offerta all'architettura e alla gestione dei dati e delle relazioni tra essi, rispetto alla scelta più comune del data base relazionale. La sperimentazione ha inoltre messo in evidenza la naturale apertura alle logiche del web semantico di una banca dati così fortemente orientata all'usabilità e alle relazioni tra risorse.

¹ <https://www.museogalileo.it>

² *Bibliotheca Perspectivae*, progetto curato da F. Camerata e S. Casati, sostenuto con il contributo di: Ministero dell'Istruzione, dell'Università e della Ricerca e Università degli studi di Firenze - Dipartimento di Filosofia.

<http://www.museogalileo.it/esplora/biblioteche/bibliotecadigitalenew/collezionidigitalinew/iperspectivaebd.htm>.

³ <http://www.museogalileo.it/esplora/biblioteche/biblioteca.html> e

<http://www.museogalileo.it/esplora/biblioteche/bibliotecadigitalenew.html>

⁴ <https://neo4j.com/>

BIBLIOGRAFIA

- [1] Brath R., Jonker, D., 2015. *Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data*, New York, WILEY.
- [2] Butini, F., Casati, S., 2014. *Some remarks about Museo Galileo's digital collections*, in *Collaborative research practices and shared infrastructures for humanities computing: 2nd AIUCD Annual Conference, AIUCD 2013, Padua, Italy, 11-12 December 2013*, proceedings of revised papers / Maristella Agosti and Francesca Tomasi (eds), Padova: CLEUP, 2014. p. 143-148.
- [3] Casati, S., 2015. *La Biblioteca digitale del Museo Galileo*. In: *Biblioteche oggi* (genn.-febr. 2015) p. 45-51.
- [4] Robinson, I, Webber, J., 2013. *Graph Databases: New Opportunities for Connected Data*. Sebastopol, O'Reilly.
- [5] Trudeau, R.J., 1994. *Introduction to graph theory*. New York, Dover Publications.

CEED: a Cooperative Web-Based Editor for Critical Editions

Simone Zenzaro¹, Daniele Marotta², Amos Bertolacci³

¹ Scuola Normale Superiore, Pisa, Italy - simone.zenzaro@sns.it

² Scuola Normale Superiore, Pisa, Italy - daniele.marotta@sns.it

³ Scuola Normale Superiore, Pisa, Italy - amos.bertolacci@sns.it

ABSTRACT

In this paper some of the results of the ERC Advanced Project PhiBor project are presented. PhiBor's main objectives are to produce a comprehensive inventory of manuscripts of Avicenna's *Ilāhiyyāt*, his main work on metaphysics, and to create on their basis a TEI P5 compliant digital critical edition. In order to permit scholars with little or no technical skills to produce a digital edition, we are developing CEED: a web-based critical edition editor. The main goal of CEED is to provide a user-friendly editor that lets scholars perform their work without requiring the acquisition of technical skills. The editor produces TEI P5 digital editions by leveraging the well known word processor approach. CEED environment is also meant to ease cooperation between scholars, since a project can be shared by several different contributors.

KEYWORDS

TEI, XML, Editor, Digital Scholarly Edition, Critical Edition

1. INTRODUCTION

Research in digital humanities is having an increasing impact on the preservation, analysis, and diffusion of cultural heritage, and scholars are showing a growing awareness of how the application of computer science to humanities improves their research workflow.

In the context of the production of critical editions, several definitions, models, and methodologies have been proposed to express the need of overcoming the *page paradigm* by means of a proper *digital paradigm* [8].

The freedom from page constraints permits to explore new approaches. On the one hand, such new approaches have produced digital editions of great scientific value; on the other hand, most of the results cannot yet be extended to a wider scholarly community due to the technical skills still required [9].

In this sense, the Text Encoding Initiative consortium, with its XML TEI standard [12], provides scholars with a large set of guidelines in the form of XML schemas and documentation and enables them to produce almost every kind of digital texts. However, this powerful representation comes with a huge initial requirement of technical skills that prevents many scholars from adopting it [3]. Another problem of XML formats is the lack of proper visualization tools. The EVT tool designed by Roberto Rosselli Del Turco [11] is filling this gap, making pioneering steps towards the fruition of texts in TEI format.

In order to enable a growing number of scholars to benefit the advantages of digital critical editions, tools with a smooth learning curve are decisive. Desirable features of tools for the production of critical editions are: (1) low technical prerequisites for users; (2) meaningful graphical visualization of the work advancements; (3) support for collaboration between scholars; (4) reliability of data representation based on standards; (5) ability to produce both digital and paper versions of the work.

Classical Text Editor (CTE) [4] and the Oxygen XML editor [10] are two major tools that can be used to produce digital critical editions. CTE focuses on the production of camera-ready copy of a text via a What You See Is What You Get (WYSIWYG) word processor. Oxygen provides a comprehensive visual XML editor. Both CTE and Oxygen, however, present downsides: CTE does not allow cooperation by different editors on the same work, and focuses on printed editions; Oxygen requires high initial technical skills.

To overcome these difficulties, we propose Critical Edition EDitor (CEED): a cooperative web-based editor for critical editions.

2. A USER-FRIENDLY CRITICAL EDITION EDITOR

CEED is being designed and developed as part of the PhiBor project [7] hosted by the Scuola Normale Superiore under the direction of Prof. Amos Bertolacci and founded by the European Research Council. PhiBor main objectives are to produce a comprehensive inventory of manuscripts of Avicenna's *Ilāhiyyāt*, his main metaphysical work with a massive posterity, and to create on their basis a TEI P5 compliant digital critical edition.

PhiBor research has already identified about 250 manuscripts and acquired copies of almost all of them. The transcription of each of these numerous witnesses during the project timespan is, of course, unfeasible; therefore it has been decided to produce a critical edition by means of an editor that would allow to record textual variations in the apparatus, and to add notes and references.

CEED provides a web application environment for the creation of digital critical editions that requires very little preliminary technical skills. ("digital critical editions" are meant here as "prêt-à-porter" editions [9], without disregarding more peculiar types of editions). A *project* in CEED identifies a critical edition and the information related to it (e.g. witnesses, manuscript images, bibliography, transcriptions). The creator of a project becomes its owner and may decide to involve other scholars in the edition permitting collaborative work. The collaboration between scholars is supported by

access management for users that share a project. Concurrent editing of the same resources has been planned but is not currently available. Being a web application, initial setup and installation are not required; it is possible to work and cooperate remotely, and the content of a project is easily retrievable. The editing process combines the classic word processors approach with contextual graphical user interfaces (GUIs). Although CEED produces TEI P5 editions, it requires no knowledge of the TEI standard, minimizing times and costs of user training. Texts are organized in *documents* that identify different transcriptions of the original chosen text and translations of it. Each document can be separated into meaningful working *units*. A *unit* presents, side by side, the portion of text under consideration and manuscripts images for its study and collation (see Figure 1).

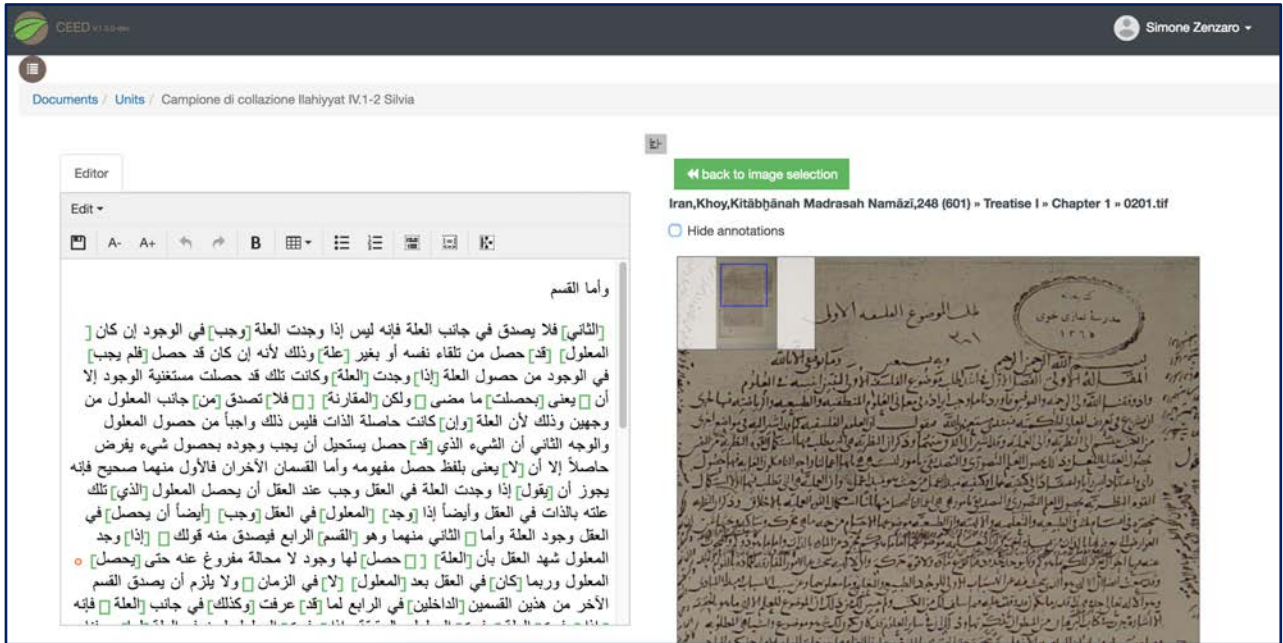


Figure 1. Working on the text

The basic action on texts is to highlight and make easily visible segments that carry interesting textual phenomena. Text phenomena managed by CEED are readings, annotations, editorial interventions, and structural information (page numbers, headings, chapters, treatises, etc.).

With regard to textual variations, CEED currently can manage both simple manuscript variant readings and variant readings with scribal interventions, like cancellations, corrections etc. (see Figure 2). Besides these two types of “objective” readings, CEED allows to record also “subjective” editorial interventions into the text by the editor(s). In the former context, CEED supports the description of different hands, causes of variations, and languages. Within the readings with scribal interventions, also the time sequence of the interventions, i.e. their different phases, is recorded (e.g. a correction that follows a cancellation); if any of these phases is attested also by another witness, it is possible to record the concordance. A CEED module is responsible for the definition and management of witnesses (manuscripts and printed books) to guarantee uniform reference to them. An exporter module produces the TEI P5 version of a project, which can serve for preliminary inspection or for the issue of the final digital edition.

From the software engineering point of view, CEED is being developed as an Angular application [1] written in Typescript [13]. The application communicates with a REST service that unifies the internal JSON [5] data model and the related operations. The REST service manages the manuscript images and the critical edition data. The images are stored inside an International Image Interoperability Framework (IIIF) [6] compliant image server. A BaseX database [2] stores the critical edition data.

3. FUTURE DEVELOPMENT

CEED is currently under development. Some core features are already available (project and work groups management; readings and text structure annotations; multilingual support; notes), while others are in a prototypal state.

The modular structure of CEED permits to incrementally add features in order to provide new functionalities. In particular, we are working on: a module for the *visualization preview* of the edition by integrating EVT; a *dashboard* that summarizes statistical data of text phenomena and users activity; *bibliography* and *authors* references; and a visual *exporter* to TEI P5. Although the constant information exchange with the PhiBor working group gives valuable direction of improvement, a proper assessment of usability will be planned as soon as we reach the completion of most of the core features.

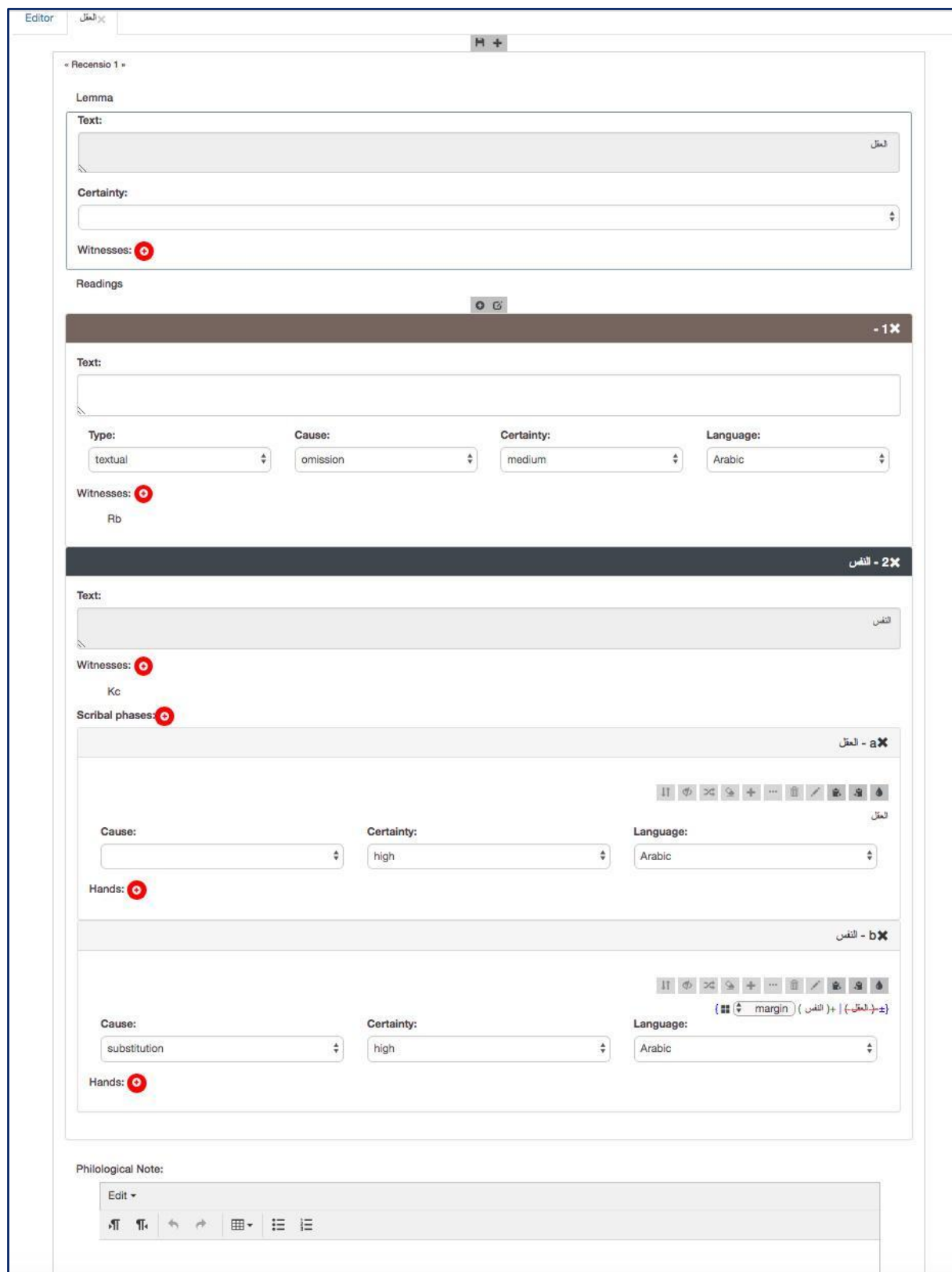


Figure 2. Readings definition

4. REFERENCES

- [1] Angular. Framework. <https://angular.io>.
- [2] BaseX: the XML Database. <http://basex.org>.
- [3] Burghart, M., Rehbein, M. 2012. "The Present and Future of the TEI Community for Manuscript Encoding", *Journal of the Text Encoding Initiative*, Issue 2. <http://jtei.revues.org/372>.
- [4] Hagel, S. 1997-2017. "Classical Text Editor. Software". [9.3]. <http://cte.oeaw.ac.at>.
- [5] JSON (JavaScript Object Notation). Data-interchange format. <http://www.json.org>.

- [6] International Image Interoperability Framework (IIIF). <http://iiif.io>.
- [7] PhiBor 2014-2019. *Philosophy on the Border of Civilizations and Intellectual Endeavours: Towards a Critical Edition of the Metaphysics (Ilāhiyyāt of Kitāb al-Šifāʾ) of Avicenna (Ibn Sīnā)*. ERC Advanced Grant (n. 339621). <http://www.avicennaproject.eu>.
- [8] Pierazzo, E. 2014. *Digital Scholarly Editing: Theories, Models and Methods*. <https://hal.inria.fr/hal-01182162/document>.
- [9] Pierazzo, E. 2016. “Quale futuro per le edizioni digitali? Dall’haute couture al prêt-à-porter”. In *AIUCD2016 - Book of Abstracts*, a cura di F. Boschetti, 53-54. Firenze: AIUCD. In: Quaderni di Umanistica Digitale. ISBN: 978-88-942535-0-4. DOI: 10.6092/unibo/amsacta/5559.
- [10] SyncRO Soft SR 2002-2017. “Oxygen. Software”. [19]. https://www.oxygenxml.com/xml_editor.html.
- [11] Team EVT 2017. “Edition Visualization Technology. Software”. [2.0]. <http://evt.labcd.unipi.it>.
- [12] TEI Consortium, eds 2017. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [3.2.0]. [10-07-2017]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5>.
- [13] Typescript. <https://www.typescriptlang.org>.

Close and Computational Reading of Galileo Galilei and Alessandro Piccolomini

Crystal Hall

Bowdoin College, chall@bowdoin.edu

1. A HYBRID APPROACH TO AN UNEXPECTED QUESTION

Using a computational method of analysis that facilitates both close and distant reading of texts, this paper argues that fundamental elements of stylometry can indicate meaningful patterns of literary expression across genre and time periods visible at the level of representative passages. In the process of stylometric analysis of Galileo Galilei's treatise on comets, *Il Saggiatore* (Rome, 1623) in a small corpus of 63 early modern Italian texts, a surprising result recurred: while the analysis was designed to test a hypothesis about the *Saggiatore's* similarity to Ludovico Ariosto's epic poem *Orlando furioso* (Ferrara, 1532), a different text had as consistently high textual similarity to Galileo's epistolary treatise as the *Furioso*. In terms of word usage frequency, amount of common vocabulary similarity, and quantity of shared bigrams, trigrams, fourgrams, and fivegrams, Alessandro Piccolomini's *Della istituzione di tutta la vita dell'uomo nato mobile, e in città libera* (Venice, 1542) is the work, in the corpus analyzed, with the most overall similarity to the *Saggiatore*. Yet, Piccolomini's views on civic duty and civil life would seem to have very little to offer for Galileo's debate about the nature of comets nearly a century later. In fact, Piccolomini typically only appears in scholarship related to precursors of Galileo's dialogues, with no implication that the Florentine natural philosopher had read the Sienese humanist's works. (Cox, 1992)

To what relationship between these two texts, then, do the quantitative results gesture? Answering this question probes the consequences of accumulating a multitude of "tiny strokes" identified through computational text analysis (Burrows, 2002). In addition, it is an opportunity to investigate the expressive instantiations of the textual features most often used in stylometric analysis and authorship attribution (Koppel et al., 2009). Rather than taking the step of statistical inference of authorship, the analysis considers observations and measurements, hopefully avoiding the persistent distrust of such models in literary analysis and the need for an apology for applying a quantitative approach at all (Hoover, 2008; Underwood, 2015). Moreover, such an investigation requires a method that can connect the broad, distant reading offered by high-dimensional space and dendrograms to the deep, close reading of figurative language and rhetorical devices.

Such a complementary approach strengthens the argument to be made about early modern Italian cultural expression and compensates for the possibility that the small corpus resulted in false positive similarities by default. Out of attentiveness to the constraints of that corpus, the analysis consistently considers the similarity measures of other texts in order to provide context for the results. When combined with a close reading of the passages that score highest in stylistic similarity and ngram matches, the method ultimately reveals the patterns of humanistic rhetorical moves that both Galileo and Piccolomini expressed with the same grammatical structures and vocabulary.

2. THE CORPUS

The texts used for this project are diplomatic digital editions of early modern Italian printed volumes made available by the Biblioteca Italiana platform (<http://www.bibliotecaitaliana.it/>) and Early English Books Online (EEBO). The corpus has been expanded by transcribing additional titles in house. The texts represent literary, historical, and philosophical genres published from 1516 through 1675.¹ They range in length from a short libretto (1,300 words) to long treatises (150,000 words) and lengthy narrative poetry (max. 325,000 words).

3. METHODS OF COMPUTATIONAL ANALYSIS

In order to better understand the extent of the stylistic similarity between Galileo's vernacular texts and Piccolomini's, this research employs several text mining methods using R for stylometric analysis and string matching. Through hierarchical clustering of passages based on Euclidean distance between relative term frequencies for the 100 most frequent words (MFWs), I identified portions of Galileo's text that were more similar to (ie: in the smallest cluster with) excerpts from Piccolomini's treatises and dialogue (adapted from Jockers, 2014). Features were limited to the 100 MFWs in order to prevent topic-specific vocabulary from creating a false difference between the texts. The hypothesis is that Piccolomini and Galileo can use similar styles to debate, contextualize, analyze, or present their views on the seemingly divergent themes of civic responsibility and the nature of comets. To further refine the precise identification of similar passages, while maintaining textual integrity for close reading, works in the corpus were divided into smaller documents via a rolling window of consecutive text (not a random sample). Also, consecutive text was more relevant for this question, since the goal was not to refine the accuracy of authorship attribution (as in Eder, 2015), but to explore the passages that resist such attribution through their variance from the rest of the passages written by the same author. The similarity analysis was conducted for rolling windows of five different lengths, with an overlap of 10% of that length: 100, 200, 500, 1000, and 2000 words.

¹ A current list of titles can be found at: <http://research.bowdoin.edu/galileos-library/library-info/piccolomini/piccolomini-texts/>

String matching was conducted in a very elementary fashion in R by identifying exact, not fuzzy, matches of character strings across full texts. As mentioned, this project considered ngrams from unigram through fivegram. The ngrams were matched after the text had been vectorized based on space, forced to lowercase, the letter "v" converted to "u" to standardized early modern printing practices, the conversion of accented vowels to non-accented (again for standardization), and all numbers (Arabic and Roman) removed.

4. SELECTED BIBLIOGRAPHY

- [1] Burrows, John. 2002. "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17 (3): 267-87.
- [2] Cox, Virginia. 2008. *The Renaissance Dialogue: Literary dialogue in its social and political contexts, Castiglione to Galileo*. Cambridge: Cambridge University Press.
- [3] Eder, Maciej. 2015. "Does size matter? Authorship attribution, small samples, big problem." *Literary and Linguistic Computing* 30 (2): 167-82. <https://doi.org/10.1093/lc/fqt066>
- [4] Hoover, David. 2008. "Quantitative Analysis and Literary Studies." In *A Companion to Digital Literary Studies*, edited by Susan Schreibman and Ray Siemens, Section 28. Oxford: Blackwell. <http://www.digitalhumanities.org/companionDLS/>
- [5] Jockers, Matthew. 2014. *Text Analysis with R for Students of Literature*. New York: Springer.
- [6] Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. 2009. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and Technology* 60 (1): 9-26.
- [7] Underwood, Ted. 2015. "The literary uses of high-dimensional space." *Big Data & Society* July-December: 1-6. DOI: 10.1177/2053951715602494

Dalle Stampe Antiche al Museo Virtuale.

Best practice nell'acquisizione di competenze digitali per l'archiviazione, conservazione e visualizzazione di materiale antico a stampa: il caso di un percorso di alternanza scuola-lavoro per gli studenti di un Liceo classico.

Barbara Balbi¹, Francesca De Ruvo², Alessandra Monica Mazzaro³,
Alfredo Apicella⁴, Roberto Montanari⁵

¹ Centro di Ricerca Scienza Nuova, Università degli studi Suor Orsola Benincasa-barbara.balbi@centrosenzanuova.it

² Fondazione Pagliara, Università degli studi Suor Orsola Benincasa-fondazionepagliara.unisob@gmail.com

³ Fondazione Pagliara, Università degli studi Suor Orsola Benincasa-fondazionepagliara.unisob@gmail.com

⁴ Università degli Studi di Napoli Federico II, Scuola politecnica delle Scienze di base alfredo.apicella1985@gmail.com

⁵ Centro di Ricerca Scienza Nuova, Università degli studi Suor Orsola Benincasa-
roberto.montanari@centrosenzanuova.it

ABSTRACT

Il trasferimento di competenze digitali nella formazione primaria e secondaria è tema assai discusso negli ultimi anni, ovviamente da quando le scienze umane e le cosiddette ICT non possono non dirsi interconnesse.

Sebbene non a caso si parli di umanisti digitali, nei licei classici l'informatica non è prevista dalla programmazione curricolare. D'altronde le competenze digitali sono lo strumento attraverso cui i domini degli umanisti vengono oggi sistemati, conservati e resi accessibili, e la nuova auspicata figura è quella dell'ingegnere integrale, educato cioè alle e dalle scienze umane.

Il percorso di alternanza scuola-lavoro che descriviamo, nasce su queste premesse con l'ambizione di sviluppare, attorno a un tema propriamente umanistico, quale quello della valorizzazione del patrimonio grafico, un programma centrato sull'acquisizione di "competenze digitali" per gli allievi di un secondo liceo classico statale. Il percorso, mirava a riferire un'idea alternativa dei domini scientifici cui gli studi umanistici sono destinati: quelli legati alla valorizzazione del patrimonio culturale con particolare riferimento al mondo dei virtual museum¹, ivi comprese le possibilità collegate ai serious game².

Materiali del percorso sono le stampe di un fondo inedito della Fondazione Pagliara dell'Università Suor Orsola Benincasa, aventi a soggetto la vita pubblica di Napoleone Bonaparte.

Il progetto ha condotto due fasi parallele riguardanti l'acquisizione delle competenze di base per l'utilizzo dei software e dei dispositivi per la scansione e catalogazione digitale e quelle per la creazione, gestione e visualizzazione di modelli tridimensionali. Il prodotto finale è una galleria virtuale tridimensionale, utile alla diffusione del progetto e alla costruzione di un'idea di archiviazione cross-mediale del patrimonio cartaceo.

KEYWORDS

Digital Humanities, Didattica, Uso pubblico, divulgazione e mediazione didattica delle culture umanistiche nell'era digitale, Virtual Cultural Heritage.

1. INTRODUZIONE

In tema di Digital Humanities ci occupiamo in questa sede della formazione dei giovani umanisti: il programma MIUR "Scuola Digitale" pur dando seguito a interessanti progetti, vede il ruolo degli studenti dei percorsi umanistici poco spinto verso la programmazione informatica. Il percorso di alternanza scuola lavoro che presentiamo ha toccato alcuni dei temi che attualmente definiscono il perimetro delle Digital Humanities in un'ottica eminentemente didattica: la formazione dei cosiddetti umanisti digitali, e quindi i percorsi di collegamento tra Università e istituti di formazione secondaria che integrino scienze umane e tecnologie, e la valorizzazione del patrimonio culturale, in questo caso cartaceo. Il percorso ha rappresentato un momento di appropriazione di una diversa prospettiva d'uso degli strumenti digitali per i classicisti. L'approccio laboratoriale ha consentito la pratica di ogni livello di costruzione digitale dalla scansione delle antiche stampe alla modellazione tridimensionale dell'ambiente virtuale che le avrebbe accolte.

¹ La definizione di Virtual Museum è in costante cambiamento parallelamente all'aggiornamento dei linguaggi digitali con cui esso viene proposto, ma durante la nostra esperienza ci siamo riferiti spesso al Virtual Museum come a un luogo di informazioni dinamico e multimediale in cui l'autenticità dell'oggetto è sostituita dall'autenticità dell'informazione, ponendo l'attenzione sulla sua evidente maggiore accessibilità e alla possibilità di rappresentazione e valorizzazione di patrimoni impossibili da mostrare, così come descritto in Rayward, Boyd, W.; Twidale, Michael, B., 1999, "From docent to cyberdocent: education and guidance in the virtual museum".

² Con serious game si intende genericamente ogni gioco digitale che abbia una declinazione didattica educativa, ma recentemente l'accezione è ampliata comprendendo elementi di riflessione su temi di grande impatto sociale o di complessa definizione. Per una trattazione più specifica rinviamo a Ritterfeld, U., Cody, M., & Vorderer, P. (Eds.). (2009). *Serious games: Mechanisms and effects*. Routledge.

2. MATERIALI

Le Stampe della Fondazione Pagliara.

L'Università degli Studi Suor Orsola Benincasa di Napoli ospita nei suoi ambienti la Raccolta d'Arte appartenuta a Rocco Pagliara (1856 –1914) che vide il suo ingresso nella cittadella di Suor Orsola nel 1947.

Uno dei settori più importanti dell'eterogenea collezione è il Gabinetto della Grafica che custodisce un cospicuo corpus di circa 20.000 opere a stampa³.

3. SVILUPPO DEL PROGETTO

a. Individuazione del materiale a stampa

Il lavoro di inventariazione, catalogazione e digitalizzazione è stato condotto dagli studenti del secondo anno del Liceo Vittorio Emanuele e ha interessato la serie di acquaforti su fondino realizzata da Giuseppe Benaglia, Michele Bisi, Giuseppe Longhi e Francesco Rosaspina che traduce il ciclo di trentanove dipinti a tempera su tela, del pittore milanese Andrea Appiani per la Sala delle Cariatidi del Palazzo Reale di Milano. Essa rappresenta un'importante documentazione iconografica poiché l'opera originaria venne distrutta durante i bombardamenti della Seconda Guerra Mondiale. Le stampe narrano la vita e le gesta di Napoleone I in una sequenza cronologica in cui gli episodi sono talvolta risolti attraverso raffigurazioni allegoriche. La serie, una seconda edizione rilegata dell'album *Fastes De Napoleon I er*, pubblicato a Parigi intorno al 1860 dall'editore F. Chardon in Rue Hautefeuille, fu dedicata a Napoleone III allo scopo di legittimare il suo potere attraverso il rimando all'Impero napoleonico.



Uno studente durante la fase di individuazione del materiale da digitalizzare.

b. Riconoscimento iconografico e Catalogazione

Gli studenti hanno costruito il database che accoglie in sé la redazione dell'inventario patrimoniale delle incisioni, fornito di diversi campi e visitabile.

³ Tra gli incisori rappresentati, la collezione vanta personalità artistiche che le conferiscono grande prestigio, come Albrecht Dürer, Annibale e Agostino Carracci, Jusepe de Ribera, Salvator Rosa, Rembrandt, Jacques Callot, Giovan Battista Piranesi, il Canaletto, Domenico Morelli e molti altri. Le opere ivi custodite, grazie alla loro eterogeneità, offrono un'ampia visuale sulla produzione grafica nei diversi secoli e su come si siano evoluti i temi della rappresentazione, gli stili e le tecniche di riproduzione, data l'ampia presenza di xilografie, bulini, acquaforti e litografie databili dalla fine del XV secolo sino agli inizi del XX.

Le opere appartenenti alle scuole europee quali quelle dei Paesi Bassi, la Francia, l'Inghilterra, la Germania, la Spagna ed il Portogallo completano questa composita raccolta, la cui ecletticità non si limita alla sola provenienza, ma attinge ai diversi generi rappresentati, come le stampe di costume, i soggetti sacri, l'architettura, le vedute, la cartografia, il genere musicale, teatrale e la ritrattistica. Per una descrizione della collezione si rinvia a: Istituto Universitario di Magistero "Suor Orsola Benincasa." Fondazione A. e MA Pagliara, Anna Caputi, Maria Teresa Penta. *La raccolta d'arte della Fondazione Pagliara*. Istituto "Suor Orsola Benincasa", 1985.

Tale inventario contiene le informazioni necessarie all'individuazione del materiale: la collocazione, l'incisore, il disegnatore, l'inventore, il titolo della singola opera, epoca, editore, tecnica, misure, iscrizioni, numero di inventario, stato di conservazione, aggiornamento del lavoro di digitalizzazione effettuato e le note relative all'esemplare.

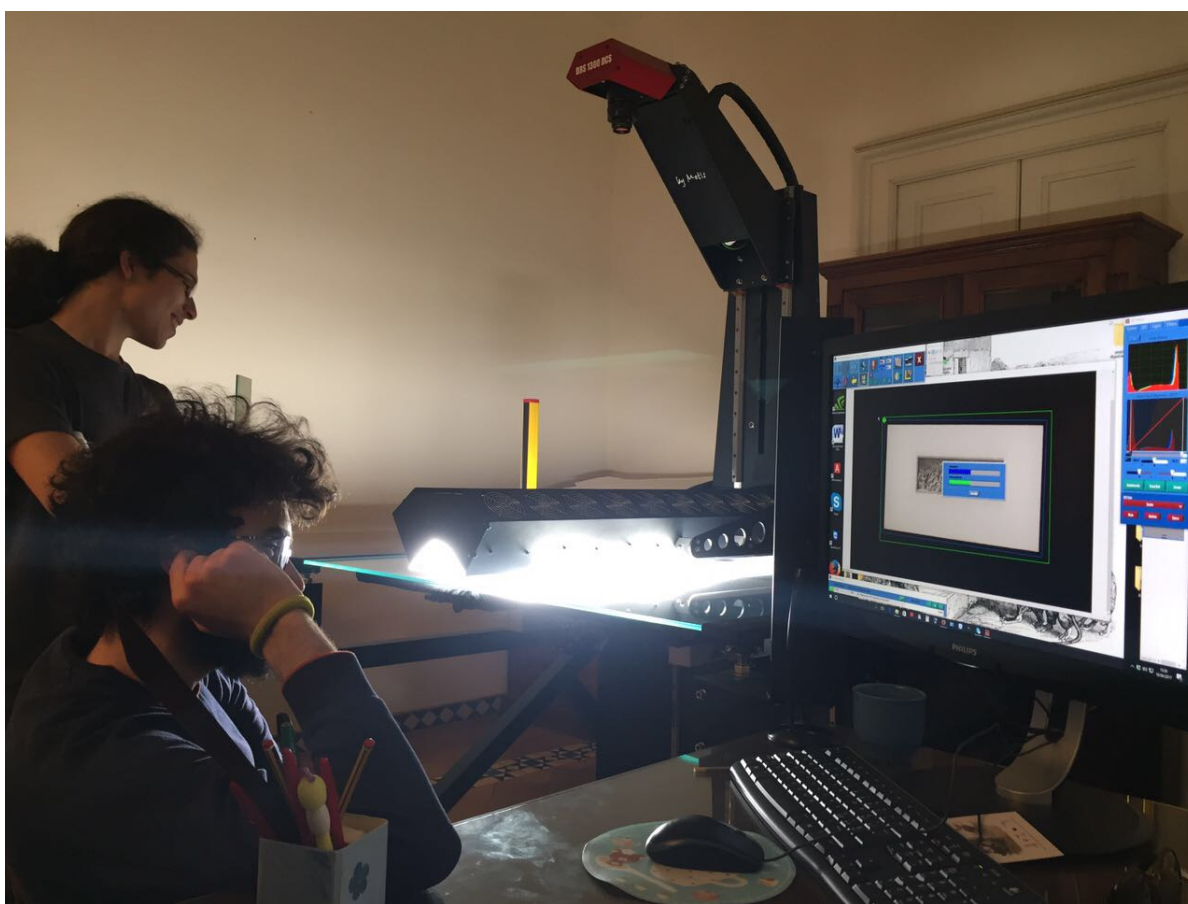
I ragazzi sono stati quindi iniziati alle norme catalografiche italiane, biblioteconomiche⁴, e artistico-documentali⁵ che normalizzano le informazioni catalografiche attraverso un linguaggio chiaro ed esaustivo che permette una efficace interrogazione da parte dell'utenza e facilita la rete dei collegamenti.

Il sistema di catalogazione informatizzata dei beni culturali insiste su due presupposti essenziali di coerenza documentale, l'uno rivolto alla specificità dell'oggetto di indagine, l'altro al recupero integrato delle informazioni.

Le competenze richieste per questo lavoro, che si inserisce nella più ampia politica di salvaguardia del patrimonio culturale sostenuto dalla Unione Europea, sono competenze spesso acquisite in contesti già avanzati nella scelta dei curricula scolastici, e che invece sarebbe importante acquisire già durante la formazione secondaria, per catturare competenze umanistiche e digitali integrate.

c. Digitalizzazione

Gli studenti hanno quindi proseguito il lavoro, attraverso la digitalizzazione delle stampe precedentemente inventariate, utilizzando uno scanner di ultima generazione che permette l'acquisizione di immagini ad altissima risoluzione⁶. Lo strumento è utilizzato specificamente per la digitalizzazione di opere d'arte come quadri, disegni, mappe, stampe e libri di diversi formati: il modello posseduto dall'Ateneo è tra i più grandi progettati e permette di scansionare opere fino alla dimensione A2.



Gli studenti alle prese con lo scanner e il software di acquisizione dell'immagine.

Gli studenti hanno così imparato ad utilizzare lo scanner e il software ad esso collegato che ha permesso loro di studiare le opere d'arte in digitale, analizzandone ogni dettaglio, dagli elementi più minuti della rappresentazione allo stato di conservazione della carta e alle tecniche di esecuzione.

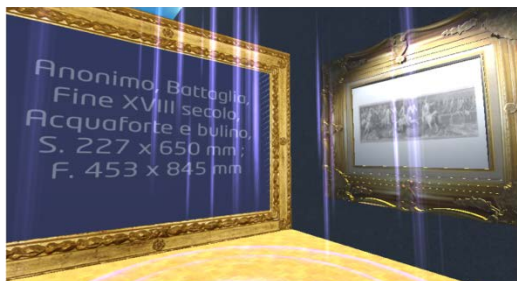
⁴ le nuove RICA (Regole italiane di Catalogazione) basate sul linguaggio ISBD (International Standard Bibliographic Description).

⁵ le nuove Normative dell'ICCD (Istituto Centrale per il Catalogo e la Documentazione).

⁶ dai 400 sino agli 800 PPI, il Light Metis DRS 1300 DCS

d. Divulgazione del patrimonio a stampa tramite la modellazione e presentazione di ambienti per Virtual Museum.

Gli strumenti utilizzati per favorire un percorso didattico che avvicinasse i giovani umanisti al dominio delle competenze ICT, necessarie alla progettazione di forme virtuali di valorizzazione del patrimonio, sono stati selezionati tra quelli open source.



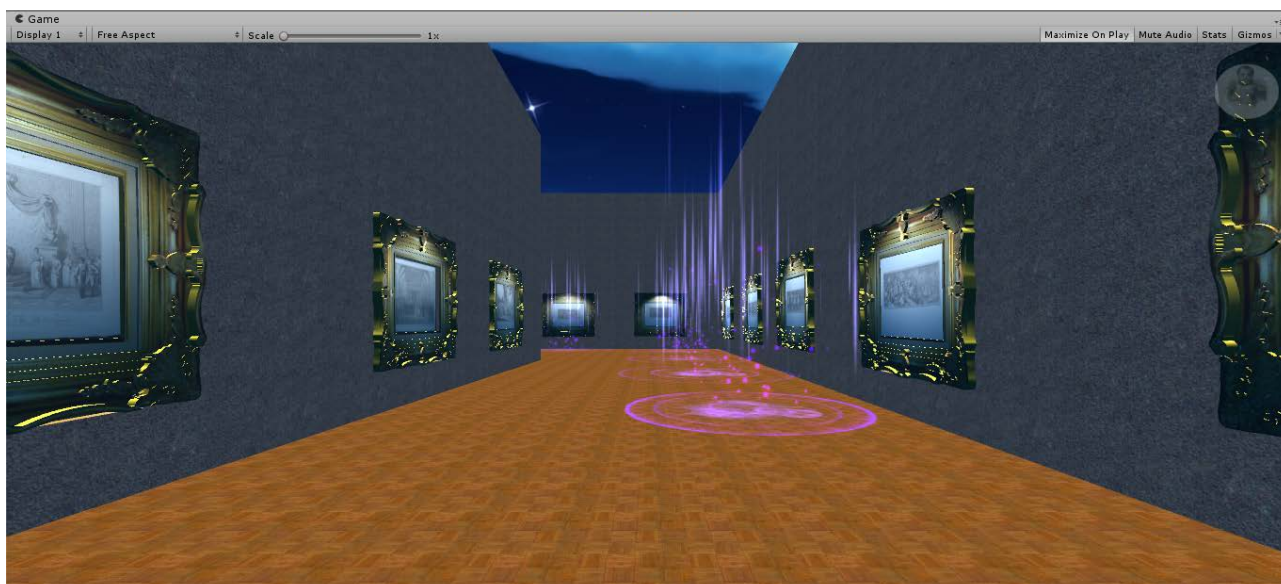
Una schermata della Galleria Virtuale e gli strumenti per navigarla in VR.

Per la modellazione dell'ambiente museale virtuale, si è quindi utilizzato SketchUp⁷, strumento freeware per la modellazione 3D, dall'uso semplificato.

Lo strumento di supporto è stato Blender⁸, software freeware che permette di arricchire il modello tridimensionale, con diverse forme di animazioni.

Per la realizzazione e l'editing delle texture GIMP⁹, con il quale è possibile produrre gli output nei più comuni formati di immagine. Per ottenere un unico file che contenesse al suo interno i dettagli delle mesh 3D, le animazioni e le texture è stato usato un plug-in free di proprietà di Autodesk, FBX Converter¹⁰.

Come strumento di presentazione della galleria così ottenuta, i ragazzi sono stati introdotti all'uso di Unity¹¹, game engine free, attraverso il quale comporre le scene e ottenere un output eseguibile da vari sistemi operativi, così da evitare il confronto con la scrittura di codici di programmazione, prediligendo invece un'attività di progettazione virtuale attraverso un'interfaccia molto simile a quella degli strumenti di modellazione 3D precedentemente elencati.



Una schermata della Galleria virtuale.

L'attività legata all'acquisizione delle competenze per l'utilizzo di questi strumenti è stata organizzata in forma laboratoriale, ossia strutturata per gruppi e per moduli consecutivi, in modo che la materia risultasse il più omogenea possibile e gli studenti riuscissero a percepire la continuità strategica tra un sistema software e l'altro.

⁷ <https://www.sketchup.com/it>

⁸ <https://www.blender.org/>

⁹ <https://www.gimp.org/>

¹⁰ <http://usa.autodesk.com/adsk/servlet/pc/item?siteID=123112&id=22694909>

¹¹ <https://unity3d.com/>

La Galleria Virtuale modellata dagli studenti per presentare le stampe scansionate è stata il risultato che il progetto ha proposto per la divulgazione e la valorizzazione del patrimonio a stampa processato: la navigazione della galleria è garantita da un link inserito nel sito del Liceo e del nostro Centro di Ricerca¹², e da un'applicazione in ambiente Android.



L'applicazione navigabile per ambiente Android.

4. CONCLUSIONI

Il percorso presentato muove dalla necessità del trasferimento di competenze digitali anche avanzate, che arruolino gli studenti dei percorsi umanistici, superando persistenti luoghi comuni legati alla rigida interpretazione dei confini disciplinari.

Le metodologie e tecnologie dei processi di digitalizzazione per la produzione e conservazione del patrimonio culturale sono domini fondamentali in cui i giovani umanisti possono contribuire con un livello di competenza informata della conoscenza approfondita dei contenuti (nella fattispecie il patrimonio culturale) e delle sue ragioni: essa è strumento per una progettazione di forme di fruizione in cui il contenitore virtuale risulti non prevaricare il messaggio culturale.

La forma laboratoriale dell'esperienza didattica è quella che meglio si presta a spiegare l'importanza della divulgazione della cultura umanistica con il supporto delle tecnologie digitali avanzate, soprattutto a favore della diffusione e valorizzazione del patrimonio culturale.

Naturalmente queste competenze risultano essere la base dello sviluppo di modelli concettuali per lo studio del patrimonio culturale.

5. RINGRAZIAMENTI

Ringraziamo la Prof.ssa Mariella De Simone per l'impegno e la classe II D del Liceo Classico Vittorio Emanuele di Napoli.

6. RIFERIMENTI BIBLIOGRAFICI

- [1] Argentin, Gianluca, Marco Gui, and Chiara Tamanini. "A scuola di competenza digitale. Il ruolo degli insegnanti nell'uso delle ICT degli studenti." *Scuola democratica* 4.1 (2013): 79-104.
- [2] Bizzarri, Giuseppe, Luca Forlizzi, and Guido Proietti. "Informatica: didattica possibile e pensiero computazionale." *Proc. of DIDAMATICA* (2011).
- [3] Benigno, Vincenza, and Stefania Manca. "L'apprendimento nella scuola del futuro. Dal laboratorio di informatica alle classi digitali: un'indagine per capire l'evoluzione dei nuovi processi di insegnamento." *Atti del Convegno DIDAMATICA* (2011).
- [4] Istituto universitario di magistero "Suor Orsola Benincasa" Fondazione A. e MA Pagliara, Anna Caputi, Maria Teresa Penta. *La raccolta d'arte della Fondazione Pagliara*. Istituto "Suor Orsola Benincasa", 1985.
- [5] Dagnino, Francesca, et al. "Serious Games, Storytelling and Intangible Heritage"
- [6] Dominici, Piero. *Dentro la società interconnessa: prospettive etiche per un nuovo ecosistema della comunicazione*. FrancoAngeli, 2014.
- [7] Fitzpatrick, Kathleen. Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, & Jeffrey Schnapp, *Digital_Humanities*. International Journal of Communication, 2014, 8: 5.
- [8] Ibba, Diana. "Progettazione e sviluppo di un Serious Game interattivo per la comunicazione di contenuti culturali." (2012).

¹² imperatore3d.altervista.org

- [9] Jodice, Mimmo, *Le savoir sur la falaise. Luoghi e storie dell'Università Suor Orsola Benincasa*. 2013.
- [10] Parrinello, Sandro, Francesca Picchio, and Monica Bercigli. "La 'migrazione' della realtà in scenari virtuali: Banche dati e sistemi di documentazione per la musealizzazione di ambienti complessi." *DISEGNARECON* 9.17 (2016): 14-1.
- [11] Rayward, W. Boyd, and Michael B. Twidale. "From Docent to Cyberdocent: Education and Guidance in the Virtual Museum" *Archives and Museum Informatics* 13.1 (1999): 23-53.
- [12] Blender. <https://www.blender.org>
- [13] Teachdigital. <http://www.teachdigital.org/2012/02/high-school-3d-modeling-projects-google-sketch-up-v-blender/>
- [14] Unity. <https://unity3d.com/showcase/gallery/non-games>
- [15] Unity. <https://unity3d.com/learn/resources/unity-serious-games-showcase>

Designing a multi-layered User Interface for EVT 2: a development report

Roberto Rosselli Del Turco¹, Chiara Di Pietro², Chiara Martignano³

¹Università di Torino, Italia - roberto.rosselidelturco@unito.it

²Università di Pisa, Italia - dipi.chiara@gmail.com

³Università di Pisa, Italia - chiaramartignano@gmail.com

ABSTRACT

EVT is a software tool to publish digital scholarly editions on the Web. Its development dates back to the start of this decade, in fact it can boast quite an interesting development history: started as a Project Based Learning initiative within the scope of a single edition project, it has evolved in a versatile tool valued and used by many researchers. The current development version has been released in October 2017: it surely is more feature-rich if compared to the first alpha version release in July 2016, but still very much critical edition-oriented. A full version supporting diplomatic / interpretative editions, complete with a new powerful image viewer and a search engine, will be released in the first half of 2018. This paper will report on EVT's progress and will also propose a few considerations about its development and the research questions that it entails, namely: the problems that derive from designing a User Interface that must take into account a great number of "objects" interacting on different layers; the new features that will be introduced in the first stable version and the impact that they will have on the final users; the soundness of the open source development model and the relationship with other projects.

KEYWORDS

XML, TEI, Digital Philology, Digital Scholarly Edition, Critical Edition, Publishing Software.

1. INTRODUCTION

Edition Visualization Technology[4][10] (EVT from now on) was born in the context of a specific digital edition project, the Digital Vercelli Book[7]. The latter aims at a facsimile and diplomatic edition of the whole Vercelli Book codex, with texts encoded in the TEI XML format[8]: when the encoding of all the manuscript texts neared completion, the developers started an inquiry on the state of the art to identify a suitable edition browsing software, but at the time (about 2010) the available tools were either proprietary, or closely linked to a specific software platform, or too complex to install and use, or so tied to a specific project that adapting them for the project's purposes would have required too much time and resources, or a mix of the above. EVT was started as a Project Based Learning initiative for students attending the degree course in Informatica Umanistica (Digital Humanities) at the University of Pisa, so that it could also have an impact with regard to the teaching side of Digital Philology, and can boast a long development history[6], with more than one false start.



Figure 1. The Digital Vercelli Book

At the end of 2013 the EVT development team had managed to produce a working tool that could be used effectively to produce the *Digital Vercelli Book* first beta edition. What's more, right from the start EVT was developed according to open criteria: distributed as open source software, based on the major standards for distribution on the Web (hence the use of HTML, CSS, JavaScript, etc.), as such naturally multi-platform and independent of specific software

environments/packages, user-friendly (both for the project editor, who should be able to publish the edition on a Web site without any server software being required, and the final user) and highly flexible with regard to User Interface layout. Since it had been designed as a general purpose tool, albeit with regard to the specific task of publishing diplomatic editions of TEI-encoded texts coupled with digitized images of the original manuscripts, it could be used with different manuscripts and texts, and in fact already during 2014 a new edition project was started that uses EVT to publish the resulting digital edition on the Web: the *Codice Pelavicino Digitale* project[2].

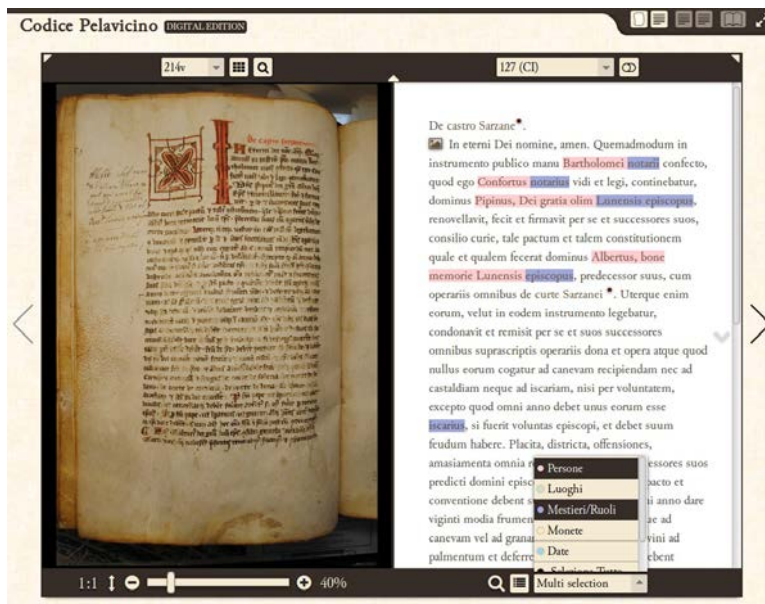


Figure 2. The Codice Pelavicino Digitale

This first version is based on an XSLT 2.0 transformation chain that takes as its input one or more documents in the standard TEI P5 format, which are turned into a client-only web-based application, using standard web technologies such as HTML5, CSS3 and JavaScript, which can be easily copied on a server and shared on the Web. Unfortunately, with time the XSLT code-base has grown considerably and it is now quite complicated to understand and modify, especially for new contributors. Furthermore, the increasing number of available widgets has made the general layout quite complex to configure, and difficult to handle with regard to the interaction between them. We needed more flexibility and modularity to make the development of new features easier and faster, and this has led to a completely new software architecture.

2. A NEW ARCHITECTURE: EVT 2.0

One of the new features to be implemented, perhaps the most important one since it was both planned for the original project (the Digital Vercelli Book) and requested by many other scholars, was support for complex, multi-witness critical editions, again encoded according to the TEI schemas and *Guidelines*. As stated in the previous section, we quickly realized that the old programming framework, especially with regard to the code handling the UI layout, was woefully unfit for the task: there was no way that we could add complex functionality to the existing code-base while adding a new layer of GUI objects to the web-application interface. For these reasons, we decided to remodel the entire structure of the software, in order to make it lighter, more manageable and more adaptable. We decided to take advantage of the Model View Controller (MVC) design pattern, which is a very popular architecture in object-oriented programming, and to implement it by means of the AngularJS[1] framework.

The first instance of the new version has been released in July 2016 and it has been the work of Chiara Di Pietro[3]. A more advanced beta version, with support for a fuller set of critical edition features, has been released in October 2017 thanks to Chiara Martignano's contribution[5], and as the previous one is available on the project home page¹.

The current EVT 2 version is almost complete with regard to critical edition support, but still needs a backporting of all of EVT 1 features (search engine, image viewer, support for lists and named entities, etc.). This will be the focus of the next development effort, with the goal of making EVT suitable to publish both diplomatic and critical editions (or a combination of the two types) encoded according to the TEI XML standard.

¹ See the development blog (<http://visualizationtechnology.wordpress.com/>) for the full announcement.

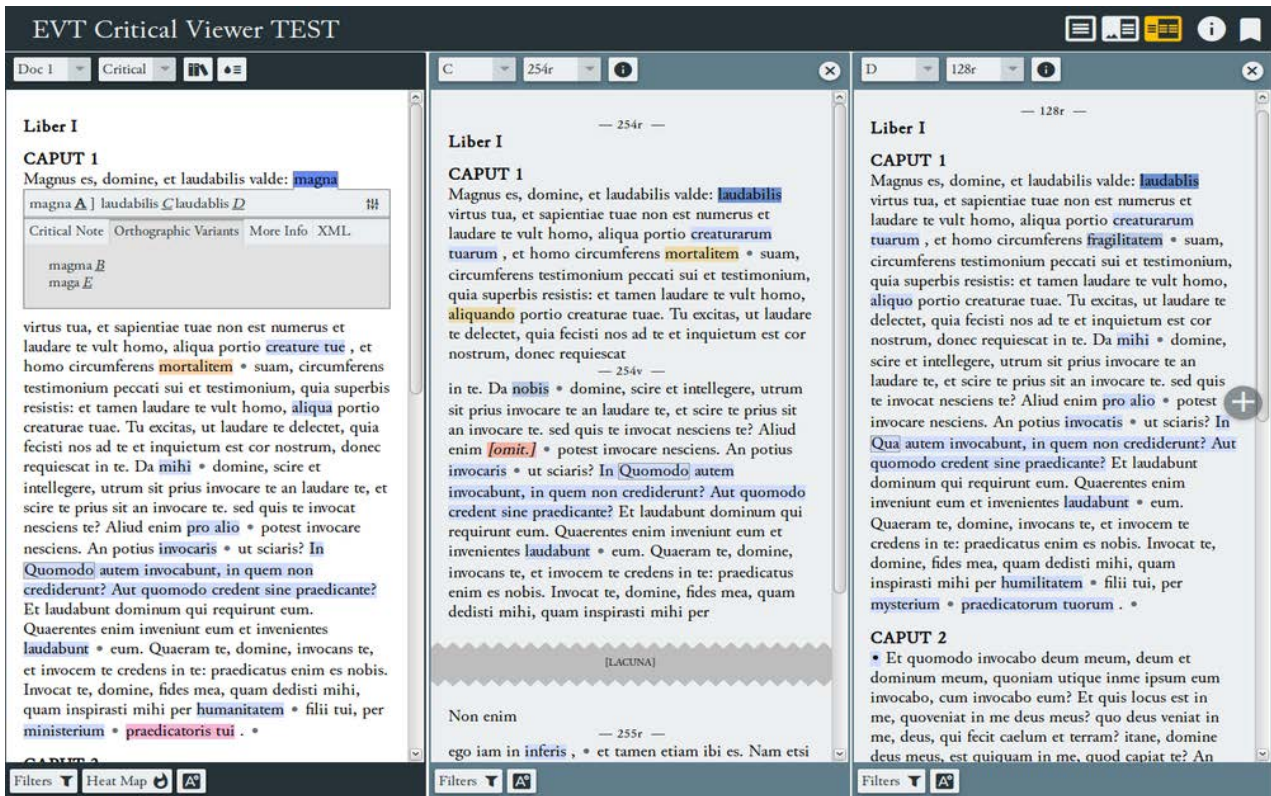


Figure 3. The first EVT 2 alpha version

At the present moment the EVT development team is working together with the researchers and scholars involved in the PhiBor - *Philosophy on the Border of Civilizations* project[9], in fact the default example used in the beta version (see Fig. 4) is based on a sample prepared as a test for their forthcoming digital edition of Aristotle's *Metaphysics* as translated by Avicenna. As it was the case for the *Codice Pelavicino Digitale*, close cooperation with a complex and rich digital edition project is essential to refine and improve the set of features that EVT can offer to the academic community.

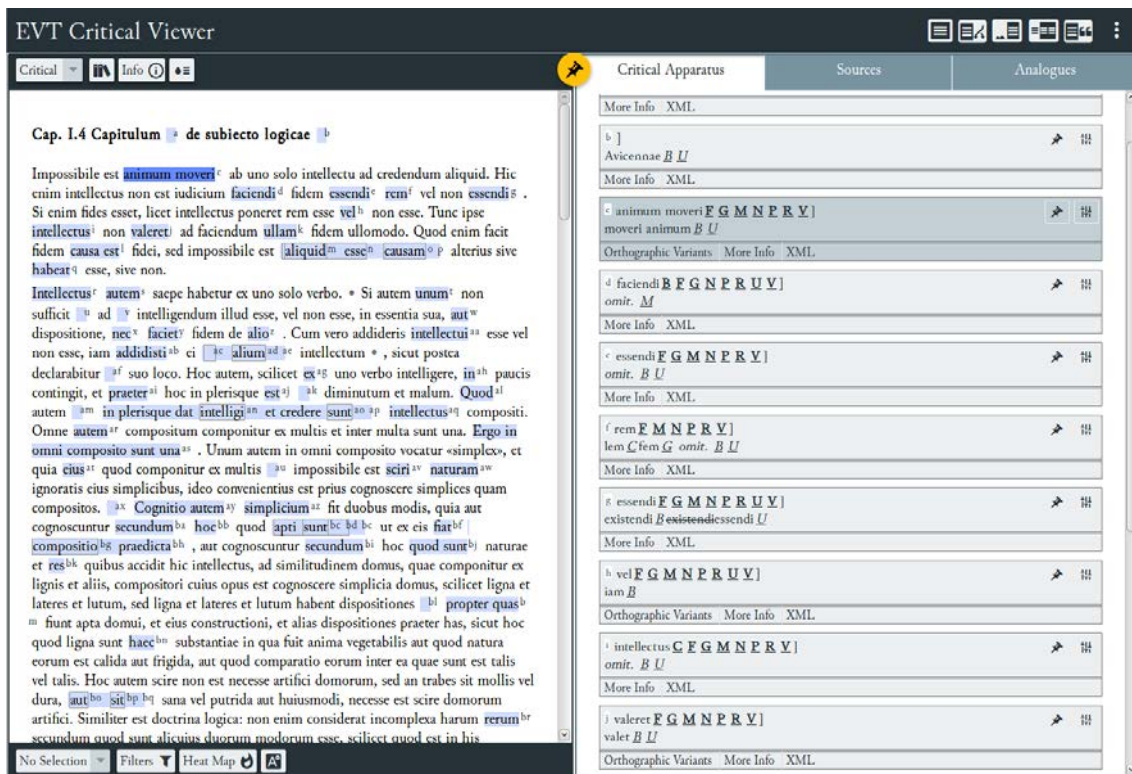


Figure 4. The current EVT 2 beta version

3. DEVELOPMENT AND RESEARCH ISSUES

The ongoing development of EVT has entailed discussion on several issues, among the most interesting with regard to research and methodology:

- designing an User Interface that must take into account a great number of “objects” interacting on different layers² has proved to be quite difficult, several issues have been fixed but many others await for an adequate solution;
- the new features that will be introduced in the first stable version will surely be interesting for final users, the risk being that of increasing complexity if the UI issues are not resolved;
- the open source development model has proved itself so far, and the relationship with other projects has been almost always very positive.

We are especially interested in receiving feedback on these topics from our users and from the general public of AIUCD 2018.

4. REFERENCES

- [1] AngularJS. <https://angularjs.org>.
- [2] *Codice Pelavicino*. Edizione digitale, a cura di E. Salvatori, E. Riccardini, L. Balletto, R. Rosselli del Turco, C. Alzetta, C. Di Pietro, C. Mannari, R. Masotti, A. Miaschi, 2014. ISBN 978-88-902289-0-2; DOI 10.13131/978-88-902289-0-2; URL: <http://pelavicino.labcd.unipi.it>.
- [3] Di Pietro, C. 2016. *EVT per le edizioni critiche digitali: progettazione e sviluppo di una nuova GUI basata sullo schema progettuale MVC*. MA thesis. Pisa: ETD - Electronic theses and dissertations repository, Università di Pisa. <https://etd.adm.unipi.it/theses/available/etd-03292016-203644/>.
- [4] *EVT - Edition Visualization Technology*. Project home: <http://evt.labcd.unipi.it/>. Development blog: <http://visualizationtechnology.wordpress.com/>. Distribution repository: <https://sourceforge.net/projects/evt-project/>. Code repository on GitHub: <https://github.com/evt-project/evt-viewer>.
- [5] Martignano, C. 2017. *Progettazione e sviluppo di un apparato critico modellato sulla tradizione a stampa in EVT*. MA thesis. Pisa: ETD - Electronic theses and dissertations repository, Università di Pisa.
- [6] Rosselli Del Turco, R. 2014. “EVT Development: An Update (and Quite a Bit of History).” *Edition Visualization Technology blog*, <https://visualizationtechnology.wordpress.com/2014/01/26/evt-development-an-update-and-quite-a-bit-of-history/>.
- [7] *The Digital Vercelli Book*, a cura di R. Rosselli Del Turco; trascrizione e codifica a cura di R. Rosselli Del Turco, R. Cioffi, F. Goria; software EVT creato da C. Di Pietro, J. Kenny, R. Masotti, R. Rosselli Del Turco. 2017. ISBN 9788875901073. URL: <http://www.collane.unito.it/oa/items/show/11>.
- [8] TEI Consortium (eds.). 2017. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.2.0]. [2017-07-10]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.
- [9] PhiBor 2014-2019. *Philosophy on the Border of Civilizations and Intellectual Endeavours: Towards a Critical Edition of the Metaphysics (Ilāhiyyāt of Kitāb al-Šifāʾ) of Avicenna (Ibn Sīnā)*. ERC Advanced Grant (n. 339621). <http://www.avicennaproject.eu>.
- [10] Rosselli Del Turco, R. et al. 2015 “Edition Visualization Technology: A Simple Tool to Visualize TEI-Based Digital Editions.” *Journal of the Text Encoding Initiative*, Issue 8 (2015). URL: <http://jtei.revues.org/1077> DOI: 10.4000/jtei.1077.

² Even if the available space for a digital edition is virtually unlimited, the physical space on an LCD screen, such as a computer monitor or a tablet, is not, since it is a 2D space just as in a traditional printed edition. It is therefore necessary to define a set of User Interface layers, such as panels sliding over the main window and/or tabbed window frames, which can be used to host the edition content (organized in smaller UI components, the “objects” hinted to in the main text) in a flexible manner.

Digitalizzare la documentazione archeologica: metodologie e “buone pratiche” per la lavorazione, gestione e il riuso dei dati.

L'archivio della Missione Archeologica Italiana in Anatolia Orientale (Arslantepe-Malatya, Turchia)

Margherita Bartoli¹, Heidi Paffrath²

¹Digilab - Sapienza Università di Roma - margherita.bartoli@uniroma1.it

²paffrath.1226869@studenti.uniroma1.it

1. INTRODUZIONE

Il contributo dei sistemi informativi ha oggi un ruolo determinante nella ricerca archeologica ed ha portato ad un vero e proprio rinnovamento metodologico della disciplina. Oggi l'archeologo collabora attivamente allo sviluppo di nuovi processi di documentazione partendo dalle proprie esigenze ed evitando così che la tecnologia lo sovrasti con le proprie regole. Molti progetti, nazionali e internazionali, si pongono infatti come obiettivo lo studio di procedure standard per la raccolta, gestione, conservazione e pubblicazione dei dati, decisamente eterogenei, che provengono dagli scavi archeologici. Proprio in questo panorama sempre più articolato nasce nel 2016 il progetto di studio per la digitalizzazione dei dati di scavo, frutto dalla collaborazione del centro Digilab¹ con la missione archeologica italiana in Anatolia Orientale (MAIAO), finalizzata alla creazione di un archivio digitale relativo al sito di Arslantepe² e alla sua pubblicazione, almeno per la parte edita, sulla piattaforma di Sapienza Digital Library (SDL).

Il suo patrimonio documentale unico, raccolto e arricchitosi nel corso degli ultimi sessant'anni di scavi, si caratterizza per la sua multiforme varietà di supporti³ e metodi d'archiviazione, ponendosi come esempio significativo degli sviluppi metodologici della ricerca archeologica negli ultimi anni e ottimo *case study* per lo studio e l'applicazione di “buone pratiche” per la gestione di dati tipologicamente così differenti.

L'archivio “fisico” della missione è oggi conservato nella facoltà di Lettere e filosofia della Sapienza Università di Roma ed è stato soggetto, negli ultimi anni, a ben due traslochi, evidenziando i rischi connessi alla sua integrità e quindi l'urgenza di attivare strategie conservative, elaborando un progetto organico per l'immagazzinamento e la gestione di questo immenso e prezioso patrimonio documentale.

2. ARSLANTEPE DIGITAL LIBRARY: SCELTE METODOLOGICHE E OBIETTIVI

Data l'enorme mole di materiali e soprattutto per permettere una sua attenta e puntuale metadattazione, il progetto ha previsto una prima fase pilota in cui si è proceduto al trattamento di un ristretto lotto di materiali documentali riservando particolare attenzione ad un settore di scavo di Arslantepe ben delimitato, sia cronologicamente sia topograficamente: la fase VIII (fase di occupazione datata 4200-3900 a.C.). Questa operazione preliminare ha avuto il duplice scopo di testare le metodologie più idonee per la digitalizzazione di materiali così diversi tra loro, nonché la definizione dei criteri minimi di metadattazione degli stessi. Sulla base di questa esperienza preliminare, si prevede di estendere il progetto all'intera documentazione di scavo.

Per questa fase iniziale si è creato un gruppo di lavoro composto da specialisti archeologi⁴ della Sapienza e dell'Università di Roma Tre, coordinati dal Centro DigiLab. Si è proceduto così alla digitalizzazione dei materiali documentali, provenienti da campagne di scavo degli anni 1992, 1997, 1998, 2003, 2006 e 2010:

- quaderni di scavo;
- disegni di oggetti su cartoncino e su carta lucida;
- diapositive a colori;
- negativi (di formato 35mm e 6x6).

¹ La direzione scientifica della prof.ssa Francesca Balossi (Sapienza Università di Roma), supportata dal dott. Francesco di Flippo (CNR) e, per parte Digilab, del prof. Gianfranco Crupi.

² Il sito si trova nel villaggio di Orduzu, in Anatolia Orientale, un'oasi nella catena dell'Anti Tauro, a circa 15 Km dalla riva destra dell'Eufrate e a 6 Km dalla città moderna di Malatya e sorge su un tell, una collina artificiale alta circa 30m con una superficie totale di 4 ettari, nata dal continuo sovrapporsi di abitati, ricostruiti ininterrottamente, a partire almeno dal V millennio a.C. Il progetto svolto nel laboratorio Digilab si affianca a quello portato avanti dalla Missione Archeologica che sta procedendo alla creazione di un database specifico per la gestione, direttamente sul sito, della documentazione raccolta durante lo scavo. Dal 1961 ad oggi, ricerche sistematiche sono condotte dalla Missione Archeologica Italiana in Anatolia Orientale (MAIAO), prima sotto la guida congiunta del Prof. Meriggi (Università di Pavia) e del Prof. Salvatore Maria Puglisi (Sapienza, Università di Roma), poi diretti dalla Prof.ssa Alba Palmieri (Sapienza, Università di Roma) e dal 1990 dalla Prof.ssa Marcella Frangipane, Professoressa ordinaria di preistoria e protostoria presso il dipartimento di Scienze dell'Antichità dell'Università La Sapienza Università di Roma. L'importanza dei risultati ottenuti da queste ricerche, ad oggi ancora in corso, ha spinto il Ministero della Cultura e del Turismo della Repubblica di Turchia ad inoltrare domanda all'UNESCO per l'inserimento del sito di Arslantepe tra quelli considerati Patrimonio dell'Umanità.

³ È composto da quaderni di scavo, planimetrie cartacee delle strutture architettoniche, dai disegni dei singoli reperti archeologici fino ad includere diversi supporti fotografici (film negativo di diversi formati, diapositive a colori, stampe positive, film 8x8, VHS...)

⁴ Studenti di lauree triennale, specialistica e master.

La sfida maggiore di questo progetto è stata quella di riuscire a coniugare le esigenze della MAIAO che utilizza già, seppur in fase di aggiornamento, un database per la documentazione più recente, con le esigenze della Sapienza Digital Library, la cui struttura (che si basa su metadati MODS) obbliga a seguire regole precise per permettere una corretta conservazione, organizzazione e recupero delle informazioni.

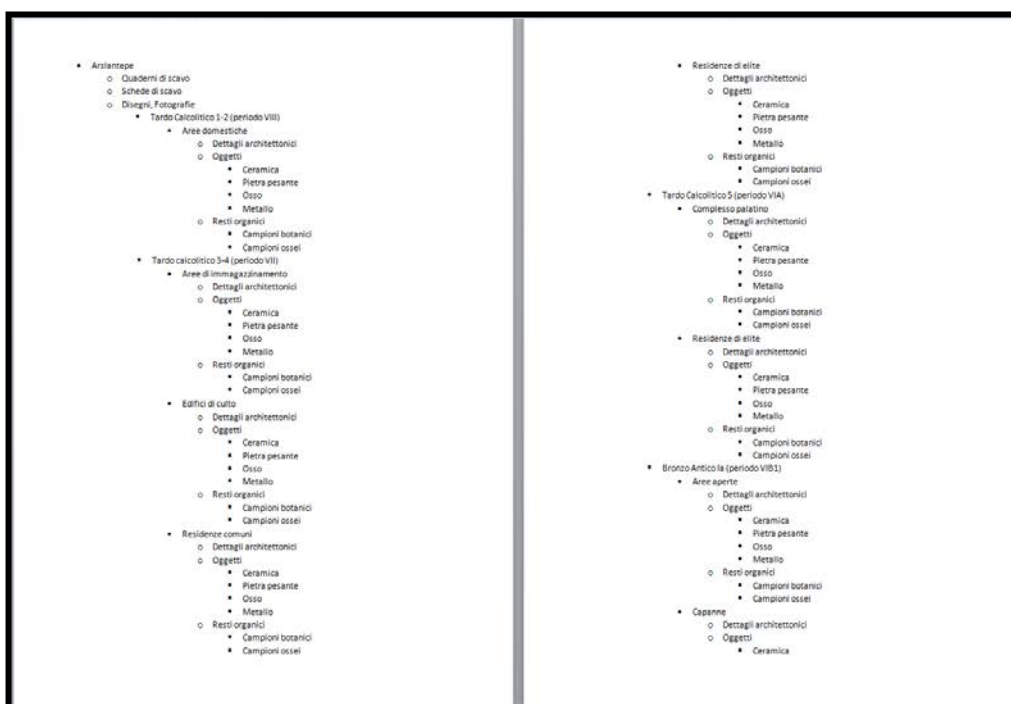


Fig. 1 Parte del complesso albero della collezione Arslantepe.

La prima fase di lavoro, che ha comportato uno stimolante confronto tra gli specialisti della missione e il gruppo di ricerca Digilab, è stata incentrata innanzitutto sull'elaborazione di un albero gerarchico della collezione (Fig.1) che ne esplicitasse i contenuti, cercando in questo modo di rispettarne e rappresentarne adeguatamente la struttura a più livelli, senza tralasciare la coerenza descrittiva che permette l'interoperabilità tra le risorse all'interno di SDL.

Il risultato di questo intenso lavoro è stato la creazione *ad hoc* di un set di metadati differenti a seconda della tipologia di materiale che, pur non rispettando uno standard esistente, ci permette, tramite mappatura, di ottenere dati interoperabili con entrambi i sistemi di gestione (SDL e Archivio MAIAO). (Tab 1)

L'analisi delle singole immagini in fase di metadattazione, l'alta risoluzione della scansione dei materiali⁵ e la correlazione tra le risorse digitalizzate, ha consentito di riconoscere e identificare anche quelle prive di informazioni, arricchendo, integrando e migliorando in maniera esponenziale la qualità dei dati.

Partendo dai risultati raggiunti è stato possibile riflettere su un eventuale riuso di questi interessanti dati. Si è perciò pensato di utilizzare le informazioni, relative alla localizzazione dei ritrovamenti nel sito, per la creazione di una piattaforma GIS collegata ad un database relazionale. Questa azione è stata oggetto di una tesi di master⁶ svolta dalla dott.ssa Heidi Paffrath, che ha previsto la creazione di un *learning tool* nella forma di un'applicazione web: un Archeo-GeoPortale (oggi ancora offline), dedicato alla presentazione di una sintesi del Periodo VIII di Arslantepe che potrebbe costituire un primo passo verso la consultazione e la condivisione dei dati prodotti con il maggior numero di utenti possibile (specialisti e non).

3. CONCLUSIONI PRELIMINARI E PROSPETTIVE DI RICERCA

La fase pilota del progetto relativo alla digitalizzazione dell'archivio della MAIAO, ha dato la possibilità di analizzare quali possono essere le problematiche riscontrabili nel trattamento di un archivio così complesso ed eterogeneo all'interno della Sapienza Digital Library, dando inoltre la possibilità di testare la scientificità delle procedure standardizzate elaborate dal gruppo di ricerca. Nell'arco di un anno di lavoro è stato possibile ottenere ottimi risultati, riuscendo a digitalizzare e metadattare 19 album da disegno (1180 disegni), 64 diari (3561 file), 3 raccoglitori di negativi (2330 negativi 35mm e 546 in formato 6x6), 4 raccoglitori di diapositive (2504 diapositive) per un totale di 900 ore di lavoro.

⁵ Scansionati con scanner Metis Gamma e Epson V700 tutti in TIFF a 600, 800 o 1200 dpi a seconda della tipologia.

⁶ Beni Territoriali e Culturali” Master di II Livello dell’Università Roma Tre “Digital Earth e Smart Governance: Strategie e Strumenti GIS per la Gestione dei Beni Territoriali e Culturali”

Questo attento lavoro di standardizzazione dei processi e la creazione di una piattaforma GIS hanno tentato di rispondere concretamente alle esigenze della ricerca e conservazione della documentazione prodotta, confermando di poter contribuire realmente alla tutela, alla fruizione e alla valorizzazione di questo sito così importante e complesso.

La digitalizzazione dei materiali e la loro corretta gestione, hanno dimostrato di poter agevolare la consultazione del materiale documentario relativo al periodo VIII di Arslantepe, facilitando la correlazione tra dati e permettendone una più semplice lettura e un immediato riuso.

Tipologia materiale	Set di metadati
<p>Quaderni di scavo (scansionati con scanner Metis come immagini in formato TIFF a 600 dpi)</p>	<p>Nome diario: nome unico del diario Nome cartella: nome attribuito alla cartella di riferimento Numero file: nome della parte di diario Anno: anno dello scavo q20 m: numero del quadrato 20x20m con cui è diviso lo scavo q5m: numero del quadrato 5x5m con cui sono divisi i quadrati 20x20m Duplicato: indicazione della presenza di un altro quaderno che descrive la stessa area di scavo Note</p>
<p>Disegni di oggetti (I singoli disegni presenti nel foglio vengono metadati singolarmente riportando di volta in volta il nome del foglio. Sono scansionati con scanner Metis in formato TIFF a 600 dpi)</p>	<p>Faldone: nome del contenitore fisico Foglio: numero del foglio Nome File: nome della singola risorsa Cod. Disegno Anno Disegno Quadrato: numero del quadrato 20x20m e/o 5x5m dove è stato rinvenuto l'oggetto Cartellino Scavo Anno: Anno in cui l'oggetto è stato scavato Descrizione: se presente, viene riportata la descrizione dell'oggetto Note Periodo: cronologia relativa dello scavo</p>
<p>Diapositive/Negativi (35mm e 6x6) (Diapositive: scansionate in formato TIFF a 1200 dpi; Negativi 6x6 in formato TIFF a 800 dpi; Negativi 35mm in formato TIFF 800 dpi)</p>	<p>Faldone: nome del contenitore fisico Nome File: nome della singola risorsa Num. rullino Anno: anno dello scatto Note Periodo: cronologia relativa al sito di Arslantepe q20: numero del quadrato 20x20m con cui è diviso lo scavo q5: numero del quadrato 5x5m con cui sono divisi i quadrati 20x20m Strato: gli strati, indicati con numeri progressivi (1, 2, 3...), coincidono con differenze realmente visibili del terreno, come: colore, consistenza o differenza nei materiali; all'ultimo strato che poggia sul pavimento della struttura viene attribuito il simbolo *, ad indicare che il materiale rinvenuto poggia direttamente sul pavimento dell'ambiente (materiale in situ). Le lettere minuscole (a, b, c...) indicano invece i "tagli", ossia divisioni arbitrarie che sono riconoscibili all'interno di uno stesso strato. Non hanno alcuna corrispondenza con differenziazioni reali; Kuyu: la presenza di un pozzetto/fossa a cui è attribuito un numero unico. Ambiente: a delimitazione di un'area, sia interna che esterna, a cui viene attribuito un numero unico Sepoltura: le sepolture e tombe (forma di unità stratigrafica), vengono indicate con la "S", le Tombe con la "T". La numerazione è progressiva per il sito intero. Homo: registra la presenza di reperti umani sepolti; vengono indicati con la "H". La numerazione è progressiva per il sito intero; R.I.: reperto isolato, ad esempio, un vaso intero o un oggetto particolare Etichetta scatola: informazioni riportate sui coperchi dei contenitori Cod. ARS: codice presente sulla diapositiva Inventario oggi.: numero inventario dell'oggetto</p>

Tab. 1 Set di metadati specifici per tipologia di materiale

4. BIBLIOGRAFIA

- [1] Anichini, F., Fabiani, F., Gattiglia, G., Gualandi, M. L., 2012 – Un database per la registrazione e l'analisi dei dati archeologici, *Map Papers 1-II:1-20*.
- [2] Ardissonne, P., Balossi, F., Borgia, E., Brienza, E., 2008 - Web GIS solutions for the analysis and valorisation of archaeological sites in the Mediterranean basin, in A. Posluschny, K. Lambers, I. Herzog (a cura di), *Layers of Perception. Proceedings of the 35th Computer Applications and Quantitative Methods in Archaeology Conference*, Berlin, Germany, April 2-6, 2007 (Kolloquien zur Vor- und Frühgeschichte, vol. 10), Habelt: Bonn.
- [3] Balossi Restelli, F., 2008 - Post-Ubaïd Occupation on the Upper Euphrates: Late Chalcolithic 1-2 at Arslantepe (Malatya, Turkey), in H. Kuhne, R. M. Czichon, F.J. Kreppner (eds.), *Proceedings of the 4th International Congress on the Archaeology of the Ancient Near East Conference*, Vol. 2, Harrasowitz Verlag, Berlin: 21-32.
- [4] Balossi Restelli, F., 2012a - At the Roots of the Late Chalcolithic Society in the Anatolian Euphrates Valley, *Origini XXXIV*: 41-58.
- [5] Balossi Restelli, F., 2012b - The beginning of the Late Chalcolithic Occupation at Arslantepe, Malatya, in C. MARRO (ed.), *After the Ubaïd: Interpreting Change from the Caucasus to Mesopotamia at the Dawn of Urban Civilization (4500 – 3500 BC)*, De Boccard, Paris: 235-260.
- [6] Balossi Restelli, F., 2015 - Hearth and Home. Interpreting Fire Installations at Arslantepe, Eastern Turkey, *From the Fourth to the Beginning of the Second Millennium BCE*, *Paléorient*, Vol. 41.1, CNRS, Paris: 127-151.
- [7] Bertoldi, S., Fronza, V., Valenti, M., 2011 – Web 2.0 e gestione integrale dei dati di scavo, in SITAR, *Potenziale archeologico, pianificazione territoriale e rappresentazione pubblica dei dati*, Atti del II Convegno, Roma, Palazzo Massimo, 9 novembre 2011, a cura di M. Serlorenzi E I. Jovine, Roma: 97-107.
- [8] Bertoldi, S., Fronza, V., Valenti, M., 2015 – Sistemi Digitali di Documentazione e Analisi Archeologica. Verso quale Direzione?, *Archeologia e Calcolatori*, 26: 233-245. http://www.archcalc.cnr.it/indice/PDF26/28_Bertoldi_et_al..pdf
- [9] Bogdani, J., 2016 – “Un archivio digitale multidisciplinare per la gestione e la conservazione di un patrimonio culturale a rischio: il progetto Ghazni (Afghanistan)” in *Archeologia e Calcolatori*, Supplemento 8: 236-245.
- [10] http://www.archcalc.cnr.it/indice/Suppl_8/31_Bogdani.pdf
- [11] Brunet, F., Cosner, A. - Naponiello, G., 2016 – “WebGIS interdisciplinari e questioni di metodo: il progetto Le fonti per la storia”, in *Archeologia e Calcolatori*, Supplemento 8: 141-149 http://www.archcalc.cnr.it/indice/Suppl_8/19_%20Brunet_et_al.pdf
- [12] Busana, M.S., Carpanese, I. - Orio, N., 2016 – “La ricerca a portata di click. Database relazionali e siti web: contenitori e visualizzatori per i dati archeologici” in *Archeologia e Calcolatori*, Supplemento 8: 246-253.
- [13] http://www.archcalc.cnr.it/indice/Suppl_8/32_Busana_et_al.pdf
- [14] Calaon, D., Pizzinato, C. 2011. “L’analisi archeologica nei processi di valutazione ambientale. Proposta metodologica in ambiente GIS”, in *Archeologia e Calcolatori*, 22: 413-439. http://www.archcalc.cnr.it/indice/PDF22/AC_22_Calaon_Pizzinato.pdf
- [15] Corti L. 1999. *I Beni Culturali e la loro catalogazione*. Torino, Paravia.
- [16] Catarci T., Di Iorio A., Schaerf M., 2014 – “The Sapienza Digital Library from the holistic vision to the actual implementation”, in *Procedia Computer Science* 38: 4 – 11
- [17] Casagrande, L., Cavallini, P., Frigeri, A., Furieri, A., Marchesini, I., Neteler, M., 2012 - *Gis open source. Grass Gis, Quantum Gis e SpatiaLite. Elementi di software libero applicato al territorio*, Dario Flaccovio Editore: Palermo.
- [18] D’Andrea, A., 2006 - *Documentazione archeologica, Standard e trattamento Informatico*, Budapest http://docenti2.unior.it/doc_db/doc_obj_18122_25-03-2011_4d8c69c14d60e.pdf
- [19] Di Nocera, G.M., 2005 – Ricognizione archeologica nel territorio di Malatya, in A. TANGIANU (ed.) *Dall’Eufrate al Mediterraneo. Ricerche delle Missioni Archeologiche Italiane in Turchia*, Istituto Italiano di Cultura di Ankara, Ankara: 55-63.
- [20] Ferrari O. 1990. “La catalogazione dei beni archeologici e le tecnologie informatiche”, in *Archeologia e Calcolatori*, 2, pp. 113-17. Roma.
- [21] Frangipane, M., 1993 - Local Components in the Development of Centralised Societies in Syro-Anatolian Regions, in M. Frangipane, H. Hauptmann, M. Liverani, P. Matthiae, and M. Mellink (eds.), *Between the Rivers and Over the Mountains*, Roma: Dipartimento di Scienza Storiche, Archeologiche e Antropologiche dell’Antichà, Università di Roma La Sapienza: 133-161.
- [22] Frangipane, M., 1996 - *La Nascita Dello Stato Nel Vicino Oriente Antico*, Laterza: Bari.
- [23] Frangipane, M., 2002 - ‘Non-Uruk’ developments and Uruk-linked features on the northern borders of Greater Mesopotamia, in S. Campbell, N. Postgate (eds.), *Artefacts of Complexity. Tracking the Uruk in the Near East*, British School of Archaeology in Iraq, London: *Iraq Archaeological Reports* 5: 123-148.
- [24] Frangipane M.(ed) 2012 - Fifty years of excavations and researches at Arslantepe Malatya (Turkey). A contribution to the study of the earliest centralised societies, *Thematic volume, Origini* 34. Gangemi, Rome.
- [25] Fronza, V., 2003 - Principi di database management in archeologia: l’esperienza senese, in R. Fiorillo, P. Peduto (a cura di), *III Congresso Nazionale di Archeologia Medievale (Complesso di Santa Sofia, Salerno, 2 - 5 ottobre 2003)*, Firenze: 629-632. - http://unisi.academia.edu/vittoriofronza/papers/92225/2003-Principi_di_database_management_in_archeologia_lesperienzasenese
- [26] Fronza, V., 2009 - L’archiviazione del dato in archeologia, in V. Fronza, A. Nardini, M. Valenti (a cura di), *Informatica e Archeologia Medievale. L’esperienza senese*, Firenze: 29-43.
- [27] Fronza, V., Nardini, A., 2009 - Un sistema informativo per la gestione dello spazio urbano, in G. Volpe, P. Favia, V Congresso Nazionale di Archeologia Medievale, Foggia-Manfredonia 30 settembre-3 ottobre 2009, Firenze. <http://unisi.academia.edu/VittorioFronza/Papers/125095/2009>
Un sistema Informativo per la gestione dello spazio urbano
- [28] Fronza, V., 2016 – “OpenArcheo2. An archaeological knowledge base” in *Archeologia e Calcolatori*, Supplemento 8: 254-262. http://www.archcalc.cnr.it/indice/Suppl_8/33_Fronza.pdf
- [29] Furieri, A., 2011, *SpatiaLite Cookbook*, sotto licenza Attribution-ShareAlike 3.0 unported (CC BY-SA 3.0) della licenza.
- [30] Gattiglia, G., 2009 - Open digital archives in archaeology a good practice, *Archeologia e Calcolatori*, Supplemento 2: 49-63.
- [31] Gattiglia, G., 2017 – From digitization to datafication. A new challenge is approaching archaeology, *AIUCD 2017 Conference, book-of-abstract*: 33-37.

- [32] Guercio M, Carloni C., 2012 – “Livelli descrittivi, relazioni e contesti di produzione nella sapienza digital library” in F. Ciotti (a cura di) *Digital Humanities: progetti italiani ed esperienze di convergenza multidisciplinare* Atti del convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) Firenze, 13-14 dicembre 2012: 51-69. http://www.editricesapienza.it/sites/default/files/5094_Digital_Humanities.pdf
- [33] Hall, B., Leahy, M.G., (eds.) 2008 - *Open Source Approaches in Spatial Data Handling*, Springer, Berlin.
- [34] Innocenti A., 2016 – “La digitalizzazione dei documenti, i database e il GIS nell'ambito della ricerca topografica: l'esempio dei possedimenti dell'abbazia di Farfa (RI) tra VIII e IX secolo” in *Archeologia e Calcolatori*, Supplemento 8: 126-131.
- [35] http://www.archcalc.cnr.it/indice/Suppl_8/17_Innocenti.pdf
- [36] Kennedy, M., Kopp, S., 1994 – *Understanding Map Projections*, ESRI, California.
- [37] Liss H., Ingvarsson-Sundström A., 2015 – “Databases and standardisation – experiences from Athens” in *Archeologia e Calcolatori* 26: 41-42.
- [38] Niccolucci, F. 2015 – “Un'infrastruttura di ricerca per l'archeologia: il progetto ARIADNE.” in *Archeologia e Calcolatori*, Supplemento 7: 41-44
- [39] http://www.archcalc.cnr.it/indice/Suppl_7/05_Niccolucci.pdf
- [40] Palmieri, A., 1969 - Recenti dati sulla stratigrafia di Arslantepe, *Origini* III: 1-76.
- [41] ADS <http://archaeologydataservice.ac.uk/>
- [42] ARIADNE <http://www.ariadne-infrastructure.eu/>
- [43] MAPPA project <http://www.mappaproject.org/web-gis-2/>
- [44] SITAR <http://www.archeositarproject.it/> - <http://webais.archeositarproject.it/webgis/login.php>
- [45] Vast-Lab – progetto ARIADNE <http://vast-lab.org/progetti/ariadne/>

Digitalizzazione delle risorse storico-artistiche e sustainability management a supporto dello sviluppo turistico-culturale in Puglia

Floriana Conte¹, Fabio De Matteis²

1 Università del Salento, Lecce, Italia - floriana.conte@unisalento.it - flor.conte@hotmail.it

2 Università del Salento, Lecce, Italia - fabio.dematteis@unisalento.it

ABSTRACT

La proposta è rivolta ad arricchire in modo significativo e innovativo l'offerta per il turismo culturale e scolastico pugliesi attraverso l'applicazione pratica della catalogazione virtuale dei beni culturali di età moderna riferibili alla Puglia sulla base di un approccio sostenibile. Essa parte da: "ASAP – Archivio digitale storico-artistico pugliese" (<http://edvara2.uniud.it/asap-archivio-digitale/>) e "Multidimensionalità, misurazione e valorizzazione della sostenibilità nella Pubblica Amministrazione" (progetti FIR-FutureInResearch in corso diretti presso i Dipartimenti di Beni culturali e di Scienze dell'Economia dell'Università del Salento rispettivamente da Floriana Conte e Fabio De Matteis). ASAP- è un archivio digitale, iconografico e documentario con accesso gratuito che ricompone immagini e fonti per la storia dell'arte in Puglia in età moderna. Il database comprende le sezioni Gli artisti, Le opere, I luoghi, Le iconografie, Le fonti, Gli studi. Quest'ultima vedrà l'implementazione della digitalizzazione in PDF di monografie, articoli scientifici e di giornale, articoli apparsi su blog, guide e letteratura di viaggio relativi alle opere d'arte schedate. Si potrebbero programmare itinerari organizzati per periodo storico e/o area geografica (per esempio, "il Cinquecento veneto a Lecce e provincia", "Le tracce di Caravaggio da Milano a Polignano a mare" ecc.) e scaricare gratuitamente bibliografia ed immagini. Il sito offrirebbe suggerimenti circa percorsi turistico-culturali gravitanti attorno alle opere digitalizzate basandosi sulla relazione tra i dati immessi e il territorio pugliese con rinvio a ristoranti, hotel, attrazioni commerciali.

La somministrazione di questionari di customer satisfaction consentirebbe di valutare la sostenibilità nel tempo dell'offerta culturale e turistica prevista. Ciò richiederebbe la definizione delle modalità di somministrazione, di raccolta e codifica dei dati, di analisi ed interpretazione dei risultati, di stesura del report finale sulla qualità percepita, di prescrizione di policy e di azioni correttive. Si valuterebbero ulteriori indicatori per la misurazione della sostenibilità turistico-culturale, al fine di definire un insieme organico di misure della sostenibilità.

PAROLE CHIAVE

Cultural Heritage, Database, FutureInResearch, History of Art, Sustainability Management, Tourism.

La proposta è rivolta ad arricchire in modo significativo e innovativo l'offerta per il turismo culturale e scolastico pugliesi attraverso l'applicazione pratica della catalogazione virtuale dei beni culturali di età moderna riferibili alla Puglia sulla base di un approccio sostenibile. Essa parte da due progetti già in corso: "ASAP – Archivio digitale storico-artistico pugliese" (<http://edvara2.uniud.it/asap-archivio-digitale/>) e "Multidimensionalità, misurazione e valorizzazione della sostenibilità nella Pubblica Amministrazione" (progetti FIR-FutureInResearch diretti presso i Dipartimenti di Beni culturali e di Scienze dell'Economia dell'Università del Salento rispettivamente da Floriana Conte e da Fabio De Matteis). L'integrazione tra progetti FIR, esplicitamente auspicata dalla Regione Puglia, amplia l'efficienza degli stessi e si affianca a già esistenti piani regionali con ricadute sul territorio: il Polo bibliomuseale salentino, il Piano ufficiale della Regione Puglia per una nuova era dei musei, Puglia digital library.

La base materiale è offerta da ASAP, archivio digitale, iconografico e documentario con accesso libero e gratuito che permette la ricomposizione virtuale di immagini e fonti per la storia dell'arte in Puglia in età moderna, ricostruita nella prospettiva dei rapporti tra Puglia, Regno di Napoli e area settentrionale dalla fine del Quattrocento al Settecento. Il database è organizzato nelle sezioni Gli artisti, Le opere, I luoghi, Le iconografie, Le fonti, Gli studi. Quest'ultima sezione mira all'implementazione della digitalizzazione in formato PDF e con link di rinvio della bibliografia relativa a ognuna delle opere d'arte schedate (monografie e articoli scientifici, articoli di giornale, articoli apparsi su blog, guide e letteratura di viaggio ecc.).

La pagina web è realizzata con e-Dvara², piattaforma di gestione dei dati sviluppata dal Dipartimento di Informatica e Matematica dell'Università degli Studi di Udine e con la collaborazione del LIDA_FOTOteca - Laboratorio informatico per la ricerca storico artistica del Dipartimento di studi umanistici e del patrimonio culturale dello stesso Ateneo. La struttura informatica (il software e-Dvara2) è già stata sperimentata in collaborazioni di portata ampia con Soprintendenze ed Enti dislocati in luoghi diversi del territorio nazionale deputati alla tutela e alla valorizzazione dei Beni culturali con i mezzi digitali: per fare un solo esempio, nel programma informatico Aristos (Archivio Informatico per la Storia della Tutela delle Opere Storico artistiche: 2002) sono state coinvolte le soprintendenze del Friuli, del Veneto, della Basilicata e della Puglia.

La messa a pieno regime di questo archivio consentirebbe di incrementare e valorizzare concretamente la rete fra biblioteche pubbliche, musei civici e grande pubblico attraverso la digitalizzazione del patrimonio bibliografico, museale e paesaggistico (per la questione dei diritti di riproduzione delle opere d'arte, pubbliche e private, italiane e straniere, sono già stati avviati contatti con la Soprintendente all'Archeologia, Belle Arti e Paesaggio per le Province di Brindisi, Lecce e Taranto, Maria Piccarreta). Il sito, inoltre, offrirebbe suggerimenti circa percorsi turistico-culturali gravitanti attorno alle

opere digitalizzate, istituendo relazioni sistematiche tra i dati immessi nel database e il territorio pugliese. Si prevede un rinvio (tramite box apposito) a ristoranti, hotel, attrazioni commerciali (negozi di artigianato tipico di vario livello: ceramiche, oreficerie, prodotti enogastronomici ecc.) ed altri luoghi di interesse culturale dislocati sul percorso collegato a un'opera. L'adattamento della pagina web di ASAP alle esigenze dell'utenza (sia locale, sia scolastica, sia turistica) permetterebbe anche il coordinamento virtuale con la rete di bookshop vetrine del Made in Puglia, offrendo riproduzioni di opere d'arte, indicazioni di libri su di esse e inoltre su cibo, design, artigianato artistico. Sarebbe del tutto inedita e attrattiva l'opportunità di pre-informarsi sull'itinerario più rispondente ai propri gusti o desideri attraverso il download gratuito dal database della bibliografia, delle immagini, del percorso suggerito e più in generale di molte informazioni utili e funzionali.

In questa prospettiva di fatto risulterebbe notevolmente arricchita l'offerta per il turismo culturale a vari livelli, consentendo ai viaggiatori di programmare a distanza itinerari mirati, organizzati per periodo storico e/o area geografica. Ecco alcuni esempi, realizzabili con facilità: "Il Cinquecento veneto a Lecce e provincia", "L'architettura lombarda tra Quattro e Seicento nel Salento", "Le tracce di Caravaggio da Milano a Polignano a mare". Fino ad oggi non esistono proposte che diano risalto contemporaneamente ai diversi aspetti turistico-culturali di un luogo pugliese.

Gli itinerari proposti potrebbero arricchirsi con la segnalazione di luoghi utilizzati o utilizzabili come scenari idonei a film, fiction e serie TV: oggi in Puglia la tutela, il recupero, la valorizzazione dei centri storici nelle città e nei piccoli borghi passa anche attraverso l'interazione tra turismo culturale tradizionale e cineturismo, se si tiene conto del vasto uso che l'industria televisiva e cinematografica sta facendo dei nostri territori. Lo dimostrano, tra le numerosi riflessioni scientifiche e divulgative sul tema, il workshop *Placetelling e valorizzazione dei borghi* che ha coinvolto alcuni studiosi di Unisalento (Fabio Pollice, Beatrice Stasi e Luca Bandirali) lo scorso 24 novembre a Cisternino nell'ambito della V conferenza de "I borghi più belli dell'area del Mediterraneo". In tale prospettiva, l'itinerario "Le tracce di Caravaggio da Milano a Polignano a mare" (prima prospettato) potrebbe facilmente estendersi ad una nuova formulazione del tipo "Le tracce di Caravaggio da Milano a Polignano a mare nel cinema da Mario Monicelli a Marco Ponti", collegando un'ipotesi di viaggio culturale di livello elevato ai gusti di un pubblico ampio che apprezza prodotti cinematografici di più facile consumo.

La fruizione virtuale del patrimonio culturale da parte di pugliesi e non pugliesi interessati all'arte e al turismo in Puglia sarebbe incoraggiata da strumenti e adeguate attrezzature specifici del database. Attraverso questi mezzi e tramite apposite app sviluppate per smartphone e tablet, si potrebbero consultare sia le fonti possedute dalle biblioteche partner sia quelle presenti in altre sedi e digitalizzate per offrire un quadro informativo più esteso. Le informazioni presenti nella pagina web sarebbero erogate in italiano e in inglese. Per coloro che si trovano in condizioni di disabilità si prevederebbero itinerari basati sull'assenza di barriere architettoniche e sulla presenza di strutture ricettive adeguate presenti sul territorio. In questi campi c'è ancora molta strada da fare, a cominciare dal settore dei trasporti pubblici, che oggi spesso non consentono ai disabili di spostarsi agevolmente e in maniera autonoma sul territorio (al contrario di quanto accade in altre regioni, in particolare in alcune nel nord Italia).

Lo sviluppo di ASAP come piattaforma sinergica destinata ai viaggiatori della cultura sostenibile in Puglia dispiegherà tutte le proprie potenzialità se la presente proposta susciterà manifestazioni di interesse da parte degli Enti e delle Istituzioni potenzialmente interessati: Università, Soprintendenza, Comune, Regione e partner di essa, biblioteche.

Appare fondamentale la ricerca della sostenibilità dell'offerta culturale proposta dal progetto "*Digitalizzazione delle risorse storico-artistiche e sustainability management a supporto dello sviluppo turistico-culturale in Puglia*", intesa come capacità di tale innovazione nella fruizione culturale di permanere nel tempo, unitamente alle ricadute socio-economiche auspiccate.

Il perseguimento della sostenibilità è previsto in maniera duplice:

1. attraverso il "supporto" che alle opere rese disponibili nel database viene da parte dell'offerta più ampia alle stesse collegata, come poc'anzi illustrato (itinerari eno-gastronomici nei luoghi in cui l'opera è o era collocata, patrimonio culturale localizzato nel territorio in cui l'opera è esposta o da cui proviene, location cinematografiche correlate alle opere catalogate per vicinanza territoriale, ecc.);
2. attraverso la ricaduta culturale che la fruizione delle opere catalogate ha sull'utenza.

Tanto il supporto quanto la ricaduta hanno come fulcro l'utenza delle opere catalogate, la quale (tanto in merito alla fruizione virtuale, quanto relativamente a quella reale) può essere distinta in almeno due macro-categorie principali di fruitori:

- l'utente-turista, inteso come il visitatore sia locale che straniero che visita per sua propria iniziativa il database o l'opera dal vivo;
- l'utente-studente, che visita il database o l'opera dal vivo su impulso dell'istituzione alla quale appartiene nelle sue vesti di studente (la scuola ha come suo fine istituzionale quello di veicolare il sapere e, pertanto, diventa interlocutore privilegiato dell'archivio digitale).

Da qui la necessità di rendere sostenibile l'offerta attraverso processi mirati per l'una o per l'altra categoria di utenti. Per l'utente-turista, come detto, si prevedono percorsi correlati all'opera d'arte ma anche di altra natura (paesaggistici, cinematografici, gastronomici, commerciali, ecc.) che consentono allo stesso di conoscere, insieme all'opera, anche il territorio pugliese.

Per l'utente-studente la pagina web contenente l'archivio digitale suggerirà, per ciascuna opera catalogata, spunti per progetti mirati all'approfondimento da parte degli studenti di diversi aspetti correlati all'opera stessa ed al territorio. Si pensi, a puro titolo esemplificativo:

- allo studio della moda, al tempo della realizzazione di un quadro sulla base degli abiti e degli accessori nello stesso rappresentati, o alla conoscenza dell'economia, del clima, delle usanze dell'epoca in funzione del soggetto ritratto;
- allo sviluppo di testi teatrali inerenti alla scena rappresentata nell'opera o alla vita del suo autore;
- all'implementazione di percorsi di alternanza scuola-lavoro presso aziende operanti nell'ambito del restauro o dell'arte con le quali si intendono sviluppare rapporti di partnership. Tali percorsi consentono agli studenti di formarsi sul campo tanto da un punto di vista strettamente artistico (se l'alternanza si realizza in aziende ed enti che si occupano di tutela, conservazione, restauro), sia relativamente a tutto ciò che rientra nell'offerta culturale (se l'alternanza avviene presso musei, pinacoteche, teatri, ecc.);
- allo studio della tecnica pittorica/scultorea utilizzata in un'opera digitalizzata anche attraverso la riproduzione di tale tecnica (laddove possibile) da parte degli studenti di istituti d'arte;
- all'approfondimento della tecnica originale comparata alle evoluzioni tecniche moderne effettuando una riproduzione dell'opera con tecniche in uso attualmente in un'ottica storicamente comparativa.

Seguendo i suggerimenti contenuti nel sito web, gli studenti – accompagnati in questo percorso dai loro docenti – avranno l'opportunità di un approccio innovativo ed interattivo allo studio delle opere catalogate nell'archivio digitale.

Tutti questi supporti che il sito web fornirebbe all'utenza (sia turistica che studentesca) renderebbero più attrattiva l'opera e consentirebbero alla stessa di veicolare più efficacemente le conoscenze a vantaggio dei fruitori: si tratta di elementi che – se opportunamente presidiati - favoriscono nel medio-lungo periodo l'offerta culturale innovativa prevista nel presente progetto e, in ultima analisi, ne favoriscono la sostenibilità.

Riferimenti prioritari nella veste di incubatori del presente progetto sono gli enti locali, considerato il ruolo di detentori di una parte rilevante del patrimonio culturale, oltre che il loro ruolo istituzionale di governo del territorio. Gli enti locali, peraltro, possono efficacemente fungere da intermediari aggreganti rispetto a tutti i soggetti coinvolti dal progetto (per esempio: organizzazioni rappresentative di aziende operanti nel settore turistico-ricettivo, scuole, aziende non profit operanti nel settore culturale, aziende informatiche che possono sviluppare app per la fruizione dell'offerta culturale proposta, ecc.).

Diventa pertanto cruciale, per una efficace realizzazione del progetto, non solo coinvolgere gli enti locali, ma cercare di far inserire il progetto (e le sue potenziali ricadute) nel ciclo di programmazione e controllo dei Comuni (in modo che lo stesso possa avere ricadute positive anche internamente all'ente stesso).

Inoltre, quale che sia l'utenza (turistica e/o scolastica) del sito web che ospita l'archivio digitale opportunamente sviluppato, un elemento indispensabile per valutare la sostenibilità nel tempo dell'offerta culturale e turistica prevista dalla presente idea progettuale è rappresentata dalla rilevazione della *customer satisfaction* (peraltro prevista per gli enti locali dalla Legge n. 213 del 2012). La somministrazione di questionari (sia on line sia cartacei presso le istituzioni coinvolte) misurerebbe la soddisfazione dell'utenza che ha usufruito delle informazioni di carattere culturale e turistico fornite dal progetto. Ciò richiederebbe: opportune modalità di somministrazione, di raccolta e codifica dei dati, di analisi ed interpretazione dei risultati, di stesura del report finale sulla qualità percepita; individuazione di policy e di azioni correttive derivanti dall'attività di rilevazione della *customer satisfaction* indispensabili per migliorare il funzionamento dell'ente locale in ambito culturale, per aumentare, di conseguenza, la soddisfazione dei cittadini e, in ultima analisi, per consolidare il consenso verso l'operato dell'amministrazione. Insieme ai questionari si valuteranno ulteriori indicatori per la misurazione e valutazione della sostenibilità turistico-culturale, al fine di definire un insieme organico di misure della sostenibilità.

Infine si prevede un'adeguata formazione (in merito a principi e strumenti utilizzabili) rivolta ai soggetti che, negli enti locali coinvolti nell'implementazione del progetto, sarebbero coinvolti nello stesso.

Risulta evidente la portata e la complessità del progetto dal momento che, partendo dalla digitalizzazione di riproduzioni di opere d'arte e delle fonti e degli studi a esse correlate, auspica una ricaduta la cui applicazione culturale sul territorio sarebbe anche pratica, mettendo in rete i principali attori pugliesi.

BIBLIOGRAFIA

- [1] *Alla scoperta del territorio: percorsi alternativi o complementari; per un nuovo turismo culturale*, Atti di un convegno europeo (20 - 21 febbraio 2003, Palazzo Incontri, Firenze), a cura di Barbara Tosti, Firenze, Edizioni Polistampa, 2003.
- [2] Axelsson, R., Andersson, K., *Social and Cultural Sustainability: Criteria, Indicators, Verifier Variables for Measurement and Maps for Visualization to Support Planning*, «Ambio», Vol. 42, 2013, pp. 215-228.
- [3] Ball, A. and Bebbington, J., *Accounting and Reporting for Sustainable development in Public Service Organizations: issues and emerging directions*, «Public Money and Management», 28(6), 2008, pp. 323-325.
- [4] Ball, A., Grubnic, S., *Sustainability accounting and accountability in the public sector*, in UNERMAN J., BEBBINGTON J., O'DWYER B., *Sustainability Accounting and Accountability*, London, Routledge, 2007.
- [5] Ballo Alagna, S., *Valorizzazione del paesaggio culturale e turismo sostenibile: il caso Mistretta e del suo centro storico*, «Archivio storico messinese», 82, 2001, pp. 105-123.
- [6] Barocchi, P., *Premessa*, in *Ricerca scientifica e sviluppo: beni culturali e nuove professionalità*, Atti delle Giornate di studio (23 novembre - 1° dicembre 1996, Il Palazzone di Cortona), a cura di Paola Barocchi, Pisa, Scuola Normale Superiore, 1996 (Centro di Ricerche Informatiche per i Beni culturali, Quaderni, VI), pp. 9-10.
- [7] Barocchi, P., *Storia dell'arte ed elaborazione automatica dei dati*, in *1° Congresso nazionale di Storia dell'arte* (CNR - Roma,

- 11-14 settembre 1978), a cura di Corrado Maltese, Roma, Consiglio Nazionale delle Ricerche, 1980, pp. 181-186.
- [8] Barocchi, P., VII. *Documents and sources*. 21, in *Automatic Processing of Art History Data and Documents*, Proceedings (Pisa, Scuola Normale Superiore, Spetmber 24-27, 1984), edited by Laura Corti and Marilyn Schmitt, Scuola Normale Superiore-The J. Paul Getty Trust, Published by Regione Toscana, 1984, pp. 235-253.
- [9] Bebbington, J. and Gray, R., *An Account of Sustainability: Failure, Success and a Reconceptualisation*, «Critical Perspectives on Accounting», 12, (5), 2001, pp. 557-605.
- [10] Canato, C., *Le ville venete: la valorizzazione turistica di un prodotto culturale*, Sommacampagna, Verona, Cierre Edizioni, 2013.
- [11] Conte, F., *Il contributo della catalogazione alla fruizione del sistema museale pugliese*, relazione al convegno internazionale *Musei*, in *Puglia, tradizione e futuro*, a cura di Raffaele Casciaro (Bari –Lecce, 17-18 novembre 2016), in *Le risorse digitali per la storia dell'arte moderna tra Puglia e Italia settentrionale: progetti e proposte a confronto*, Atti del seminario di studio (Lecce, ex monastero degli Olivetani, Padiglione Chirico, 24-25 maggio 2017), a cura di Floriana Conte, Roma, Edizioni di Storia e Letteratura (Collana "Biblioteca dell'Accademia dell'Arcadia" sottoposta ad approvazione di Comitato scientifico e referaggio anonimo), in corso di stampa.
- [12] De Matteis, F., Preite, D., *Il ciclo di sostenibilità negli enti locali*, Torino, Giappichelli Editore, 2015.
- [13] De Matteis, F., Preite, D., Striani, F., "Cultural Sustainability: some Reflections on the Financial Role of the Italian Local Government", «Journal of Public Administration, Finance and Law», vol. 11, 2017, pp. 7-16.
- [14] Dumay, J., Guthrie, J., Farneti, F., *GRI Sustainability Reporting Guidelines for Public and Third Sector Organizations*, «Public Management Review», Vol. 12, Issue 4, 2010, pp. 531-548.
- [15] Farneti, F. & Guthrie, J., *Sustainability Reporting by Australian Public Sector Organisations: Why They Report*, «Accounting Forum», 33, (2), 2009, pp. 89-98.
- [16] Kaur, A. and Lodhia, S., *The state of disclosures on stakeholder engagement in sustainability reporting in Australian local councils*, «Pacific Accounting Review: Special issue on Sustainability Accounting and Reporting», 26, (1/2), 2014, pp. 54-74.
- [17] *Le risorse digitali per la storia dell'arte moderna tra Puglia e Italia settentrionale: progetti e proposte a confronto*, Atti del seminario di studio (Lecce, ex monastero degli Olivetani, Padiglione Chirico, 24-25 maggio 2017), a cura di Floriana Conte, Roma, Edizioni di Storia e Letteratura (Collana "Biblioteca dell'Accademia dell'Arcadia" sottoposta ad approvazione di Comitato scientifico e referaggio anonimo), in corso di stampa.
- [18] Rogers, P.P., Jalal, K.F., Boyd, J.A., *An introduction to sustainable development*, Earthscan, London 2008.
- [19] Soini, K., Birkeland, I., *Exploring the scientific discourse on cultural sustainability*, «Geoforum», Vol. 51, 2014, pp. 213-223.
- [20] Strange, T., Bayley, A., *Sustainable development. Linking economy, society, environment*, Paris, OECD Publishing, 2008.
- [21] Throsby, D., *On the Sustainability of Cultural Capital*, Sidney, Macquire University Department of Economics, Research Paper 10/2005.
- [22] Tomasin, L., *L'impronta digitale*, Roma, Carocci, 2017.
- [23] World Commission on Environment and Development, *Our Common Future*, Oxford University Press, Oxford, U.K., 1987.

Do alchemists dream of electronic sheets?

Ilaria Cicola

Alma Mater Studiorum - University of Bologna, Italy - ilaria.cicola@gmail.com

ABSTRACT

How and to what extent can Digital Humanities (DH) contribute to the understanding of Medieval Arabic Alchemy, and in particular to the study of the corpus traditionally attributed to the most important, prolific and controversial figure of this field, namely Jābir Ibn Ḥayyān? How can research benefit from the enhanced memory provided by digital support? In this paper, where the Digitized Jabirian Corpus (DJC) is presented, I will try to show how DH can change how we approach this very specific and challenging field of study. I will begin by presenting the main specificities and problematics of the texts, problematizing both the linguistic and the historical aspects of the state of the art in this field of study, taking into consideration different significant views on the topic and trying to approach the problematic of memory in this subject. Then, I will proceed to an explanation of the DJC and the methodologies used both to digitalize it in a TEI-compliant format and to lemmatize it with Natural Language Processing (NLP) tools. I will analyze the challenges encountered and the limits of the tools used, such as AlKhalil lemmatizer. I will then present the results accomplished by my research, which has mostly focused on lexical analysis and Information Retrieval (IR) until now. Afterwards, I will try to understand how future enhancements and developments can lead to (i) major improvements in both the conservation and dissemination of the texts and to (ii) enlightening discoveries about authorship and text organization. Lastly, I will show how this research can benefit from new structural and syntactical analysis conducted on the corpus by means of the Univesal Dependencies (UD) model.

KEYWORDS

Arabic computational linguistics; Corpus Linguistics; Arabic Linguistics; Data Mining; Information Retrieval; Applied linguistics; Arabic Alchemy

1. INTRODUCTION

Arabic alchemy, a scientific activity that flourished between the VII and the XI century, represents an interesting field of study of the medieval Arabic thought: its texts are complex and the specific lexicon used is arduous to interpret. The most prominent Arab alchemist, Jābir Ibn Ḥayyān is considered the author of thousands of texts where the alchemical knowledge is scattered within a plethora of books (a practice called *tabdīd al-‘ilm* ‘the dispersion of the science’) in order to preserve it amongst experienced alchemists. Scholars like Berthelot [1], Wiedemann [24], Ruska [19], Holmyard [8], Kraus [10][11], Corbin [3], and lastly Lory [13] [14] [15] and Haq [7] gathered, edited and commented alchemical texts ascribed to various authors. However, despite the interest in this domain, yet there is the need for tools and a digitized corpus of this discipline¹. The realization of a digitized corpus (§4) implied studies about the authorship of the texts (§3), lexical analysis (§5) and the questions on integrity of the corpus (§6). The complexity of such a corpus, together with the study of its specific features have been the foundation to create tools fit for understanding and easily accessing the wide range of texts related to the Great Work² and represent a challenging use of Digital Humanities in Cultural Heritage.

2. HISTORICAL REMARKS

The corpus gathered in this study is ascribed to Jābir Ibn Ḥayyān, who supposedly lived in VIII century and was the disciple of the sixth Shi’i Imam, nevertheless his authorship has been and still is questioned by scholars. The datation being an issue in understanding both the theories expressed in the corpus and the cultural evolution with its transmission in the Medieval era, it is important to highlight here the different opinions about the authorship of the corpus. Lory [13:15] stresses that the presence of “*éléments ultra-chiïtes [...] dans le corpus [est] un indice flagrant que la date de leur rédaction ne pouvait être antérieure à la fin du IX siècle*”. He bases his view on Jābir’s appartenance to the extremist Shi’a (*ḡulāt*). Kraus underlines the utilisation of Greek loanwords as further evidence of the late dating [11]. Moreover, he states that the use of the ‘dispersion of the science’ is clue to the time discrepancy between the different texts of the corpus, making it so that it must be ascribed to multiple authors. Sezgin [20] and Haq [7] think differently: the first tries to tie the life of Jābir to the period of Ja‘far al-Šādiq, stating that he lived between 725 and 812. Haq follows this theory, challenging Kraus’ opinions, and states that the corpus belongs to a single author who lived in the VIII century, arguing that the translations from Greek had already started at that time. Haq considers the *tabdīd al-‘ilm* proof of the unity and continuity of the corpus, thus supporting the theory of an unique author. Lastly, Corbin and Lory embrace Kraus’ theories and Lory ultimately suggests that several layers of writing are present in the corpus: “*autour d’un noyau primitif, d’un enseignement alchimique à caractère pratique, ont été ajoutées des considerations plus théoriques et doctrinales, et enfin des introductions,*

¹ The largest Arabic corpus available can be found at <http://arabicorpus.byu.edu/>.

² Translation of the Latin *Magnum Opus*, this term describes the Alchemic work of both prime matter transmutation and personal evolution. It was first used to describe the Hermetic tradition practices, and then it was extended to other traditions, like the Arab one.

transitions et références reliant les différentes parties entre elles à l'ensemble de l'oeuvre jâbirienne"[15:21]³. This solution might be supported by the use of DH (§6).

3. THE DIGITIZED JABIRIAN CORPUS

The TEI P5 [22] compliant Digitized Jabirian Corpus (DJC) is based on three sources: i) *Names, Nature and Things*, which gathers texts from the *Book of Stones* edited by Haq [7], ii) *Muhtār Rasā'il* ('Selected Studies'), an anthology edited by Kraus [10], and iii) *L'élaboration de l'Elixir Suprême* ('The elaboration of the Supreme Elixir'), a collection edited by Lory [9]. The decision to only include edited texts was made to ensure the integrity of the corpus and to avoid major paternity issues; footnotes were excluded from the digitalization to avoid information overflow. Paragraphs are used as the basic unit for the analyzing process, since they are considered conceptual units in the DJC, therefore every <p> is numbered, so to provide exact *loci* for internal references.

The most crucial part of the digitalization was to render the peculiarities of the corpus, briefly summarized in the following list:

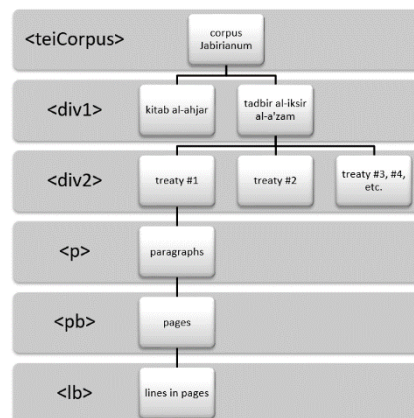


Fig.1 The TEI structure of the DJC

i) *Alchemical technical language*: no previous work has ever dealt with the tagging of the alchemical lexicon⁴. Clarity in this matter is needed and a wide corpus could prove an efficient testing ground for the definition of the nature of alchemical terms in Arabic. This work acknowledged only the core elements, chosen for their recurrence and relevance both in the corpus and in academic literature.

E.g. <term type="alch" translation="and the copper"> والنحاس </term>

ii) *Loanwords and calques*: Arabic alchemy flourished in a multilingual setting. The Great Art was practiced in Greece before arriving in the Arab world [10] [11], hence the presence of calques and loans in the formation of the Arabic alchemical lexicon (e.g. *falsafa* from the Greek φιλοσοφία 'philosophy'). Moreover Persian-related lexicon can be found (e.g. *jawhar* 'substance', borrowed from the Pahlavi *gōhr*).

E.g. <term type="alch" translation="the substance" xml:lang="fa" subtype="loan"> الجوهر </term>

iii) *Synonymy* is a fundamental characteristic of the corpus, as it represent the realization of the *mizān al-ḥurūf* ('the balance of the letters', a criterium similar to the Kabbalah, that assigns a specific value to every letter, hence each synonym relates to a specific weight of the same element).

E.g. <term type="alch" translation="tin" sameAs="ورطب_حر زاوس المشتري القلعي"> القلعي </term>

iv) *Polysemy* is another pivotal feature of the corpus. Words change meaning in relation to the context they are inserted in, as Lory clearly describes, "*Le langage alchimique exprime [...] la représentation unitaire d'un monde multidimensionnel*" [9:25]⁵.

E.g. <term type="alch" ana="#balance #day of judgment #specific gravity #metaphysical principle of monism"> ميزان </term>

v) *Quotes*: the encyclopedic style of Medieval Arabic scholars' writings explains the conspicuous use of quotes, which can occur in two styles: a) simple reference to the author (e.g. *alà ra'ī Balinās* 'following Apollonius view'); or b) textual quotes from books (e.g. The Koran). The only problem encountered with this tag is that <quote> can't overlap <p>, because of hierarchy issues.

E.g.: قال بليناس

<quote type="author" ana="Apollonius of Tyana">

أقول وأصفُ الحكمة التي أَدتُّ بها بعد خروجي من السَّرْب وأخذ

<lb n="7" ed="H"/>

³ "around an original core of practical Alchemic teachings, were [later] added more doctrinal and theoretical reflections, and finally [were added] introductions, transitions and references connecting the different of the Jābirian work to each other".

⁴ Ferrario [5][6] studied the different synonyms for gold and mercury, but did not try to tag them in an automatic lexicon. Scotti [21] created a system for tagging alchemical recipes, but his scope was to render the structure of the texts only.

⁵ "Alchemical language expresses [...] the unitary representation of a multidimensional world"

الكتاب واللوح. إن الذي يعمّ الأشياء
كلّها الطبائع التي هي البسيطة لا
المركبة، وإذا

<lb n="8" ed="H"/>

كان الشيء عاماً فمجال أن لا يكون له
</quote> - كميّة

vi) *Hypertextuality*: the *tabdīd al-ilm* (the aforementioned technique of scattering pieces of information in many different texts, so to challenge the adept to be fully aware of the totality of the writings and to become a master in information retrieve and use), is both one of the most significant aspects and one the most challenging obstacles in the understanding of the Great Work texts, because of the reader's impossibility to interlink texts or to materially access them. With this perspective in mind, this work proposes a new approach to the interaction with the corpus, using the digitalization as a tool to destructure it and to grant easy access to the information. Since there are striking analogies between the operating principles of the Jabirian corpus and of the contemporary hypertexts [12], the works so far done in this direction points out to the possibility to treat this corpus as an hypertext.

E.g.

<div2 type="treaty" n="4">
الجزء الأول من كتاب الاحجار على رأي
بليناس</div2>
<ref target="#1.1"/>

In the tag <ref> the attribute "target" specifies where the reference is located. Of course, the corresponding book has been tagged with an according ID:

<div2 @xml:id="1.1" type="section"
n="1">
الجزء الأول من كتاب
الاحجار على رأي بليناس</div2>

4. LEMMATIZATION PROCEDURES AND RESULTS

Lemmatization's first step was to extract all the words from the XML source file by a Python routine which excluded the tags, creating a raw version of the corpus. After a first trial analysis with AraMorph [2], it was evident that the lexical database didn't align with the DJC⁶, and we decided to switch to AlKhalil [16] [17] lemmatizer. This morphosyntactic parser has a more classical orientation both in its result analysis and database lexicon, making it more suitable for the DJC. The results XML file produced by AlKhalil with the multiple analysis of every word of the DJC was then processed with a Python script in order to create an XLS file in which the analysis could be displayed, chosen and corrected if needed. This file was created merging the corpus information (i.e. editor, page, line, paragraph), context (phrase) and the words' analysis.

Ed	Page	line	Paragr	Phrase	word #:	Arabic	Analysis	corr	note
94357	L	7	1	161	1	بسم الله الرحمن الرحيم	بسم		
94358	L	7	1	161	2	بسم الله الرحمن الرحيم	الله		
94359	L	7	1	161	3	بسم الله الرحمن الرحيم	الرحمن		
94360	L	7	1	161	4	بسم الله الرحمن الرحيم	الرحيم		
94361	L	7	2	161	1	الحمد لله شكرا لما وهب لنا ومنه	الحمد		
94362	L	7	2	161	2	الحمد لله شكرا لما وهب لنا ومنه	له		
94363	L	7	2	161	3	الحمد لله شكرا لما وهب لنا ومنه	شكرا		
94364	L	7	2	161	4	الحمد لله شكرا لما وهب لنا ومنه	لما		
94365	L	7	2	161	5	الحمد لله شكرا لما وهب لنا ومنه	وهب		
94366	L	7	2	161	6	الحمد لله شكرا لما وهب لنا ومنه	لنا		
94367	L	7	2	161	7	الحمد لله شكرا لما وهب لنا ومنه	ومنحنا		
94368	L	7	2	161	8	الحمد لله شكرا لما وهب لنا ومنه	من		
94369	L	7	2	161	9	الحمد لله شكرا لما وهب لنا ومنه	فضله		
94370	L	7	2	161	10	الحمد لله شكرا لما وهب لنا ومنه	بغير		

Fig. 2 A screenshot of the XLS file used for the lemmatization

⁶ For an interesting and complete example of implementation of Aramorph, see Nahli (2013) [18].

AlKhalil proposed multiple analysis for 92,5% of the words in the corpus and this percentage might be considered higher if typing errors are taken into account. Main issues with the AlKhalil lemmatizer are: i) fluctuation in the analysis of the *hamza*, which affected the analysis of certain verbal forms (e.g. *af'ala*,) superlatives, and sometimes of the imperative verbs; ii) numerals, which are often not correctly analyzed, (e.g. the uninflected word *ašara* 'ten' in compounds like *itnā- 'ašara* 'twelve'); iii) compounds built on prepositions (e.g. *kaylā* 'lest') were not always correctly analyzed; and lastly iv) some words were not included in the lemmatizer database (e.g. *sirgīn* 'manure', *bin* 'son', *lāhūt* 'deity'). We had to cope with these problems, normalizing the texts so to be able to analyze them and sometimes adding to the database new entries. This notwithstanding, the overall result of the analysis is definitely accurate.

The lemmatization of the corpus allowed for the creation of automatically generated concordances of the corpus. The concordances are sorted in Arabic alphabetical order, by root, lemma and word form. A superscript number on the lemma represents the number of occurrences of the word in the corpus. The word (highlighted in bold) is then listed and its context displayed together with its placement in the corpus. The aim of the creation of concordances is to give the possibility to other researchers to approach the corpus from different perspectives, so that it may be possible to seize new nuances in the meanings or in the overall structure in future developments of the project.

عثر	عثر: عثر الذي هذا الكتاب أهداه، وجميع ما في هذا الكتاب L.22.115
أخذها	أخذها: أخذها : عثان أهداه . وهذا فيه ما تطم بما قد تقدم من الشرط L.1.111
أخذ	أخذ: أخذها : فيه تديريين، أهداهما مقدم قد يضل أهداه، والثاني في التمرقة L.1.111
أجل	أجل: أجل : بوجه التثني وتدخل عليه باقي نفسه وروحه فيبدأ أجل ، ما في L.1.113
أجل	أجل: أجل : فإنه تحتاج إلى الحجر أو لا من أجل أنه تحتاج إلى معرفة فعل لار L.1.113
أخذ	أخذ: أخذ : الأثران العربية مع غيرها أن يكون أهداه ناصدا شعرا ذاتها فيها L.37.113
أخذ	أخذ: أخذت : أخذت: نارا الشجر فادخلت مكان لار الحجر الأظم أو أرغفه أو ماله أو L.1.117
أخذ	أخذ: أخذ : وفي غيره، إن لا يحتاج إلى وقت بعينه، فيبدأ أهداه ما يصحب به L.25.117 ؛ وقد تكررت تلك في أهد الكتاب المائة والاثني عشر وهو كتاب L.3.118
أخذها	أخذها: أخذها : الأثران لهد "الخصال شتى أهداهما إنك تجد فيه من الطبايع L.1.117 ؛ أهداهما الأهد، والثاني المعروف بالتدبير الثاني، وقد تكررتاه L.22.117 ؛ في هذا المعنى والضمح الإعتاد واليقوع الكامل في L.1.117

Fig. 3 A page from the concordances

5. WHAT NOW?

This work gathered in total 103166 words, of which 19471 are word forms; 6370 alchemical words were tagged and represented - when necessary - as synonyms or polysemantic words; 195 quotes and 50 “hypertextual” links were found. These numbers show that the grand total of the DJC is most certainly a good source of data for further analysis. The aim, in the long run, is to add a syntactical analysis, using the Universal Dependencies (UD)⁷ model. Syntactic annotation would allow to deepen and plausibly confirm the theory of Lory on the stratification of the corpus we introduced in §2, thus enhancing the research not only on a textual level, but also providing new insights on the historical sedimentation of the corpus, in facts, by mapping the different syntactic structures we might be able to understand which parts are “core”, “commentaries” and lastly “connection” passages.

As shown, the DJC corpus is a good example of textual processing and conservation of cultural heritage. It illustrates how DH can provide a new type of access to the text, enhancing the capacity of the researcher to explore and read it. In the Digital Era, philology, the practice of “*leggere bene*” [23], should be reconstructed and integrated in a novel scenario: one where memories can be processed without limits of space, nor time. A scenario characterized by both high and low entropy (huge amounts of data, but strictly organized) in which DH can possibly become a sort of *deus ex machina*, and avoid that all relevant information “will be lost in time, like tears in rain”.

6. REFERENCES

- [1] Berthelot, M. 1893. *La chimie au Moyen Age*. Amsterdam: Otto Zeller Osabrück Philo press.
- [2] Buckwalter, T. 2002. *Buckwalter Arabic Morphological Analyzer Version 1.0*, *Linguistica Data Consortium*, University of Pennsylvania, LDC catalog No.: LDC 2002L49
- [3] Corbin, H. 2003 [1981]. *Le livre des sept statues*. Paris: L’Herne
- [4] Ferrario, G. 2007. *Al-Kimiya. Notes on Arabic Alchemy*. In: *Chemical Heritage Review*, 25, 3.
- [5] Ferrario, G. 2009a. *An Arabic Dictionary of Technical Alchemical Terms: MS Sprenger 1908 of the Staatsbibliothek zu Berlin (fols. 3r–6r)*. In: *Ambix: Journal for the Society for the History of Alchemy and Chemistry*, 56, 1, (pp. 36-48).
- [6] Ferrario, G. 2009b. *Understanding the Language of Alchemy: The Medieval Arabic Alchemical Lexicon in Berlin, Staatsbibliothek, Ms Sprenger 1908*. In: *Digital Proceedings of the First Annual Lawrence J. Schoenberg Symposium on Manuscript Studies in the Digital Age*, Vol. 1, No. 1, Article 2. Retrieved from: <http://repository.upenn.edu/ljsproceedings/vol1/iss1/2>.
- [7] Haq, S. N. 1994. *Names, Nature and Things*. Dordrecht: Kluwer Academic Publishers

⁷ <http://universaldependencies.org/introduction.html>

- [8] Holmyard, E. J. 1928. *The Arabic Works of Jābir Ibn Ḥayyān*. New York: EP Dutton.
- [9] Jābir ibn Ḥayyān, Lory P. (ed.). 1983. *Dix traités d'alchimie. Les dix premiers Traités du Livre des Soixante-dix présentés, traduits de l'arabe et commentés par Pierre Lory*. Paris: Sindbad.
- [10] Kraus, P. 1935. *Jābir Ibn Ḥayyān. Muḥtār Rasā'il, Essai sur l'histoire des idées scientifiques dans l'Islam*. Tome I et II. Paris - le Caire.
- [11] Kraus, P. 1942-43. *Jābir Ibn Ḥayyān. Contribution à l'histoire des idées scientifiques dans l'Islam*. Le Caire: Mémoires de l'Institut d'Égypte.
- [12] Lancioni, G. 2008. Variants, links and quotations: classical arabic texts as Hypertexts. In: Bredi, D., Capezzone, L., Rostagno, L. & Dahmash, W. (Eds.) *Scritti in onore di Biancamaria Scarcia Amoretti*, Volume II, Roma: II, (pp. 245-265).
- [13] Lory, P. 1989. *Alchimie et mystique en terre d'Islam*. Paris: Folio Essais.
- [14] Lory, P. 1988. *L'élaboration de l'Elixir Suprême - Quatorze traités de Jābir ibn Ḥayyān sur le Grand Œuvre alchimique*. Damascus: Publications de l'Institut Français d'Études Arabes.
- [15] Lory, P. 2003. *Introduction*. In: Corbin H., *Le livre des Sept Statues* (pp. 9-28). Paris: L'Herne.
- [16] Mazroui A., Lakhouaja A., Meziane A., Ould Abdallahi Ould Bebah M., Boudlal A., Shoul M. 2010. *AlKhalil Morpho Sys 1*. In: Proceeding of the International Arab Conference on Information Technology, Benghazi. Retrieved from: <http://itpapers.info/acit10/Papers/f653.pdf>
- [17] Mazroui A., Boudchiche M., Ould Abdallahi Ould Bebah M., Lakhouaja A., Boudlal A. 2016. *AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer*. In: Journal of King Saud University – Computer and Information Sciences DOI: 10.1016/j.jksuci.2016.05.002.
- [18] Nahli, O. 2013. *Computational contributions for Arabic language processing Part I. The automatic morphologic analysis of Arabic texts*, *Studia graeco-arabica* 3, pp. 195-206. Pisa: Pancini.
- [19] Ruska, J. 1924. *Arabic alchemists*. Heidelberg.
- [20] Sezgin, F. 1967. *Geschichte des Arabischen Schrifttums*, Volume IV. Leiden: Brill.
- [21] Scotti, A. 2003. *Ipotesi per la creazione di un repertorio digitale relativo alle ricette mediche alchemiche*. In: Crisciani, C. e Paravicini Bagliani, A. (eds), *Alchimia e medicina nel Medioevo*, (pp. 337-370). Firenze: Edizioni del Galluzzo.
- [22] Tei Consortium. 2014. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.7.0. Last updated on 16th September 2014, revision 13036. Retrieved from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>
- [23] Varvaro, A. 2012. *Prima lezione di filologia*. Roma-Bari: Laterza.
- [24] Wiedemann, E.G. 1922. *Zur Alchemie bei den Arabern*. Erlangen.

Edizioni digitali di opere a tradizione complessa: il caso del Conte di Carmagnola

Beatrice Nava

Università di Bologna, Italia - beatrice.nava2@unibo.it

ABSTRACT

La seguente proposta di intervento ha lo scopo di presentare un progetto di edizione digitale della tragedia manzoniana *Il Conte di Carmagnola*, concentrandosi sulle specificità filologiche e sulle strategie da adottare a livello di codifica di un'opera a tradizione complessa, che si è potuta recentemente giovare della scoperta di un testimone manoscritto inedito (copia censura). L'obiettivo è quello di discutere una tipologia di *digital scholarly edition* non ancora rappresentata nel panorama delle DH, nel quale non paiono riscontrarsi esempi di edizione in grado di sostenere – in particolare per quanto riguarda la codifica e fruizione dell'apparato – opere con tradizioni di difficile sistemazione critica. Verranno messe in luce, attraverso un breve esempio tratto dall'edizione critica di riferimento della tragedia (Bardazzi 1985), le difficoltà di formalizzazione e rappresentazione dell'apparato, che attente operazioni di codifica consentirebbero di chiarire e semplificare. In particolare si evidenzierà, fornendo anche un esempio di apparato codificato in XML TEI, come le operazioni di *markup* permettano di risolvere problemi di individuazione e resa della diacronia degli interventi variantistici dell'autore, che ritorna a più riprese sui propri manoscritti, modificandoli anche di molto. Ciò permette inoltre di riflettere sui limiti del linguaggio XML TEI per la rappresentazione di casi complessi come quello del Carmagnola, e dunque di molti dei testi centrali della nostra letteratura nazionale.

PAROLE CHIAVE

Edizione critica digitale, *Il Conte di Carmagnola*, filologia d'autore, codifica di apparati complessi, XML TEI, overlapping

1. INTRODUZIONE

Uno sguardo al vasto panorama delle edizioni digitali attualmente realizzate o progettate¹ permette di cogliere come, in questo campo, gli sforzi maggiori si siano finora orientati verso la riproduzione documentale e la possibilità della collazione automatica di testimoni o edizioni. Tali risorse si concentrano in maniera minore sulla fruizione di apparati critici: nei casi, percentualmente non maggioritari, in cui venga presentata una qualche forma di apparato variantistico, esso tende ad essere realizzato solo per tradizioni non eccessivamente ricche o situazioni testuali semplici, ad esempio per opere soggette a rimaneggiamenti correttori puntuali, come varianti sostitutive secche.

Constatata dunque l'attuale carenza di modelli di edizione digitale capaci di sostenere il livello di complessità filologico-critica di molte delle opere del nostro panorama letterario, il presente intervento ha lo scopo di discutere il progetto di edizione digitale di una tragedia caratterizzata appunto da vicende elaborative particolarmente intricate: *Il Conte di Carmagnola* di Alessandro Manzoni.

2. IL CASO DI STUDIO

La scelta dell'opera è motivata da stringenti ragioni filologiche: essa infatti, a seguito dello studio della copia per la censura (MS Ital. 72, Houghton Library, Harvard University), rinvenuta e segnalata nel 1990 (Severino 1990, 3-19), ma rimasta inedita, necessita di un completo lavoro di riedizione. La nuova acquisizione ha permesso di indagare le ultime fasi elaborative della tragedia e di realizzare l'edizione critica dei testimoni più immediatamente coinvolti dalla conseguente redistribuzione variantistica: i due manoscritti braidensi riportanti l'ultima stesura integralmente autografa (Manz.B.X.2, per gli Atti I-V, e Manz.B.VS.XI.4, per le prose introduttive) e l'*editio princeps*, stampata presso la tipografia milanese di Vincenzo Ferrario nel 1820.

Ma il *Carmagnola* è un candidato ideale per la realizzazione di un'edizione digitale soprattutto per la complessità della sua situazione testuale, difficilmente esprimibile mediante supporto cartaceo e che non avrebbe alcun vantaggio dall'essere digitalizzata secondo le modalità dominanti sopra accennate.

La difficoltà maggiore presentata dal testo riguarda l'individuazione della diacronia degli interventi variantistici dell'autore e il rapporto reciproco tra le stratigrafie correttive: Manzoni torna infatti a più riprese sui propri materiali manoscritti, testando su quanto già prodotto modifiche e correzioni in vista di riscritture successive. Questo modo di procedere genera un'articolata stratificazione di fasi elaborative difficilmente distinguibili: la realizzazione di una edizione digitale permette in questo senso una notevole semplificazione, consentendo di suddividere in modo visivamente più agevole il processo elaborativo dell'autore in merito a ciascuna porzione testuale.

3. ESEMPI DI EDIZIONE

Per comprendere meglio la questione si fornisce un breve *specimen* di testo e apparato secondo la formalizzazione realizzata da Giovanni Bardazzi, curatore dell'edizione critica di riferimento del Carmagnola (Bardazzi 1985), e la mia

¹ Si vedano almeno il catalogo di Greta Franzini <http://www.filologiadautore.it/wp/catalogo-delle-edizioni-di-manoscritti-moderni-xvi-xxsec/> e quello di Patrick Sahle <http://www.digitale-edition.de/>. Cfr. anche Italia, Bonsi 2016 e Pierazzo, 2015.

proposta di edizione cartacea. L'esempio, pure non tra i più complessi presentati dall'opera, evidenzia una grande quantità di interventi autoriali su una piccola porzione testuale, fornendo un saggio del lavoro manzoniano.

<p>Testimoni: ManzB.VS.XI.4 = B; MsItal. 72 = Cc ed. princ. = C</p> <p>Alessandro Manzoni, <i>Il Conte di Carmagnola</i>, Milano, Ferrario, 1820, p. 21: [...]f) un terreno paludoso, in mezzo al quale passava una strada elevata a guisa d'argine: e fra le paludi s'alzavano qua e là delle macchie poste su di un terreno più sodo: il Conte pose agguati in queste, e si diede a provocare il nemico [...]</p> <p>Apparato Bardazzi: un <i>prec. da</i> un lungo argine <i>subito biff.</i> in <i>prec. da</i> che non <i>subito biff.</i> passava <i>seg. da</i> un argine <i>subito biff.</i> alta <i>prec. da</i> volta <i>adsubito biff.</i> a guisa <i>seg. da</i> d'argine: paludi> d'argine: il Conte pose imboscate [<i>parola così divisa</i>: imbos-cate] in molte macchie le quali erano sparse qua e là [<i>seguono due tre lettere indecifrabili</i>] > d'argine: le > d'argine: nelle paludi v'erano molte > v'erano qua e là macchie di > macchie e alture [?] di > alture [?] da non <i>subito canc.</i> fra <i>subito calc. su n[elle]</i> paludi <i>seg. da</i> v'erano delle macchie <i>subito biff.</i> SU BP PER C? alta] <i>canc. in favore dielevatasprs.</i> ci pose agguati,] <i>corr. in</i> pose agguati in queste, <i>sprs.</i></p> <p>Apparato Nava: un terreno...queste.] B¹ un lungo argine² un terreno paludoso che ave<va>in mezzo al quale passava un argine una strada^avolta a d<>^balta elevata (<i>sps.</i>) a guisa d'argine: ^afra le mac<chie> ^bil Conte pose imboscate in molte macchie ^cle paludi avevano sparse qua e là d<>^de nelle paludi v'erano ^amolte ^bqua e là macchie ^adi ^be alture ^adi ^bda cui e fra le paludi v'erano delle macchie s'alzavano qua e là delle macchie poste su un terreno più sodo: il Conte vi pose agguati, → pose agguati in queste, Cc=B (<i>ma err. vi pose poi canc.</i>)</p>

Tabella 1: confronto apparati cartacei

Come riportato in **Tabella 1**, Bardazzi frammenta notevolmente le lezioni varianti, rendendo l'apparato difficile da decifrare e impedendo l'individuazione del testo definitivo prodotto dall'ultima stratificazione correttoria. Se l'apparato non è di fatto il luogo deputato alla presentazione immediata del testo finale stabilito, è pur vero che una formalizzazione più attenta alla sua visualizzazione – come avviene nella seconda proposta, che permette di individuare attraverso il corpo maggiore l'ultima stratificazione variantistica – semplifica la lettura dello strumento critico e permette una visione diacronica dell'evoluzione del testo, proponendo la successione degli interventi autoriali fino all'ultimo momento compositivo. Tuttavia, di fronte a situazioni come quella proposta, l'apparato cartaceo rimane comunque di non facile lettura; mentre le operazioni di codifica consentono una gerarchizzazione di fasi più chiara, permettendo di esplicitare la diacronia degli interventi senza perdere le connessioni sintattiche e semantiche delle lezioni con il contesto. La marcatura consente dunque di descrivere analiticamente il processo scritto e creativo dell'autore, operazione necessaria per un'opera così ampiamente rimaneggiata.

Le possibilità di codifica per questo tipo di apparato sono molteplici, e si potrebbe procedere sia stabilendo sezioni relativamente ampie di testo da riferire a un unico apparato, all'interno del quale operare con i marcatori di volta in volta adeguati, sia aprendo un elemento <app> per ogni singola porzione di testo variante. Si propone qui un esempio di questa seconda strategia di marcatura (Parallel Segmentation Method)², che combina, all'interno dell'apparato, introdotto per ciascun lemma, l'elemento <subst> con gli attributi @del e @add. Ciascun intervento dell'autore è numerato in sequenza, mediante un valore numerico progressivo assegnato all'attributo @seq. Viene segnalata anche l'indicazione topografica delle varianti e un esempio di intervento editoriale, con indicazione di responsabilità e relativo grado di certezza.

<pre> <teiHeader> <fileDesc> <titleStmt> <title>Il conte di Carmagnola, edizione digitale</title> <author>Alessandro Manzoni</author> <respStmt> <resp>Codificato da</resp> <persName xml:id="B.N.">Beatrice Nava</persName> </respStmt> </titleStmt> <sourceDesc> <p><bibl><author>Alessandro Manzoni,</author> <title>Il Conte di Carmagnola,</title> <editor xml:id="Ed_Ferrario">Vincenzo Ferrario,</editor> <pubPlace>Milano</pubPlace><date>1820</date></bibl> stabilito nell'edizione critica di <bibl> <editor xml:id="G.B.">Giovanni Bardazzi</editor> </pre>
--

² L'esempio di *markup* è limitato, per ragioni di brevità, alla descrizione di tre lezioni tratte dallo *specimen* sopra riportato, preceduto dalle sezioni del TeiHeader necessarie a comprendere gli attributi e i valori utilizzati.

```

<pubPlace>Milano</pubPlace>
<date>1985</date></bibl></p>

<listWit>
<witness xml:id="B_prose">Manz.B.VS.XI.4</witness>
<witness xml:id="B">Manz.B.X.2</witness>
<witness xml:id="Cc">Ms. Ital.72</witness>
</listWit>

</sourceDesc>
</fileDesc>

<encodingDesc>
<variantEncoding method="parallel-segmentation location="internal"/>
</encodingDesc>

<body>
<p xml:id="f.21" resp="#G.B.">3 I due eserciti si trovarono divisi da

<app>
<lem>un terreno paludoso</lem>
<rdg wit="#B_prose">
<subst>
<del seq="1">un lungo argine</del>
<add place="inline" seq="2">un terreno paludoso</add>
</subst>
</rdg>
<rdg wit="#Cc">un terreno paludoso</rdg>
</app>

<app>
<lem>in mezzo al quale passava</lem>
<rdg wit="#B_prose">
<subst>
<del seq="1">che aveva</del>
<add place="inline" seq="2">in mezzo al quale passava</add>
</subst>
</rdg>
<rdg wit="#Cc">in mezzo al quale passava</rdg>
</app>
[... ]
<app>
<lem>e fra le paludi</lem>
<rdg wit="#B_prose">
<subst>
<del seq="1">fra le<choice><sic>macch</sic><supplied resp="#B.N." cert="high">ie</supplied></choice></del>
<del seq="2">il Conte pose imboscate in molte macchie </del>
<del seq="3">le paludi avevano sparse qua e là d</del>
<add seq="4">e nelle paludi v'erano</add>
<del seq="5">molte</del>
<add seq="6">quà e là macchie</add>
<del seq="7">di</del>
<addseq="8">e alture</add>
<del seq="9">di</del>
<del seq="10">da cui</del>
<addseq="11">e fra le paludi</add>
</subst>
</rdg>
</app>
</p>

```

Tabella 2: esempio marcatura apparato

Partendo da una marcatura base così analitica, si potrebbero poi sviluppare strumenti che permettano una animazione della sequenza scrittoria. Una visualizzazione di questo tipo⁴, basata sull'elaborazione di una codifica che espliciti il susseguirsi

³ Ciascun comma delle prose introduttive dell'opera è marcato attraverso l'indicazione del numero di pagina e della lettera stabilita da Bardazzi nella propria edizione critica.

⁴ Cfr. La Dama Boba (<http://damaboba.unibo.it>) e Samuel Beckett Digital Manuscript Project (<http://www.beckettarchive.org>).

degli interventi autoriali, facilita la ricostruzione genetica del processo di scrittura e la comprensione dello sviluppo diacronico del testo. Oltre a questa interessante possibilità, l'edizione digitale si configura come un fondamentale strumento di supporto alla ricerca filologica e critica. Infatti lo stato attuale degli studi sul *Carmagnola* lascia aperti consistenti problemi ecdotici e di sistemazione del testo, come, ad esempio, la presenza di lezioni diverse tra la copia censura e la *princeps*, non documentate da alcuna evidenza testimoniale⁵. Si rende dunque necessaria un'annotazione ragionata che consenta poi, ad appositi *tool* di ricerca e interrogazione, di estrarre tutti i luoghi testuali interessati da una specifica criticità, così da facilitare notevolmente lo studio di questioni cruciali per una comprensione definitiva dell'opera. In particolare, attraverso la creazione di una tassonomia che includa le questioni critiche ancora aperte e l'assegnazione di @xml:id univoci a ciascuna casistica specifica, si potrebbero poi progettare gli strumenti di *query* finalizzati all'*information retrieval* in modo da permettere il recupero immediato di tutti i *loci* interessati. Oltre a una maggior immediatezza rispetto a quanto si potrebbe ottenere con la creazione di tavole sinottiche in cartaceo, i fruitori dell'edizione digitale avrebbero anche la possibilità – attraverso gli opportuni operatori – di combinare e raffinare i propri elementi di ricerca.

```
<encodingDesc>
<classDecl>
<taxonomy>
<category xml:id="gap_censura_princeps">
<category xml:id="intervento_con_doppia_attestazione">
<categoryxml:id="intervento_parigino">
</classDecl>
</encodingDesc>
```

Tabella 3: esempio di tassonomia per la codifica di questioni critiche rilevanti

4. CONCLUSIONI

Avendo evidenziato concretamente i vantaggi dell'edizione digitale di un'opera a tradizione complessa, verranno infine discusse le criticità delle strategie adottabili per la marcatura dell'apparato, che emergono immediatamente anche dalla trattazione del semplice esempio riportato, se si pensa alla necessità di elaborare una codifica così analitica per l'intera opera, che presenta anche casi molto più complessi. La formalizzazione dell'apparato del *Carmagnola* permette anche una riflessione di grande interesse circa l'applicabilità dello standard XML TEI a un caso in cui le strutture varianti tendono a superare, anche di molto, le porzioni strutturali gestibili all'interno di questo linguaggio, generando i ben noti problemi di *overlapping* (Schmidt 2010 e 2012). Sono ancora in fase di studio le possibili marcature da adottare per affrontare la questione rimanendo all'interno dello standard in uso, ma si evidenzia un duplice criticità: il rischio di una sovrabbondanza di tag (dovuta, per esempio, all'uso di <anchor> e alla necessità di disseminare il testo di @xml:id), con possibili problemi di gestione in fase di visualizzazione e recupero dell'informazione, e la gestione delle tempistiche di realizzazione di una marcatura così complessa.

L'intervento proposto permette dunque di discutere una tipologia testuale oggi non rappresentata nel panorama delle DH, riflettendo contestualmente sulle strategie di marcatura più adatte per descriverla.

5. BIBLIOGRAFIA

- [1] Bardazzi, G. 1985, *Il Conte di Carmagnola*, edizione critica a cura di G. Bardazzi, Milano, Fondazione Arnoldo e Alberto Mondadori
- [2] Italia P., Bonsi C. 2016, Edizioni Critiche Digitali. Digital Critical Editions. Edizioni a confronto/comparingeditions, Roma, Sapienza Università Editrice. ISBN: 978-88-9377-003-3
- [3] Manzoni, A. 1820, *Il Conte di Carmagnola*, Milano, Tipografia Ferrario
- [4] Pierazzo, E. 2015, *Digital Scholarly Editing: Theories, Models and Methods*, Burlington - Ashgate, Farnham – Surrey. ISBN: 978-1-4724-1211-9.
- [5] Schmidt, D. 2010, *The inadequacy of embedded markup for cultural heritage texts*. «Literary and Linguistic Computing», 25, 3, pp. 337–356. DOI=https://doi.org/10.1093/lc/fqq007
- [6] Schmidt, D. 2012, *The role of markup in the digital humanities*. «Historical Social Research» 37, 3, pp. 125-146.
- [7] Severino, R. 1990, *A carte scoperte. Manzoni e altri contributi critici e filologici sulla cultura italiana in America*, Roma, Edizioni dell'Ateneo, pp. 3-19
- [8] Tomasi, F. 2013, *Le edizioni digitali come nuovo modello per dati di autorità concettuali*, JLIS.it., II. DOI= http://dx.doi.org/10.4403/jlis.it-8808
- [9] Tomasi F. 2014, *L'ecdotica e le tecnologie. Tra applicazioni, principi e un esperimento*, «AnuarioLope de Vega. Texto, literatura, cultura» XX, pp. 84-98. DOI=https://doi.org/10.5565/rev/anuariolopedevega.84
- [10] Manoscritti manzoniani consultati: Biblioteca Nazionale Braidense, Milano Manz.B.VS.XI.4; Manz.B.X.2. Houghton Library, Harvard, MS Ital. 72 (riproduzione digitale).
- [11] Catalogo mss. manzoniani: Bassi, D. 1931, *Lettere inedite di Alessandro Manzoni*, «Aevum», V, pp. 35-60.

⁵ Le principali questioni critiche emerse dal nuovo studio sul *Carmagnola*, bisognose di specifica codifica, saranno brevemente riassunte in sede di intervento; si fornisce qui solo un esempio di possibile tassonomia.

From space to space: Image-based digital 3d-reconstruction of Indian Nagara architecture

Gerald Kozicz¹, Ludwig Grimm²

¹ Institute of Architecture and Media, TU Graz - Austria - kozicz@tu-graz.at

² Institute of Architecture and Media, TU Graz - Austria - ludwig.grimm@tu-graz.at

1. INTRODUCTION

The proposed paper is based on an ongoing research project (FWF P28509) which deals with the documentation, representation and analysis of a particular type of North Indian sacral architecture, known as Nagara-Temples. Because of many different reasons, including bare physical, but also historical and political ones, the general given stage of documentation of this type of architecture is comparatively poor. Furthermore, the area of the western Himalayas, where the Nagara-Temples are situated, is highly endangered by natural disasters like floods, avalanches and landslides. Accordingly, the need for precise documentation of the current building status is evident from a perspective of cultural heritage preservation.

As an interdisciplinary project integrating cultural studies and digital media research, one of the goals is to develop new ways of digital documentation and representation in order to allow for more precision, confirmability and a better communication in the culture-theoretical analysis and to digitally preserve the current condition of the architectural subjects precisely. At the same time, the methods developed in the technical part of the project must respect the conditions and limitations of the given source situation and fieldwork practice to be applicable in cultural studies.

The organic nature of the Nagara architecture is especially interesting, but also challenging in the context of digital documentation. Its formal manifestation makes it difficult to distinguish structural from ornamental features as a high level of nested integration between super- and sub-forms is observable within a fractal grammar of shape. Very Precise, three-dimensional models with high resolutions were needed to represent the architectural qualities appropriately. At the same time, the complexity and cost of the used technology must be kept in an appropriate relationship to the project size and funding. In order to reach this goal, various digital tools had to be integrated in a custom methodology.

2. TECHNICAL METHODOLOGY

2.1 PHOTOGRAMMETRY

The digital photogrammetric processing automatically generates models from series of photographs or video footage, which olds each feature of the object from at least two perspective positions. Based on image-feature-matching, bundle-adjustment calculations and forward intersections, a point-cloud model is generated of which a mesh model can be derived. In comparison to the laser-scan it is less precise, but has many advantages in terms of the practical usability especially for smaller projects. Additionally, the photogrammetry allows natively for automated texture generation, which is impossible with a pure laser scan. Most importantly, one does not need a special hardware-setup other than a DSLR-camera (possibly with a tripod) to take the source-photos and a relatively powerful workstation-pc in order to process them. However, a camera-drone is not mandatory, but very helpful to cover viewing angles from above, or larger objects/areas in general. As in general the features of photogrammetry are very well suited for the needs of the ongoing project, the decision to use it as primary technology was made.

While being modest on the needed hardware in-situ, a special software for the processing is needed. Today, there are various competitors on the market, reaching from simple, fully automated consumer-software without editing abilities to professional products with further analysis features and scripting-interfaces. Which of the available solutions might be the best choice is of course dependent on the special needs of the aimed usage and the available budget. In the course of the ongoing project, different products have been tested. Without further elaborating on the details of the differences between those, Agisoft Photoscan proved to be most suitable for the purposes of the project.¹

2.2 SOURCE FOOTAGE AND MANUAL EDITING

The necessary series of images, serving as source for the photogrammetric processing, has to fulfil certain parameters in order to achieve acceptable results. The lighting situation is crucial to the process, but is often unalterable in uncontrolled environments. Furthermore in many cases, the actually necessary positions for photographs cannot be reached, resulting in incomplete or distorted parts of the generated model. By using a camera-drone for capturing the footage, those “blind spots” might be significantly reduced, dependent on the actual site situation. Still, manual remodelling of certain parts, editing, cleaning and manual texture processing is practically unavoidable to achieve consistent results.

3. POSSIBLE USAGE OF THE GENERATED MODEL

The direct outcome of the photogrammetric processing is a high-density polygon model, including diffuse color textures. Based on the 3d-model, the original photographs can be orthorectified (liberated from perspective distortion) and stitched to generate high resolution raster imagery of orthogonally projected facade views.

¹ cf. Loguidice, *The pitfalls*

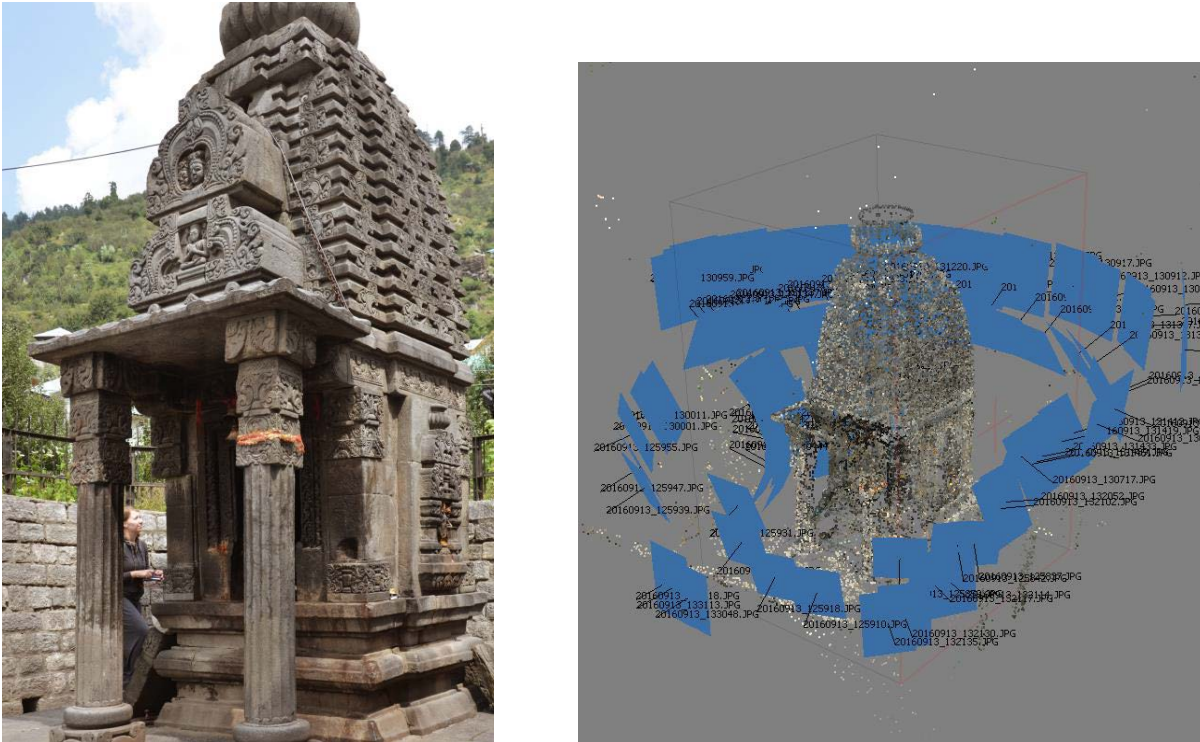


Fig. 1 and 2: Photograph of the Jagatsuk Gauri Shankar shrine and photogrammetric processing

The high-poly model can be transferred to other 3d-applications in order to manually post process and edit it. This is necessary to clear common issues of the automated process. Dependent on the density/resolution of the model, this is effectively possible only by using digital sculpting software, like Pixologic Zbrush or Autodesk Mudbox, which have been designed for editing very dense meshes. By taking the model to a digital content creation package like Autodesk 3ds Max or Maya, raytraced renderings from any perspective view can be generated.²

For cultural-historical analysis, the naturalistic high-polygon model can serve as precise reference of geometry and also texture, being always available once generated. It serves also as documentation of the current condition in terms of digital cultural heritage preservation. Taking it as reference, any form of abstracted model, for example representations of estimated construction stages, structural and shape-grammatical elements or iconographic relationships can be modelled manually.

With more advanced post-processing techniques, a low-resolution model, which still holds a lot of naturalistic detail, can be derived from the original high-poly. This model can be rendered in real-time consumer applications, making it available for public display, whether stationary or on mobile devices. Even Virtual Reality applications, making the model virtually accessible in the original scale, are possible.

4. REFERENCES

- [1] Litomisky, Krystof. *Consumer RGB-D Cameras and their Applications*. PhD diss., University of California, 2012.
- [2] Loguidice, Bill. *The pitfalls and potential of inexpensive 3D scanning solutions*. Last modified: June 17, 2017. <https://arstechnica.com/gadgets/2017/06/>
- [3] Maxzin, Joerg. *Auferstandenes Kulturerbe: Wie eine verbrannte Holzskulptur dank 3D gerettet wurde*. In: Digital Production, 04/2017. München: DETAIL Business Information GmbH.
- [4] Technische Hochschule Deggendorf (Hrsg.). *Lukas aus der Asche – Auferstandenes Kulturerbe aus dem 3d-Labor*. Lindenberg im Allgäu: Kunstverlag Josef Fink, 2016.

² cf. Maxzin, *Auferstandenes Kulturerbe*, 77f



Fig. 3 and 4: Shaded high-density polygon model and raytraced rendering

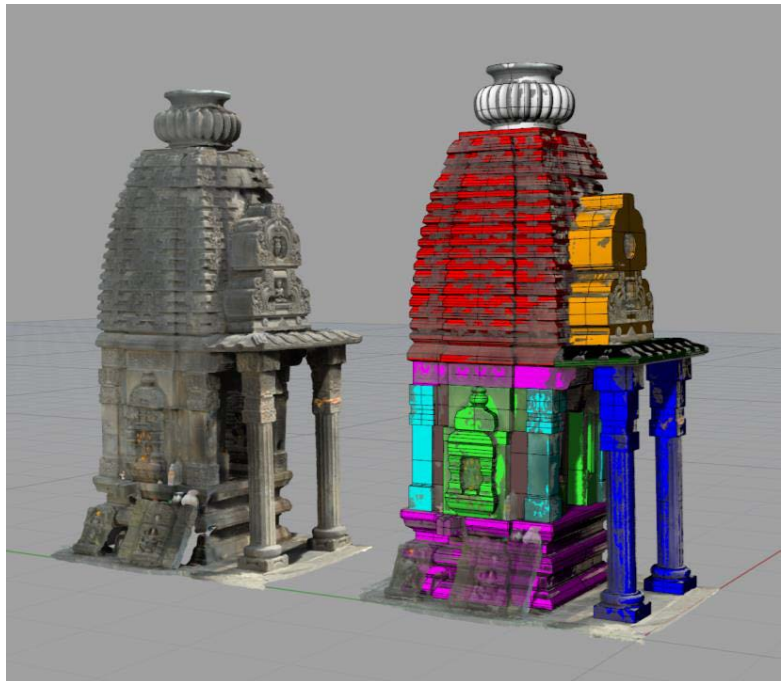


Fig. 5 and 6: Orthorectified elevation view and abstracted analytical model

From the Clavius Correspondence to Linked Data: the CoW-LD Project

Angelica Lo Duca¹, Andrea Marchetti²

¹ Institute of informatics and Telematics - CNR, Italy - angelica.loduca@iit.cnr.it

² Institute of informatics and Telematics - CNR, Italy - andrea.marchetti@iit.cnr.it

ABSTRACT

Over the last years, a great effort has been made in the field of Digital Humanities to digitize documents and collections in different formats. As a result a great amount of data has been made available thus providing new opportunities to build applications allowing the navigation of collections. In this paper, we focus on the Clavius Correspondence, a set of 266 handwritten letters, sent to Christopher Clavius, a known mathematician of the Sixteenth Century, from other important scientists of the same period. In order to bring all these precious manuscripts on the Web, the Clavius on the Web (CoW) project was built. CoW enriched all these letters with lexical and semantic annotations. In addition, the project implemented a Web platform where all this material was made available and annotations within letters could be done. Among the many activities of CoW, one focused on building a Linked Dataset about all contextual resources associated to manuscripts, people and places involved in those letters. This paper describes the CoW Linked Dataset (CoW-LD), built around these manuscripts. CoW-LD was built partially manually and partially through automatic programs, which enriched the dataset with external resources, such as those contained in DBpedia and GeoNames. As a result, CoW-LD contains also letters annotations and events associated to the authors of the letters, as well as aliases to external sources.

KEYWORDS

Linked Data, Digital Humanities, Europeana Data Model, Christopher Clavius.

1. INTRODUCTION

One of the preliminary activities of Digital Humanities consists in digitizing documents and manuscripts and organizing them in collections. After that, the next activity consists in making them accessible to common users through Web portals or online catalogs [1]. Many projects and initiatives have been proposed following this line, such as the Levy Dead Sea Scrolls Digital Library¹, which shows the famous Qumran scrolls in a Web portal. Another recent initiative is the Clavius on the Web project (CoW) [2], which aims at making accessible on the Web the Clavius Correspondence, owned by the Historical Archives of the Pontifical Gregorian University (APUG)² in Rome. Christopher Clavius (1537-1612) was a jesuit mathematician and astronomer and one of the most important characters in the scientific scene of the late Sixteenth Century. These manuscripts consist of two volumes of correspondence (266 letters) and seven volumes of works, some of these ones printed and some not still published. The importance of the correspondence becomes clear just looking at the authors of the letters: Galileo Galilei, Tycho Brahe, Joseph Scaliger, Guido Ubaldo Dal Monte and many others. Within CoW many modules were implemented to visualize, annotate and extract knowledge from manuscripts.

In this paper, we describe the Clavius on the Web Linked Dataset (CoW-LD)³ which represents all the contextual resources associated to the Clavius Correspondence as Linked Data. CoW-LD contains many links to external sources, i.e. DBpedia [3], GeoNames⁴, Wikipedia⁵, VIAF⁶, Treccani⁷ and Google Books⁸. The dataset is released under the CC BY-SA 4.0 license⁹ and is published as a Linked Data node, which can be queried through a SPARQL endpoint¹⁰.

2. RELATED WORK

Many important initiatives have been rising to publish collections of documents and manuscripts as Linked Data. DM2E [4] is a project which provides links to manuscripts and their contextual resources from fifteen Cultural Heritage institutions across Europe. Other important initiatives are the Europeana Linked Open Data [5], the Burckhardtsource project [6], BookSampo dataset [7] for fiction literature, the Amsterdam Museum [8] and the Rijksmuseum Collection [9].

¹ <<http://www.deadseascrolls.org.il/>> (last access 2017/09/29).

² <https://www.unigre.it/Archivio/archivio_en.php> (last access 2017/09/29).

³ <<http://claviusontheweb.it/lod>> (last access 2017/09/29).

⁴ <<http://www.geonames.org/>> (last access 2017/09/29).

⁵ <<https://www.wikipedia.org/>> (last access 2017/09/29).

⁶ <<https://viaf.org/>> (last access 2017/09/29).

⁷ <<http://www.treccani.it/>> (last access 2017/09/29).

⁸ <<https://books.google.com/>> (last access 2017/09/29).

⁹ <<https://creativecommons.org/licenses/by-sa/4.0/>> (last access 2017/09/29).

¹⁰ <<http://claviusontheweb.it/dataset/snorql/>> (last access 2017/09/29).

6. BIBLIOGRAPHY

- [1] BORGMAN, Christine L. *What are digital libraries? Competing visions*. Information Processing & Management, 1999, 35.3: pp. 227-243.
- [2] PEDRETTI, Irene et. al. *The Clavius on the Web Project: Digitization, Annotation and Visualization of Early Modern Manuscripts*. In Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem (AIUCD '14), 2014, ACM, New York, NY, USA.
- [3] AUER, Soren. et al. *Dbpedia: A nucleus for a web of open data*. The Semantic Web, 2007, 722-735.
- [4] BAIERER, Konstantin, et al. DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities. Semantic Web, 2017, Preprint: 1-13.
- [5] ISAAC, Antoine; HASLHOFER, Bernhard. *Europeana linked open data–data. europeana. eu*. Semantic Web, 2013, 4.3: 291-297.
- [6] DI DONATO, Francesca, et. al. Biblioteche digitali semantiche. il progetto burckhardtsource. org. Bibliotime, 2013, 16.1.
- [7] MÄKELÄ, Eetu, et. al.. Fiction literature as linked open data - The BookSampo dataset. Semantic Web, 2013, 4.3: 299-306.
- [8] DE BOER, Victor, et al. *Amsterdam museum linked open data*. Semantic Web, 2013, 4.3: 237-243.
- [9] DIJKSHOORN, Chris, et al. *The Rijksmuseum Collection as Linked Data*. Semantic Web, 2014, Preprint: 1-10.
- [10] ISAAC, Antoine. et al. *Europeana data model primer*. 2013. Available at: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primr_130714.pdf (last access 2017/09/29).
- [11] BRICKLEY, Dan et. al. *FOAF vocabulary specification 0.99*. 2007. Available at: <http://xmlns.com/foaf/spec/> (last access 2017/09/29).
- [12] MILES, Alistair, et. al.. *SKOS simple knowledge organization system reference*. 2009. Available at: <https://www.w3.org/TR/skos-reference/> (last access 2017/09/29).
- [13] MCGUINNESS, Deborah L., et al. *OWL web ontology language overview*. W3C recommendation, 2004, 10.10: 2004. <https://www.w3.org/TR/owl-features/> (last access 2017/09/29).
- [14] CLAVIUS, Christoph. *CORRISPONDENZA*. Edizione critica a cura di. Ugo Baldini e Pier Daniele Napolitani. VOLUME II. 1570 - 1592. Parte I: lettere e testi.

Gli *Acta Eruditorum* come esempio di valorizzazione di una collezione digitale

Stefano Casati¹, Federica Viazzi²

¹Museo Galileo, s.casati@museogalileo.it, ²Fondazione BEIC, federica.viazzi@beic.it

KEYWORDS

Convergenza/divergenza tra archivio, biblioteca, museo - Valorizzazione

Le biblioteche digitali del Museo Galileo di Firenze e della Fondazione Biblioteca Europea di Informazione e Cultura di Milano hanno iniziato a collaborare nel 2011 sulla spinta delle attività sostenute dalla Biblioteca Digitale Italiana. Da allora hanno realizzato numerosi progetti, convinti che per creare un circolo virtuoso di contaminazioni e arricchimento reciproco la cooperazione con realtà diverse dalla propria rivesta un ruolo di fondamentale importanza.

La condivisione della ricerca di un'offerta informativa che superi la semplice rappresentazione in formato immagine dei documenti costituisce indubbiamente il cardine che sostiene l'intesa tra le due istituzioni. La collaborazione si è infatti sviluppata e rafforzata nel tempo proprio nella ideazione di metodologie e sistemi innovativi per arricchire la presentazione digitale del supporto analogico.

Le odierne strategie di ricerca delle informazioni impongono di offrire una diversificazione delle esperienze informative e di fornire molteplici vie d'accesso alle risorse, con l'obiettivo di aggiungere valore alle usuali ricerche e di conseguenza ai risultati ottenuti. La logica dell'*information retrieval* dovrà sempre più integrarsi anche con l'appagamento della curiosità e degli interessi degli utenti remoti, che, se soddisfatti, saranno ulteriormente incentivati a tornare sul portale non solo quando avranno necessità di effettuare ricerche bibliografiche.

Con la proposta di un intervento rivolto al tema della valorizzazione delle collezioni digitali intendiamo fornire una breve *summa* della filiera che ha visto protagonista un'opera tra le più rilevanti dell'età moderna: gli *Acta Eruditorum*, dalla campagna di digitalizzazione dei volumi, sino alla decisione di valorizzarli "liberandone" le immagini e condividendole nei progetti Wiki.

Acta Eruditorum è una delle prime riviste scientifiche di area germanica¹ e il più antico giornale letterario tedesco. Noto ben oltre la ristretta cerchia della comunità degli storici della scienza, per un secolo intero - dal 1682 al 1782 - aggiornò gli intellettuali e gli studiosi europei sulle ultime scoperte scientifiche e sulle novità letterarie dell'epoca. Fondato da Otto Mencke e Gottfried Leibniz, era interamente scritto in latino e conteneva estratti da nuovi scritti, note e recensioni, brevi saggi e trattati, soprattutto sulle scienze naturali e la matematica, ma anche su varie altre discipline quali diritto, storia e teologia. Vi contribuirono numerosi filosofi naturali e letterati², come Leibniz, Robert Boyle, Leonhard Euler, Pierre-Simon Laplace, tanto per citarne alcuni tra i più famosi.

Negli *Acta* apparvero studi di eccezionale importanza, come quello di Leibniz sui fondamenti del calcolo infinitesimale³ e la rivista divenne uno dei principali veicoli della comunicazione del sapere durante l'Illuminismo⁴. La particolarità delle linee editoriali degli *Acta* è che la maggior parte delle recensioni e degli articoli sono anonimi rendendo questa monumentale raccolta di fatto un'opera collettiva imputabile ai *Collectores Actorum Eruditorum*.

Il Museo Galileo e la Fondazione Beic, consapevoli dell'importanza della rivista, da circa tre anni hanno intrapreso una strada comune volta alla valorizzazione dell'opera sia nel suo insieme sia dei singoli contributi che contiene.

L'intervento mirerà ad illustrare le procedure seguite per mettere a disposizione del pubblico il *corpus* degli *Acta Eruditorum*, esplicitando quali siano state le scelte tecniche, scientifiche e procedurali che hanno definito le linee guida a monte del progetto di digitalizzazione e le scelte adottate riguardo agli standard di metadato, con le conseguenti differenze di rappresentazione dei documenti. L'accento verrà poi posto sul lavoro svolto per creare gli oggetti digitali. I metadati bibliografici sono stati compilati con l'obiettivo di offrire una navigabilità il più possibile fluida (con dei link dall'opac verso i contributi e le opere dai quali sono stati estratti gli articoli e vice-versa per creare una struttura fondamentalmente circolare della conoscenza). Complementari ai metadati bibliografici, dei metadati strutturali ben architettati possono permettere di rappresentare graficamente l'organizzazione logica e fisica del documento, mentre il lavoro di registrazione degli articoli e delle recensioni vuole garantire l'accesso diretto ai singoli contributi.

Mentre il lavoro di digitalizzazione, metadato e pubblicazione della collezione può dirsi concluso, l'attività di promozione e divulgazione è invece in corso d'opera. Per iniziare si è scelto di valorizzare prima di tutto le immagini relative al significativo apparato iconografico dei volumi degli *Acta Eruditorum*, arricchiti da numerose tavole xilografiche e calcografiche che rappresentano esemplarmente il rapporto che si instaurò fra arte e scienza nell'Europa del periodo.

¹ "Acta Eruditorum", The Euler Archive, ultimo accesso 12 settembre 2017, <https://eulerarchive.maa.org/publications/journals/ActaEruditorum.html>.

² "Laeven Acta Eruditorum", heiDOK – The Heidelberg Document Repository, ultimo accesso 12 settembre 2017, https://archiv.ub.uni-heidelberg.de/volltextserver/16568/1/Laeven_Acta_Eruditorum.pdf.

³ "Acta Eruditorum", Treccani, ultimo accesso 12 settembre 2017, <https://www.treccani.it/enciclopedia/acta-eruditorum>

⁴ Laeven, Augustinus Hubertus. 1990. The "Acta Eruditorum" under the Editorship of Otto Mencke. The History of an International Learned Journal between 1682 and 1707. Amsterdam: APA-Holland University Press

Il senso profondo dell'intesa quindi, è da individuare nella comune adesione alla "filosofia GLAM" (Galleries, Libraries, Archives and Museums⁵) e all'apertura verso gli interessanti e proficui orizzonti dell'universo Wiki. In linea questa filosofia e con le nuove tendenze in materia di licenze e condivisione di immagini e metadati, i *files* relativi all'apparato iconografico degli *Acta Eruditorum* sono stati caricati sulla piattaforma Wikimedia Commons liberi da diritti così da poter essere diffusi e riusati.

Attraverso un lavoro di ricerca storico-bibliografica ancora in fase di sviluppo, le immagini vengono inserite in Wikipedia nelle varie lingue all'interno delle voci relative all'argomento trattato, all'autore dell'articolo e comunque dove possano risultare pertinenti (per esempio all'interno della voce relativa all'Accademia alla quale era affiliato un determinato scienziato). Per fare un esempio concreto dell'impatto positivo dato dall'inserimento di un'immagine all'interno di una voce di Wikipedia, le illustrazioni tratte dagli *Acta* vengono visualizzate oltre 500mila volte ogni mese nelle voci di Wikipedia⁶.

BIBLIOGRAFIA

- [1] Stefano Casati, Fabrizio Butini, *Some remarks about Museo Galileo's digital collections*, in *Collaborative research practices and shared infrastructures for humanities computing: 2nd AIUCD Annual Conference, AIUCD 2013, Padua, Italy, 11-12 December 2013*, proceedings of revised papers / Maristella Agosti and Francesca Tomasi (eds), Padova: CLEUP, 2014. p. 143-148.
- [2] Stefano Casati, *La Biblioteca digitale del Museo Galileo*, In *Biblioteche oggi* (genn.-febr. 2015) p. 45-51.
- [3] Chiara Cauzzi, Valeria De Francesca, Lisa Longhi, Federica Viazzi, *Conoscersi per riconoscersi: la partecipazione come specchio del bibliotecario*, in *AIBStudi* (2016), DOI: <http://dx.doi.org/10.2426/aibstudi-11462>
- [4] Federica Viazzi, *Chi siamo? La descrizione fisica e logica del documento*, in *Biblioteche Oggi Trends* (dicembre 2016), p. 33-42, DOI: 10.3302/2421-3810-201602-033-1
- [5] Stefano Casati, Camilla Rotoli, *La Biblioteca digitale del Museo Galileo e il progetto GLAM*, in *Biblioteche oggi* (luglio-agosto 2017) p. 33-36.

⁵ "Progetto: GLAM/Progetti", Progetto: GLAM, Wikipedia, ultimo accesso 12 settembre 2017, <https://it.wikipedia.org/wiki/Progetto:GLAM/Progetti>.

⁶"GLAMorgan", BaGLAMa 2, ultimo accesso 12 settembre 2017, <https://tools.wmflabs.org/glamtools/glamorgan.html?&category=Acta%20Eruditorum&depth=6&year=2017&month=8>.

Historical Images and the Recovery of the Past. *The Medieval Kingdom of Sicily Image Database*

Caroline Bruzelius¹, Paola Vitolo², Joseph Williams³

¹Duke University, NC, c.bruzelius@duke.edu

²University "Federico II" of Naples, Italy - paolavitolo@libero.it

³PhD researcher - . joseph.c.williams@duke.edu

1. INTRODUCTION AND PROJECT GOALS

South Italy was a site of artistic and architectural experimentation. This was particularly true in the Middle Ages, when this geographically disparate area was forged by its Norman conquerors into a kingdom (1130), which was subsequently ruled by Hohenstaufen, Angevin, and Aragonese dynasties. The multicultural heritage forged in the Kingdom of Sicily represents therefore a rich architectural and decorative patrimony for both Mediterranean and European cultures, and was in many ways unique in the history of art and architecture.

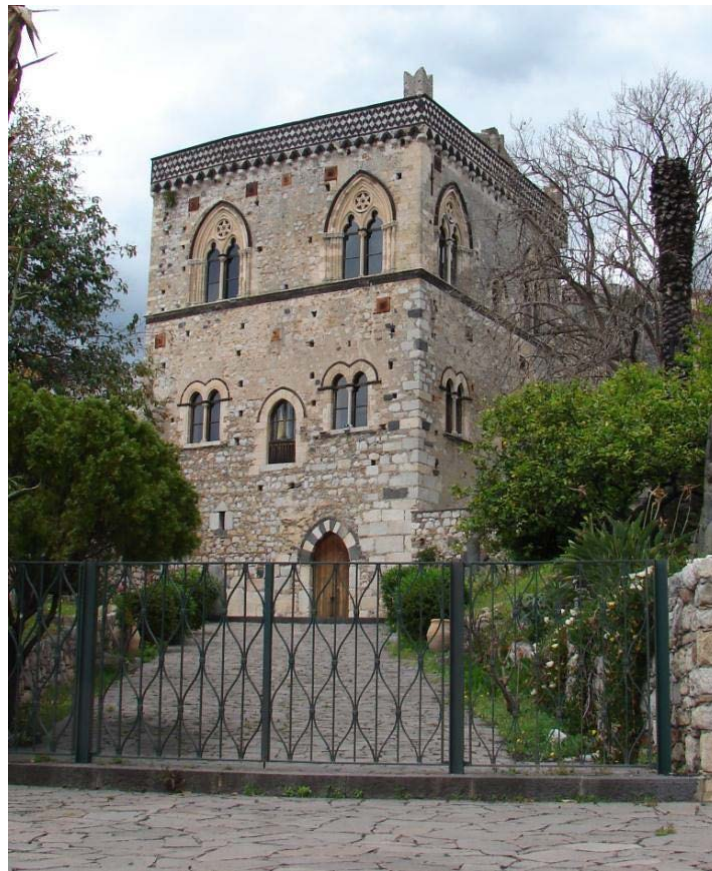


Fig.1. Taormina, Duchi di Santo Stefano Palace (14th century).

This artistic legacy comes down to us in profoundly modified form, however. The devastation of earthquakes and volcanic eruptions, Baroque re-decoration, radical modern restoration, urban transformation and war destruction have led to huge losses.

In addition, urban expansion has changed the visual impact of the monuments on the urban landscape. As a result of so much devastation, we now see these important monuments through the filter of innumerable changes and modifications, many of which are re-interpretations of the original structures that were driven by post-war ideological or economic reasons. Nothing can compensate for the destruction of monuments. But historical images (photographs, drawings, maps, paintings...) produced as a visual record of travel, as practical exercises in professional training, or for documentary purposes by artists, architects, travellers, photographers and scholars of the past can to some extent mitigate the losses, permitting us to “see” monuments and urban or natural landscape prior to the profound changes of the last centuries.

The Medieval Kingdom of Sicily Image Database, online since October 2016, was therefore created to provide the users with as much information as possible on the original appearance of the important sites of South Italy, as well as to document the process of their rediscovery by scholars and travellers from the 18th up to the mid-20th centuries. Our database makes accessible to the community of a wide range of disciplines, as well as to the local residents and travellers, a large body of

information and knowledge that is dispersed in Italian, European and American collections, images that are for the most part unknown and unpublished.



Fig. 2. Messina Cathedral destroyed in the earthquake of 1908

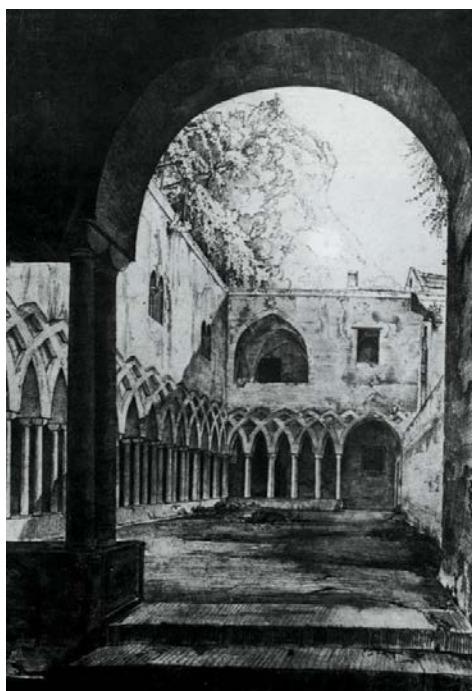


Fig. 3. Johann Heinrich Schilbach, *Amalfi, Capuchin convent (San Pietro a Toczolo)*, 1825. Darmstadt, Hessisches Landesmuseum

2. PROJECT STORY AND STRUCTURE

This project was initiated in 2011 by Caroline Bruzelius with funding from a National Endowment for the Humanities (USA), and now forms part of the activities of the Duke University (NC, USA) Wired! Lab. It is presently directed by Bruzelius and Paola Vitolo (University “Federico II” of Naples, Italy), with Joseph Williams (PhD researcher) as project manager, David Tremmel (Duke University) as data manager and database & web developer, William Broom (Duke University) as project coordinator, John Taormina (Duke University) as metadata and image management consultant. It has also involved numerous collaborators over the years at different stages of their academic career.

The Medieval Kingdom of Sicily Image Database
A Visual Resource of Historical Sites c. 1100 - c. 1450

Home Browse & Search Map of the Kingdom Copyright and Reuse Contact Us English Italiano

About

This database is a collection of historic images that represents the medieval monuments and cities of the Kingdom of Sicily collected from museums, libraries, archives and publications.

The historic Kingdom of Sicily encompassed the Island of Sicily and the areas of lower Lazio, Campania, Abruzzo, Molise, Apulia, Basilicata, and Calabria. The kingdom was created in the twelfth century by the Normans, was inherited by Frederick II of the Hohenstaufen dynasty, and conquered by the French in 1266. In the fifteenth century it came under Aragonese control. The rich and international artistic legacy of towns, castles, churches, monasteries and their decoration is a testament to the dynamic social and political history of South Italy. With its unique geographic position and multicultural heritage, the Kingdom of Sicily was a melting pot of artistic and architectural concepts.

Why are historic views particularly important for these monuments, and why have we created this database? Our repository gathers images created for the most part prior to the destruction or significant alteration of sites and monuments, as wars, earthquakes, extensive restoration, as well as simple neglect, have profoundly affected this rich artistic patrimony. Urban expansion has also often changed the landscape and transformed the role and meaning of monuments within cities or their periphery.

The database features historical images in a range of media, including drawing, painting, engraving and photography. The dates of the images range from the late-sixteenth through the mid-twentieth centuries. These images and available descriptive information about their production can be found under the "Browse & Search" tab. Sites are organized topographically by location. Each record contains a list of associated Images: a brief description of the site, and details about its creation and relevant sources when known. More recent published images, including architectural plans, photographs and reconstructions, are collected within individual site entries under the "Visual Docs" tab.

This database is a work in progress and is by no means exhaustive or complete. There exist many more images produced by travelers, scholars, artists, architects, photographers, and soldiers, who traveled or studied in southern Italy that we have not been able to catalogue. The website editors will be grateful to receive information on collections and publications that can contribute to the project. To offer suggestions, use the "Contact Us" link above.

This web site is hosted by Trinity Technology Services (TTS) at Duke University. We are indebted to TTS for all of the support they have provided for this project.

Fig.4. The Medieval Kingdom of Sicily Image Database Homepage

At the outset of the project, the team created a master list of roughly 800 geo-referenced sites, developed a database structure, and established the criteria for the selection and cataloguing of images. The team was sent into libraries, museums and archives in Europe and USA and, as many images are now available online, also established links to the digital collections of Museums and cultural institutions.

The database consists of two linked parts: Historical Images and Works/Sites. For each individual image, data records include five sections with specific types of information: Overview (description and basic information on chronology, title, image type, notes and analysis), Image (reproduction of the image with information on format and resolution), Creation (dates, technique, measurements, artist's name and dates), Location (Repository and link to online collections), Research (bibliography and links to Google search and Books), Cross reference to Works entries.

In the Works and Sites entries we provide information on chronology, creation, patronage, artistic and architectural feature, condition, restorations and refurbishment campaigns.

The database is a work in progress and was designed as an expandable resource that can be continuously added to, enlarged and improved as new resources, collections, as well as new types of scholarship, emerge.

The screenshot shows the website interface for the Medieval Kingdom of Sicily Image Database. The main heading is "The Medieval Kingdom of Sicily Image Database" with the subtitle "A Visual Resource of Historical Sites c. 1100 - c. 1450". A search bar is located in the top right corner. The navigation menu includes "Home", "Browse & Search", "Map of the Kingdom", "Copyright and Reuse", and "Contact Us". The sidebar on the left contains "Cities & Sites", "Images by Artist/Creator", "Images by Collection", and "Advanced Image Search". The main content area is titled "Image Details" and shows the entry for "Agrigento, Cathedral, Porta di Ponte". Below the title are tabs for "Overview", "Image", "Creation", "Location", and "Research". A link "Send us additional information about this image" is visible. A table of metadata is displayed, and a small image thumbnail is shown on the left.

Field Name	Contents
Image number	35074
Description	View of the western side of the town
Given Title	S. Pietro, Porta del Ponte
Site(s) Depicted	Agrigento, Cathedral (San Gerlando)
Type	Drawing
View Type	Full View
Period	19th Century
Analysis	Western side of the town with apsidal view of the cathedral and the Porta del Ponte (destroyed).
Notes	From Leo von Klenze's series made during his 1823 trip to Rome and Sicily with King Ludwig I of Bavaria.
Cataloger	Gabriella Cianciolo

Fig.5. *The Medieval Kingdom of Sicily Image Database*. Image entry for “Agrigento, Cathedral, Porta di Ponte” (nr. 35074)

3. METHODOLOGICAL PROBLEMS

Limits and problems related to the use of historical images have been clear from the very beginning to the creators of the database. In particular:

1. Each image must necessarily be approached critically, and cross-referenced with other kind of written and/or figurative sources. Individual views may not be reliable representations of the exact condition of places and buildings at a specific date depending on personal or artistic purposes.
2. As the creators of the website are not in a position to make judgments on the significance of each image for individual research, we have decided to be as inclusive as possible, gathering and cataloguing as many images of buildings and their decorative programs as seem relevant to us. It is our expectation that those whose work focuses on individual sites or topics will verify that an image is historically correct, as the cataloguers are not experts on all subjects.
3. This research initiative potentially has an endless duration and therefore has no claim of being exhaustive.
4. Since the acquisition of images proceeds through collections, the collection of material is not necessarily representative of the total work of an artist or a traveler, or of the places and monuments they depicted.

4. FUTURE PLANS

The editors of the database, in looking towards the future, have in mind first of all to expand the image collection, as well as to enlarge the geographical representation and the typology of materials. At present the team is developing an interactive mapping function that will show the location of sites in relation to the roads and ports of the region. The new mapping component will be searchable by types of institutions as well as artistic styles. It will interact with the material on each specific site in the database, visualizing the itineraries of Grand Tour artists. Finally, the map will permit the integration of this type of inquiry with the study of the itineraries of artists, architects, scholars and travellers of modern periods so that we may be able to reflect upon the accumulated knowledge derived from travel and movement through space.

We hope and expect that this project will have a significant impact on research, restoration, and the appreciation of the historic patrimony of South Italy. We are deeply concerned with the issue of monuments in the transmission of memory and identity, and the importance of their continued preservation and for restorations informed by the full range of documentary resources.

5. SITOGRAPHY

The Medieval Kingdom of Sicily Image Database (<http://kos.aahvs.duke.edu/index.php>)

Wired! Lab (<http://www.dukewired.org/>)

6. BIBLIOGRAPHY

- [1] Paola Vitolo, *Un contributo allo studio del patrimonio artistico e architettonico dell'Italia meridionale: il progetto The Medieval Kingdom of Sicily Image Database*, in *Sicily through foreign eyes: travelling architects. La Sicilia nello sguardo degli altri: architetti in viaggio*, ed. Paola Barbera and Maria Rosaria Vitale, Conference proceedings (Università degli studi di Catania, Dipartimento di Architettura, sede di Siracusa, May 18-19, 2017), in press.
- [2] Paola Vitolo, *Il Medioevo, il paesaggio, le città: evocazione, interpretazione, documentazione. Il progetto The Medieval Kingdom of Sicily Image Database*, in *La città, il viaggio, il turismo. Percezione, produzione, trasformazione*, ed. Gemma Belli, Francesca Capano, Maria Ines
- [3] Pascariello, introduction by Alfredo Buccaro and Fabio Mangone, AISU (Associazione Italiana di Storia Urbana), Conference proceedings (Naples, September 7-9, 2017), E-book 2017 (available on: www.iconograficiatitaeuropea.unina.it, section *Il Sud d'Italia tra schizzi e appunti di Viaggio. L'interpretazione dell'immagine, la ricerca di una identità*, ed. Bruno Mussari, Giuseppina Scamardi).
- [4] Caroline Bruzelius, *Visualizing the Medieval Past: The Kingdom of Sicily Image Database project*, in *Quei maledetti Normanni. Studi offerti a Errico Cuzzo per i suoi settant'anni da colleghi, allievi, amici*, ed. Jean-Marie Martin and Rosanna Alaggio, Ariano Irpino 2016, 109-116.
- [5] Caroline Bruzelius, Paola Vitolo, *The Kingdom of Sicily Image Database*, «Archeologia e Calcolatori», 27 (2016), 107-130.
- [6] Paola Vitolo, *The Kingdom of Sicily database project*, "Virtual Museum of Archeological Computing", <http://archaeologicalcomputing.isma.cnr.it/itineraries/> > Projects, 2016.
- [7] Caroline Bruzelius, *Teaching with Visualization Technologies. How does Information Become Knowledge?*, "Material Religion," 9/2 (2013), 246-253.

IDEA e la conservazione dei dati epigrafici di EAGLE

Pietro Liuzzo¹, Franco Zoppi², Antonio E. Felle³, Giuseppe Amato⁴

¹ Università di Amburgo - pietro.liuzzo@uni-hamburg.de

² CNR-ISTI - franco.zoppi@isti.cnr.it

³ DISUM – Università di Bari “Aldo Moro” - antonio.felle@uniba.it

⁴ CNR-ISTI - giuseppe.amato@isti.cnr.it

1. LA DIGITALIZZAZIONE DELLE ISCRIZIONI: METODI E TECNICHE

Poche discipline possono vantare di aver digitalizzato la quasi integrità del patrimonio documentario di loro interesse come l'epigrafia latina e greca (Orlandi et al. 2014, 2017). Con trent'anni di esperienza nella digitalizzazione delle iscrizioni, l'epigrafia ha ottenuto questo risultato, ma ha anche imparato molto sulla sostenibilità di tale impresa e sul continuo impegno che essa comporta.

Nel corso del progetto EAGLE (European Network for Ancient Greek and Latin Epigraphy) è stato per esempio realizzato un modello, basato sull'affermato standard TEI/EpiDoc (Amato et al. 2013, Manghi et al. 2015, Felicetti et al. 2016) in grado di garantire una mappatura a CIDOC-CRM e all'EMD (European Metadata Model). Questo lavoro ha reso possibile non solo lo sviluppo del portale EAGLE con le sue funzionalità fino ad allora solo auspiccate dalla comunità internazionale degli epigrafisti, ma ha anche permesso di raccogliere un significativo consorzio di partner e di produrre risorse aggiuntive, come ad esempio la Virtual Exhibition “Signs of Life” che raccoglie link a progetti, 3D, infografica in una introduzione per non addetti ai lavori (Liuzzo et al. 2017).

2. EAGLE E LA CONSERVAZIONE DEI DATI EPIGRAFICI

Il portale EAGLE è il cuore e il principale punto di accesso ai contenuti raccolti e preservati dal progetto non solo per i membri del preesistente *Electronic Archive of ancient Greek and Latin Epigraphy* (Epigraphische Datenbank Heidelberg - EDH¹, Epigraphic Database Roma - EDR², Epigraphic Database Bari - EDB³, Hispania Epigraphica Online - HEOI⁴) ma anche per altri partner quali Ubi erat lupa, Last Statues of Antiquity, Ancient Graffiti Project e molti altri.⁵

- In primo luogo l'aggregatore EAGLE raccoglie dati in EpiDoc XML, esportati dai vari progetti e database. Questo è un ricco set di informazioni, sebbene sia solo una parte dei dati archiviati nei database originali. Tutti questi dati non solo sono ricercabili allo stesso tempo, ma sono anche correttamente collegati tra di loro di modo da raccogliere edizioni multiple di un singolo testo sotto un unico risultato, grazie agli ID prodotti, mantenuti e controllati dal progetto Trismegistos. I dati sono accessibili a tutti e tutti possono contribuire i propri dati ad EAGLE, sebbene queste operazioni non siano accessibili se non a coloro che siano in possesso di competenze specifiche sia dal lato dell'architettura dei dati che da quello dell'epigrafia.
- Grazie ad un avanzato sistema di riconoscimento delle immagini, basato su tecniche di deep learning e metodi di aggregazione delle features (Amato 2016), è possibile cercare visivamente, a partire da un'immagine, il vasto archivio di foto di iscrizioni disponibile
- Sono state inoltre realizzate traduzioni dei testi incisi, sebbene presenti - nonostante gli sforzi - ancora soltanto per pochi dei testi originali digitalizzati.

3. IDEA: SOSTENIBILITÀ DELL'EPIGRAFIA DIGITALE

Si comprende come sia dunque disponibile ad oggi per tutti una gamma molto ampia di risorse e servizi sotto l'ombrello di EAGLE, tutti molto apprezzati dagli utenti di EAGLE: ma queste risorse e questi servizi sono da preservare, da aggiornare e da sviluppare costantemente. Per affrontare in modo organico questo problema - anche se non con la presunzione di risolverlo completamente - è stata fondata l'*International Digital Epigraphy Association* (IDEA), il cui scopo principale è appunto di mantenere quanto realizzato dal progetto EAGLE, ma non solo: l'intento di IDEA è anche quello di conservare unito il consorzio allargato grazie al progetto europeo e di pensare creativamente alle nuove prospettive per l'epigrafia, come ad esempio quelle recentemente delineate dai colleghi dell'*Epigraphic Database Heidelberg* (Feraudi-Gruénais and Grieshaber 2016).

IDEA ha come principale punto di forza il network costituito durante il progetto EAGLE, in quanto esso è probabilmente il prodotto più importante anche per la conservazione a lungo termine dei dati raccolti nei primi trent'anni di quella che oggi si definisce *Digital Epigraphy*. IDEA si propone di sostenere attivamente (come ha già fatto e continua a fare) nuovi progetti e progetti in corso di sviluppo, offrendo infrastrutture, funzionalità e distribuzione delle informazioni ai suoi membri. IDEA punta quindi non solo alle proprie risorse ma anche a sostenere quelle di altri: non intende essere un progetto “accentratore” ma, al contrario, un protettore dell'ecosistema delle diverse risorse attualmente disponibili, allo scopo di

1 <http://edh-www.adw.uni-heidelberg.de/home>

2 <http://www.edr-edr.it/default/index.php>

3 <http://www.edb.uniba.it/>

4 <http://eda-bea.es/>

5 <https://www.eagle-network.eu/eagle-project/partners/>

rendere il lavoro più efficiente con risultati più accurati e dunque più utili per tutti gli utenti, specialisti e non dell'epigrafia (Felle 2016).

Tutto questo comunque non è che una parte di quanto resta da fare: le risorse digitali che finora sono state realizzate non sono finalizzate solo ad archiviazione, o sterili conteggi, ma al loro uso creativo per produrre edizioni digitali affidabili, citabili, leggibili, aperte ai contributi degli specialisti ma anche accessibili per non addetti ai lavori.

Attualmente IDEA è in grado di offrire supporto sia tecnico che scientifico per la aggregazione, l'arricchimento e la conservazione di dati epigrafici digitali provenienti da fonti diverse.

Dal punto di vista tecnico, questo significa fornire le risorse hardware e software per i) ospitare l'insieme di servizi che compongono il portale EAGLE e l'infrastruttura di aggregazione ed indicizzazione sottostante; ii) implementare il "mapping" tra i modelli dei dati adottati da diversi content provider ed un modello di riferimento comune così come è stato realizzato nell'ambito del progetto EAGLE; iii) realizzare i "workflow" che eseguono l'integrazione dei dati e iv) effettuare una operazione di disambiguazione basata sugli identificatori prodotti da Trismegistos.

I dati così integrati vengono mantenuti in una infrastruttura di archiviazione costantemente monitorata e aggiornata con contenuti nuovi o semplicemente nuove versioni di contenuti già presenti che dovessero essere prodotti dai *provider*. Tutti i dati sono facilmente ricercabili sia usando metodologie tradizionali basate su interrogazioni testuali o strutturate, sia visivamente, usando altre immagini come interrogazioni.

Tutte le operazioni di gestione corrente (backup, manutenzione degli indici, ecc.) vengono effettuate sui sistemi di IDEA in modo automatico o, dove se ne presenti la necessità, con l'intervento del personale sistemistico di IDEA che ha in carico l'infrastruttura.

IDEA offre anche interessanti opportunità tecniche e scientifiche. Oltre alla ovvia ricchezza rappresentata dal materiale gestito tramite il portale EAGLE, a disposizione di tutti i ricercatori e appassionati della materia, IDEA è in continua ricerca di nuove soluzioni avanzate a problemi aperti quali, per esempio i) il miglioramento, l'integrazione, e l'armonizzazione semi-automatica dei metadati, ii) la generazione automatica di descrizioni ed annotazioni del materiale digitale disponibile controllata da un sistema di revisione e controllo specialistico interno, iii) la creazione semi-automatica di relazioni fra i contenuti indicizzati, iv) la visualizzazione organizzata ed ordinata dello stato dei lavori di digitalizzazione, edizione e ricerca in generale, a supporto delle discipline storiche e archeologiche interessate, v) l'integrazione con altre risorse digitali quali Wikidata.

Anche in questo primo anno di vita dell'associazione sono stati portati a termine alcuni progetti di utilità per tutti gli interessati, come un API per l'accesso agli ID allineati di Trismegistos (già in uso da parte di EAGLE stesso e dell'Epigraphic Database Heidelberg) e una app dimostrativa per l'edizione di nuovi corpora epigrafici online.

4. BIBLIOGRAFIA

- [1] Amato, G., P. Bollettieri, C. Gennaro, P. Manghi, and A. Mannocci 2013. 'AIM InfrastructureSpecification', Deliverable 4.1 of the EAGLE European Project, Grant agreement 325122
- [2] Amato, G., Falchi, F., and Vadicamo, L. "Visual recognition of ancient inscriptions using Convolutional Neural Network and Fisher Vector" *Journal on Computing and Cultural Heritage (JOCCH)* 9 (4), 21, 2016
- [3] Felicetti, Achille, and Francesca Murano. "Scripta Manent: A CIDOC CRM Semiotic Reading of Ancient Texts." *International Journal on Digital Libraries*, 2016. doi:10.1007/s00799-016-0189-z
- [4] Felle, Antonio E. 2016, "Oltre EAGLE: l'International Digital Epigraphy Association (IDEA). Una presentazione in anteprima", *Archeologia e Calcolatori* 27, 353-355 (http://www.archcalc.cnr.it/indice/PDF27/18_Note_Recensioni_Felle.pdf).
- [5] Feraudi-Gruénais, Francisca, and Frank Grieshaber. "Digital Epigraphy Am Scheideweg? / Digital Epigraphy at a Crossroads?" Düsseldorf, 2016. doi:10.11588/heidok.00022141.
- [6] Liuzzo, Pietro, Francesco Mambriani, and Philipp Franck. "Storytelling and Digital Epigraphy-Based Narratives in Linked Open Data." In *Mixed Reality and Gamification for Cultural Heritage*, 507-23. Springer, Cham, 2017. doi:10.1007/978-3-319-49607-8_20.
- [7] Manghi, P., A. Mannocci, M.A. Sicilia, J. Gomez Pantoja, J. Rubiro Fuentes, E. Rivero Ruiz, and F. Zoppi 2015. 'EAGLE metadata model specification – Second Release', Deliverable 3.1 of the EAGLE European Project, Grant agreement 325122.
- [8] Mannocci, A., V. Casarosa, P. Manghi, and F. Zoppi 2014. 'The Europeana network of ancient Greek and Latin epigraphy data infrastructure', *Metadata and Semantics Research - 8th Research Conference, MTSR 2014, Karlsruhe, Germany* (2014), 286-300.
- [9] Orlandi, Silvia, Raffaella Santucci, Francesco Mambriani, and Pietro Maria Liuzzo, eds. "Digital and Traditional Epigraphy in Context," Vol. 36. *Antichistica*. Roma: Sapienza Università Editrice, 2017. doi:10.13133/978-88-9377-021-7.
- [10] Orlandi, Silvia, Raffaella Santucci, Casarosa, and Pietro Maria Liuzzo, eds. *Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference*. Rom, 2014. <http://archiv.ub.uni-heidelberg.de/propylaeumdok/2337/>.

Identification of Documents in Written Cultural Heritage Digitalization – Description of Implemented Solution

Tomáš Klimek,¹ Tomáš Psohlavec², Olga Čiperová³

¹ Ph.D. National Library of the Czech Republic, Czech Republic - tomas.klimek@nkp.cz

² Aip Beroun s.r.o., Czech Republic - tomas.psohlavec@aipberoun.cz

³ Aip Beroun s.r.o., Czech Republic - olga.ciperova@aipberoun.cz

ABSTRACT

The present paper deals with the issue of physical document identification in the domain of written cultural heritage digitalization which, in the context of historical documents, occupies a different place compared to modern literary sources. The paper briefly presents an implemented solution to the unique and unambiguous identification of digitized historical documents - Index of Digitized Written Cultural Heritage Resources (IDHR). IDHR was created in relation to the development of a digital library, Manuscriptorium, giving access to historical documents (containing over 400,000 items) from more than 100 institutions all over Europe. IDHR functions as an autonomous system, independent of the digital library. The system enables to identify items of historical documents and of all of their digital representations via persistent identifiers. It uses information contained in descriptive metadata - these concern standard basic information about physical location of an item (repository, settlement, country, shelf marks) as well as information concerning the relevant digitalization centre, about the data repositories from which digital copies are made accessible, and identification of the copy itself. Also, the time scope during which the information used for the creation of the persistent identifier were valid, is recorded. The system has been designed to serve both the end users as well as historical book resources administrators and digital content creators. There are follow-up modules and systems which are linked with the core part of IDHR and which closely cooperate with it. In contrast to the existing solutions, IDHR enables to record all kinds of changes that may occur due to gifts, exchanges or sales of documents, repository transfers or reorganization of relevant institutions. IDHR gathers information about the location of digital copies in individual digital libraries or data repositories as well as about the conditions under which the documents can be viewed or further used. It also enables to create persistent repository identifiers. The system has been designed to cover the worldwide production of written cultural heritage digitalization and it has the ambition to provide a persistent overview of digitized historical documents and of all their related digital representations. Assuming that the system is further implemented successfully, we will be able to profit from a database which will provide a solid basis for work with digitized written cultural heritage.

KEYWORDS

Written cultural heritage - unique identification of documents - digital library - manuscripts - medieval literature - register.

1. INTRODUCTION

Unique identification of a specific document in its physical form is extremely important for digitalization of historical written cultural heritage; this being particularly true for medieval manuscripts. The reason is that research based on medieval literature cannot stem from the same identification process which we are used to when working with modern literary texts. [1] Medieval literature perceives the idea of authorship in a different way [2], identification of a text in different readings and in individual surviving documents proves to be very difficult. Many researchers speak about the fluidity of medieval texts containing a lot of variations, hypertexts, paratexts, compilations and integrally incorporated commentaries. [3] Therefore, every surviving document should be treated as unique, even though it can contain texts known from other sources, and its digitalization has to take into consideration every single item, in contrast to digital libraries for modern documents based on a title as such. [4]

When we were building the digital library Manuscriptorium (www.manuscriptorium.com), which gives access to historical documents (containing over 400,000 items) from more than 100 institutions all over Europe [5], we could not avoid coping with the problem of how to identify a physical historical document in digital environment satisfactorily; meaning such identification which would also provide an overview and basic information about all its available digital copies and metadata. To that end, we created a system independent of the digital library, however initially using Manuscriptorium as a data source. Regarding the fact that the issue of identification of physical documents in the area of written cultural heritage has not yet been treated in its complexity, we have designed the registration system, which shall be briefly described in the present paper, with the ambition to cover the worldwide production.

2. IDENTIFICATION OF AN ITEM ACCORDING TO ITS PHYSICAL PLACEMENT AND TECHNICAL SOLUTION OF SUCH IDENTIFICATION

In order to be able to find the right digital copy, we have to know how to search for the required document not only in terms of a bibliographical (descriptive) unit but also in the sense of a real physical item. However, the original document might have changed its repository, its owner, its title as well as its author identification. Moreover, its identification data stated in literature may not be valid any longer or the information required for a digital copy may not correspond to the way it is recorded for the purposes of research.

The absence of any official system of unique identifiers adapted to specific needs of historical documents description initially led to the creation of a specific identifier referred to as PhysID (physical identification). It was developed by the digital library Manuscriptorium (www.manuscriptorium.com) and it has been successfully used for 20 years on a national basis. Nevertheless, unique identification of a specific physical document based purely on this tool can be extremely difficult, if not impossible. Therefore, information contained in descriptive metadata must always be used when identifying a specific physical item in electronic systems. Descriptive metadata involve information mainly concerning physical storage place (repository, settlement, country) and its shelf mark. The PhysID, originally created for Manuscriptorium, comprises three parts: firstly, a unique storage place identifier of the document (LocID), secondly, standardized shelf mark or any other clear identification inside the repository, and thirdly, control code (code CRC32). Creation of the identifier also takes into consideration all kinds of specificities in the way shelf marks and other inventories of the items are reported in different libraries and archives such as the usage of upper and lower-case letters or other specific signs. [6] Even though such an identifier is unique, it cannot be considered as being persistent, i.e. it may not be (and often is not) permanently valid. If there is a change in the information contained in the identifier (change in repository, title, owner, shelf mark), the identifier of the given document changes as well.

3. BENEFITS OF INDEX OF DIGITIZED WRITTEN CULTURAL HERITAGE RESOURCES

The above-mentioned reasons led us to create Index of Digitized Written Cultural Heritage Resources (IDHR), a system enabling ongoing identification of items of historical documents and their digital representation via persistent identifiers. The system has been designed to serve both the end users as well as historical book resources holders and digital content creators.

The solution that we chose for IDHR contains, besides the original PhysID data, further information concerning the institution which carried out the process of digitalization, information about the data repositories from which digital copies are made accessible, and identification of the copy itself, e.g. the serial number. Also, the time scope of time during which the information used for the creation of the persistent identifier were valid is recorded. Any change in information serving to identify the documents is then put on record into a section called *Concordances*.

There were digital databases covering digitized modern archives (e.g. the Czech website registrdigitalizace.cz) before the IDHR came into existence. However, their main objective was to make the content of a document available in electronic format regardless the form of its medium (book). These databases were primarily designed to eliminate duplicity in digitalization of concrete titles and to provide information about the state of digitalization.

The system of digital copies identification in the area of written culture heritage (IDHR) obligatorily assigns persistent identifiers to every single physical item, to all of its digital copies and to all of its relevant electronic data (e.g. full texts, all kinds of descriptive records either converted from printed catalogues or created directly in digital environment).

4. PARTS OF THE SYSTEM AND FOLLOW-UP MODULES: DIGITAL CONCORDANCE, URI RESOLVER, REPOSITORY REGISTER

There are follow-up modules and systems which are linked with the core part of IDHR and which closely cooperate with it – *Digital Concordance*, *URI Resolver* and *Repository Register*. The first of them, *Digital Concordance*, ensures persistence of references, it records all kinds of changes that may occur due to gifts, exchanges or sales of the documents, repository transfers or reorganization of relevant institutions. Sometimes, changes may be related to the fact that institutions are cancelled or newly opened, there may also be changes due to war and natural disasters. *Digital Concordance* enables to find out new places of repositories, owners and shelf marks. The record in the main IDHR database can thus be accompanied by a substantial number of concordant records.

Another autonomous database is *URI Resolver*, which gathers information about the location of digital copies in individual digital libraries or data repositories as well as about the conditions under which the documents can be viewed or further used. Any possible changes of data repositories or changes of URL sites of digital libraries are recorded here.

The last external part of IDHR, *Repository Register*, has been designed to create repository identifiers – LocID. They accompany the identifier of the physical document (PhysID). The system provides its services to IDHR as well as to digital libraries and individual listers. The system ensures that repositories and their identifiers are administered centrally and that the content of the database is correct and complete.

5. CONCLUSION

A satisfactory solution to the issue of physical document identification is an essential prerequisite for the further successful development of digitalization of written cultural heritage. The IDHR system is ready to be filled in with relevant data and in the future, it has the ambition to provide a persistent overview of digitized historical documents and of all their related digital representations worldwide. Assuming that the system is implemented successfully, we will be able to profit from a database which will provide a solid basis for work with digitized written cultural heritage.

6. REFERENCES

- [1] Scharpe, R. 2003. *Titulus. Identifying Medieval Latin Texts. An Evidence-Based Approach*. ISBN: 978-2-503-51258-7. Turnhout.
- [2] Škarka, A. 1986. *Literatura bez autorů a bez generací*, in Idem, *Půl tisíciletí českého písemnictví*, Prague, p. 11.

- [3] Uhlř, Z. 2010. *Digital Codicology and Contextual Editing of Medieval Manuscripts*. Revista arhivelor. 01/2009. 86(2):36-54. <http://www.arhivelenationale.ro/images/custom/image/serban/2012/RA%202%202009%20pdf%20cu%20watermark/06%20uhlir,%20zdenek-ingleza.pdf>. The idea of text layers is promoted by Thomas O'Sullivan: O'Sullivan, T. 2011. *Defining and Editing the Shifting Shapes of Sermons. Texts Worth Editing*. The Seventh International Conference of the European Society for Textual Scholarship. Pisa, 26 November 2010 [http://67.23.4.192/Pisa_Presentation.wmv] [25 April 2011]. Compare to: Echard, S. – Partridge, S. 2004. *The Book Unbound. Editing and Reading Medieval Manuscripts and Texts*. ISBN 0-8020-8756-6. Toronto - Buffalo – London; Ziolkowski, J. 2001. *Text and Textuality, Medieval and Modern*. Der unfeste Text. Hrsg. Sabel, B. – Bucher, A. ISBN 3-8260-1968-7. Würzburg, pp. 109 – 131; D. Caie, G. 2000. *Hypertext and multiplicity: the medieval example*. The Renaissance text. Theory, Editing, Textuality. (Ed.) Murphy, A. ISBN 0 7190 53986. Manchester – New York, pp. 30 – 43.
- [4] Tredinnick, L. 2006. *Digital Information Contexts: Theoretical Approaches to Understanding Digital Information*. ISBN 1 84334 169 7. Oxford.
- [5] [Http://www.manuscriptorium.com](http://www.manuscriptorium.com); Uhlř, Z. 2015. Digitization is not only making images: manuscript studies and digital processing of manuscripts. *Book Science*. 51, pp. 148 – 162; Uhlř, Z. – Knoll, A. 2009. Manuscriptorium digital library and ENRICH project: Means for dealing with digital codicology and palaeography. *BoD*, pp. 67 – 78.
- [6] Details for setting a unique physical document identifier are available on Manuscriptorium web pages: <http://www.manuscriptorium.com/cs/content/definice-digitalniho-dokumentu-pro-potreby-visk6>.

Il Progetto Traduzione del Talmud Babilonese: il Ruolo della Tecnologia e della Linguistica Computazionale

Emiliano Giovannetti¹, Davide Albanesi¹, Andrea Bellandi¹, David Dattilo²,
Michael Dollinar², Alessandra Pecchioli³, Clelia Piperno²

¹ Istituto di Linguistica Computazionale "A. Zampolli" - CNR, Italia - nome.cognome@ilc.cnr.it

² Progetto Traduzione Talmud Babilonese S.c.a r.l., Italia - nome.cognome@talmud.it

³ Progetto Traduzione Talmud Babilonese S.c.a r.l., Italia - alepec3@gmail.com

ABSTRACT

L'obiettivo principale del Progetto Traduzione del Talmud Babilonese è produrre la traduzione del Talmud in italiano. La traduzione, affidata ad un team di circa 80 studiosi, è condotta con l'aiuto di Traduco, un software preposto ad agevolare tutte le fasi di lavoro previste dal progetto, dall'attribuzione degli utenti alle sezioni da tradurre, fino al supporto all'impaginazione finale. La presenza di una piattaforma collaborativa digitale che già, di per sé, costituisce una innovazione nell'ambito dei grandi progetti di traduzione, è arricchita da algoritmi per il trattamento automatico del testo e della lingua, in costante evoluzione, attraverso i quali il traduttore, il revisore o lo studioso possono contare su funzionalità sempre più avanzate.

PAROLE CHIAVE

Linguistica Computazionale, Traduzione di Testi Religiosi, Traduzione Assistita dal Calcolatore, Traduzione Collaborativa

1. INTRODUZIONE

Il "Progetto Traduzione del Talmud Babilonese" (PTTB) nasce da un Protocollo d'Intesa tra la Presidenza del Consiglio dei Ministri, il MIUR, il CNR, e l'UCEI - Collegio Rabbinico Italiano. La traduzione del Talmud in italiano è affidata ad un team di traduzione formato da circa 80 studiosi, fra traduttori esperti, traduttori in formazione, istruttori, revisori di contenuto e revisori editoriali.

2. IL RUOLO DELLA TECNOLOGIA

L'aspetto innovativo del progetto, che lo differenzia da tutte le altre traduzioni del Talmud e di altri testi antichi avvenute nel mondo, è la presenza di Traduco, la piattaforma digitale preposta ad agevolare tutte le fasi di lavoro previste dal progetto, dall'attribuzione di utenti a specifiche porzioni del testo da tradurre fino alle ultime fasi di impaginazione [4] [7]. Traduco è sviluppato, e costantemente aggiornato, dal 2012 (anno di avvio del progetto), da parte di un team di informatici e linguisti computazionali presso l'Istituto di Linguistica Computazionale del CNR di Pisa.

Una rassegna dei sistemi informatici di CAT (Computer-Assisted Translation) sviluppati in ambito di ricerca accademica e industriale (tra cui: Across¹, Déjà Vu², memoQ³, SDL Trados⁴, Similis⁵, Transit NXT⁶, Wordfast⁷, OpenTM⁸, OmegaT⁹, Transolution¹⁰, Matecat¹¹, TinyTM¹²) evidenzia la presenza di sistemi massimamente orientati alla velocizzazione del processo di traduzione, e realizzati per la traduzione di testi di natura prevalentemente manualistica e legislativa.

Opere complesse e articolate, come il Talmud babilonese, necessitano tuttavia di applicazioni specializzate in grado di assistere una molteplicità di utenti (traduttori, revisori, redattori e supervisori) in un delicato e articolato processo interpretativo-traduttivo [3].

Pur progettato secondo criteri di modularità e riusabilità, Traduco è stato sviluppato a partire dai requisiti utente provenienti dal team dall'Unione delle Comunità Ebraiche Italiane e del Collegio Rabbinico Italiano. A partire da tali requisiti l'ILC-CNR ha implementato la piattaforma e, come, introdotto più avanti, ne ha arricchito le funzionalità attraverso tecniche di trattamento del testo e della lingua. Traduco consente di marcare il testo su base semantica (e creare automaticamente rubriche tematiche per nomi di rabbini, idiomi, espressioni linguistiche, misure, nomi di piante, ecc.), aggiungere note di varia natura e riferimenti bibliografici, distinguere tra traduzioni letterali e aggiunte esplicative. Permette inoltre agli utenti di organizzare il proprio flusso di lavoro strutturando gerarchicamente il testo in traduzione e avvalendosi di strumenti di

¹ <http://www.my-across.net/en/> (ultimo accesso: 20/12/2017)

² <http://www.atril.com/> (ultimo accesso: 20/12/2017)

³ <https://www.memoq.com/> (ultimo accesso: 20/12/2017)

⁴ <http://www.sdl.com/> (ultimo accesso: 20/12/2017)

⁵ <http://www.similis.fr/> (ultimo accesso: 20/12/2017)

⁶ <http://www.star-ts.com/> (ultimo accesso: 20/12/2017)

⁷ <http://www.wordfast.com/> (ultimo accesso: 20/12/2017)

⁸ <http://www.opentm2.org/> (ultimo accesso: 20/12/2017)

⁹ <http://www.omegat.org/> (ultimo accesso: 20/12/2017)

¹⁰ https://bitbucket.org/fredrik_corneliusson/transolution/ (ultimo accesso: 20/12/2017)

¹¹ <https://www.matecat.com/> (ultimo accesso: 20/12/2017)

¹² <http://tinytm.sourceforge.net/en/index.html> (ultimo accesso: 20/12/2017)

supporto che facilitano il processo di traduzione, inserire commenti e citazioni, effettuare ricerche complesse all'interno del testo e, soprattutto, ricevere suggerimenti alla traduzione sulla base del testo già tradotto. Traduco, in quanto realizzato come applicazione Web, è intrinsecamente collaborativo: utenti connessi da luoghi diversi (e con ruoli diversi) possono lavorare insieme, simultaneamente, alla traduzione dello stesso testo. Traduco, inoltre, include un pannello di supervisione dal quale è possibile monitorare in tempo reale lo svolgimento del lavoro effettuato, sia in termini di percentuale di completamento della traduzione del testo, che di lavoro svolto da ogni singolo utente.

Di particolare rilevanza, soprattutto nelle ultime fasi di lavoro che preludono alla stampa dei volumi cartacei, è il componente di supporto all'editoria. Attraverso tale modulo del sistema il redattore e l'editore hanno la possibilità di vedere a schermo, in tempo reale, la pagina nella forma che assumerà una volta stampata.

Traduco gira su connessione sicura (https) ed è dotato di sistemi a ridondanza multipla e geograficamente distribuiti in grado di garantire la massima sicurezza delle traduzioni conservate nel suo database.

3. LA RICERCA AL SERVIZIO DELLA TRADUZIONE

Come si richiede ad un sistema sviluppato all'interno di un istituto di ricerca, Traduco è in costante evoluzione. Nonostante sia già maturo per supportare i cinquanta utenti del sistema nelle varie fasi di traduzione del Talmud, il team di progetto è continuamente al lavoro per studiare nuove soluzioni che ne potenzino e arricchiscano le funzionalità.

Traduco, dal momento che è sviluppato dall'Istituto di Linguistica Computazionale del CNR, integra algoritmi per il trattamento automatico del testo e della lingua. Le funzionalità di elaborazione e di accesso al testo, basate su tecnologie in linea con lo stato dell'arte, sono costantemente aggiornate ed estese per dotare il traduttore, revisore o lo studioso di soluzioni sempre più efficienti ed efficaci. Al cuore di Traduco, come in ogni sistema di CAT, c'è il componente di suggerimento delle traduzioni: il sistema è in grado di fornire all'utente, in modo automatico, suggerimenti di possibili traduzioni prodotte sulla base del repertorio di traduzioni già effettuate, raccolte in una "memoria di traduzione" (MT), sia dall'utente stesso che da altri utenti. La MT, dopo cinque anni di lavoro, è oggi così ricca da consentire al sistema di suggerire automaticamente, una volta su due, una traduzione esatta. Per migliorare la qualità dei suggerimenti alla traduzione che il sistema propone la MT è stata arricchita con informazione semantica attraverso un approccio basato sulla semantica distribuzionale. In particolare, la MT è stata arricchita con coppie di parole che l'algoritmo distribuzionale adottato [1] ha individuato come semanticamente correlate (sinonimi, varianti ortografiche, flessioni morfologiche, ecc.) e la misura di similarità utilizzata è stata conseguentemente modificata [5].

Inoltre, è in corso l'integrazione di tecniche di analisi automatica della lingua ebraica volte al potenziamento del suggeritore, che potrà confrontare le parole delle frasi in traduzione su base lemmatica [9].

Sono state anche sperimentate tecniche di analisi stilistica volte a ridurre il lavoro di revisione successivo alla traduzione [8]. Per questo studio sono state quantificate le modifiche che ogni traduzione ha sostenuto nel suo "ciclo di vita" (dalla prima traduzione inserita dal traduttore all'ultima validata dal curatore) e sono state, quindi, confrontate con un gran numero di caratteristiche linguistiche (testuali, lessicali, morfologiche, sintattiche) calcolate sulle ultime versioni. I risultati ottenuti hanno mostrato una effettiva correlazione tra l'entità delle modifiche effettuate e la conformità ad alcune specifiche caratteristiche linguistiche (in particolare di natura morfo-sintattica che includono verbi), dimostrando che supportare i traduttori nel processo traduttivo seguendo uno stile desiderato potrebbe effettivamente ridurre il numero di interventi necessari (e, quindi, risparmiare tempo) da parte di revisori, redattori e curatori.

4. PROSPETTIVE

Attraverso opportuni adattamenti, e con relativa facilità, è possibile modificare Traduco affinché supporti nella traduzione di altri testi e in altre lingue. Molti dei componenti di trattamento del testo, infatti, sono agnostici alla lingua, e quelli specificamente costruiti per l'analisi automatica di una determinata lingua possono essere "riallenati" per essere adattati al trattamento di nuovi idiomi. La seconda fase del progetto avviata nel 2017 prevede che, oltre alla traduzione del Talmud, si lavori, nei cinque anni successivi, anche alla formazione. In particolare, tra le varie attività educative previste, si capitalizzerà sull'esperienza maturata nel corso dei primi cinque anni di progetto per dotare Traduco di componenti software specificamente orientati all'insegnamento delle lingue ebraica e aramaica e allo studio del Talmud, anche attraverso l'adozione di approcci lessicografico-computazionali e per la rappresentazione della conoscenza [2]. Si prevede, infatti, di strutturare la terminologia talmudica in una risorsa termino-ontologica che consenta di rappresentare, allo stesso tempo, i termini chiave del Talmud e le entità nominate (persone, luoghi, ecc., già disponibili nei glossari tematici) e i relativi concetti. Il piano terminologico sarà implementato attraverso il modello lessicale *lemon*¹³ [6] e quello concettuale, almeno inizialmente, con il linguaggio SKOS¹⁴ (Simple Knowledge Organization System).

5. RICONOSCIMENTI

Questo lavoro è stato condotto nel contesto del progetto di ricerca TALMUD tramite la partnership scientifica tra S.c.a r.l. "Progetto Traduzione del Talmud Babilonese" (PTTB) e ILC-CNR sulla base del "Protocollo d'Intesa" tra la Presidenza Italiana del Consiglio dei Ministri, il Ministero dell'Educazione, dell'Università e della Ricerca, l'Unione delle Comunità Ebraiche Italiane, Il Collegio Rabbinico Italiano e il Consiglio Nazionale delle Ricerche (21/01/2011).

¹³ <http://lemon-model.net/> (ultimo accesso: 20/12/2017)

¹⁴ <https://www.w3.org/TR/skos-reference/> (ultimo accesso: 20/12/2017)

6. BIBLIOGRAFIA

- [1] Baroni, M., and Lenci, A. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.
- [2] Bellandi, A., Bellusci, A., Carniani, E., Giovannetti, E. 2014. Content Elicitation: Towards a New Paradigm for the Analysis and Interpretation of Text. In *Proceedings of the 13th IASTED International Conference on Software Engineering*, Innsbruck.
- [3] Bellandi, A., Albanesi, D., Benotto, G., Giovannetti, E., G. Di Segni. 2015. When Translation Requires Interpretation: Collaborative Computer-Assisted Translation of Ancient Texts. In *Proceedings of LaTeCH 2015*, pp. 84-88, Beijing, China, July 30, 2015.
- [4] Bellandi, A., Bellusci, A. 2015. Towards a Translation Platform as a Bridge Between Ancient and Modern Languages. In A. Bozzi (ed.), *Digital Texts, Translations, Lexicons in a Multi-Modular Web Application: Methods and Samples*, B.A.R. II, Vol. LX Leo Olschki editore, Firenze.
- [5] Bellandi, A., Benotto, G., Di Segni, G., and Giovannetti, E. 2016. Investigating the Application and Evaluation of Distributional Semantics in the Translation of Humanistic Texts: a Case Study. In *Proceedings of the 2nd Workshop on Natural Language Processing for Translation Memories (NLP4TM 2016)*. 28 May 2016. Portorož (Slovenia). pp. 6-11.
- [6] Declerck, T., Buitelaar, P., Wunner, T., McCrae, J., Montiel-Ponsoda, E., and Aguado de Cea, G. 2010. Lemon: An Ontology-Lexicon model for the Multilingual Semantic Web. In *Proceedings of the W3C Workshop: The Multilingual Web - Where Are We?*, 26-27/10/2010, Madrid.
- [7] Giovannetti, E., Albanesi, A., Bellandi, A., and Benotto, G. 2017. Traduco: A collaborative web-based CAT environment for the interpretation and translation of texts. *Digital Scholarship Humanities* (2017) 32 (suppl_1): i47-i62, Oxford Press University.
- [8] Giovannetti, E., Albanesi, D., Bellandi, A., Dattilo, D., Dell'Orletta, F. 2017. Stylometry in Computer-Assisted Translation: Experiments on the Babylonian Talmud. To appear in *Proceedings of the Fourth Italian Conference on Computational Linguistics (Clic-it 2017)*, 11-13 December 2017, Rome, Italy.
- [9] Pecchioli, A. 2017. Elaborazione del linguaggio naturale (NLP) in Ebraico: il caso dell'analisi linguistica automatica applicata all'ebraico mishnaico del Talmud. XXXI Convegno AISG 2017 - Nuovi studi sull'Ebraismo, Ravenna, 4-6 settembre 2017.

Integrazione di QR-code e localizzazione di prossimità in visite educative museali basate su pervasive gaming

Giovanni Luca Dierna¹, Alberto Machi²

¹ Istituto di Calcolo e Reti ad Alte Prestazioni, Laboratorio di Robotica e Social Sensing - CNR, Italia - giovanniluca.dierna@icar.cnr.it

² Istituto di Calcolo e Reti ad Alte Prestazioni, Laboratorio di Robotica e Social Sensing - CNR, Italia - alberto.machi@icar.cnr.it

ABSTRACT

L'articolo descrive i primi esperimenti relativi all'attività di prototipazione e sviluppo di un'applicazione su terminale mobile di ausilio alla scoperta e contestualizzazione di reperti durante visite tematiche di un museo. Lo scopo è di carattere educativo ed è quello di far conoscere, tramite il gioco, i reperti museali e la loro contestualizzazione storica, geografica e funzionale. Tutto ciò viene realizzato integrando tecnologie digitali web, codici QR ed un'infrastruttura tecnologica composta da radiofari BLE (beacon), posizionati in prossimità di oggetti di interesse del museo. In particolare vengono qui illustrati i primi risultati scaturiti da un'attività di sperimentazione e test sul campo che ne metta in evidenza aspetti positivi, criticità, possibili miglioramenti e questioni legati agli aspetti di *user interface* e *interaction design*.

PAROLE CHIAVE

Caccia al Reperto, *pervasive gaming*, *gamification*, QR-code, BLE beacon, localizzazione di prossimità, sperimentazione.

1. INTRODUZIONE

L'articolo mostra i primi risultati della sperimentazione di itinerari educativi museali attraverso il *pervasive gaming*, cioè l'apprendimento, l'intrattenimento con il gioco e il coinvolgimento emotivo, allo scopo di conoscere il patrimonio culturale attraverso l'integrazione della scoperta / visione di manufatti e racconti [1,2,3,4,5].

Ciò può essere infatti raggiunto attraverso la cosiddetta *gamification* ovvero l'uso delle dinamiche tipiche dei giochi e delle tecniche di game design, in contesti esterni ai giochi, in particolare nel campo della valorizzazione e della diffusione del patrimonio storico, artistico, scientifico e ambientale, in particolare mediante il coinvolgimento ludico della "caccia al tesoro" [6].

Gli esperimenti mirano a fornire al visitatore un'esperienza di *roleplaying* coinvolgente in riferimento al contesto di riferimento, finalizzato all'apprendimento [7].

Sono stati previsti tre livelli di visita (di cui due in contesti non prettamente scolastici) e di fruizione (in funzione dei livelli di approfondimenti scelti):

- la visita di un ragazzino accompagnato dalla sua famiglia, con un basso livello di approfondimento dei contenuti dell'app (cui corrisponde il primo test), condotta in occasione della Giornata Nazionale delle Famiglie al Museo 2017;
- la visita di un gruppo di ragazzini, eventualmente accompagnati dai loro genitori, con un livello intermedio di approfondimento dei contenuti dell'app (cui corrisponde il secondo test);
- la visita di una scolaresca di una classe V della scuola primaria, accompagnata dagli insegnanti, con rivisitazione dei contenuti in classe via web (test da effettuare).

Gli esperimenti si inseriscono nella sperimentazione didattica descritta in [16], nata nell'ambito del Progetto Scuola-Museo, promosso dall'Assessorato Regionale dei Beni Culturali e dell'Identità Siciliana della Regione Sicilia, e svoltasi presso il Museo Archeologico Regionale di Camarina, in provincia di Ragusa (Italia), e sono consistiti in un percorso di visita museale legato alla (ri)scoperta di manufatti e reperti archeologici e alla loro osservazione e contestualizzazione. Le illustrazioni nel seguito si riferiscono al tema "Usanze funerarie".

2. IDEA E STATO DELL'ARTE

L'idea che sta alla base della sperimentazione è quella di rendere i contenuti museali e più in generale dei beni culturali *family friendly*, e in particolare facilmente fruibili da un pubblico in età scolare, mediante l'uso di tecnologie innovative e appropriate [8,9,10,11].

E' noto infatti che attività miranti allo stesso scopo sono spesso messe in pratica dai musei e da organizzazioni di volontari e vengono svolte prevalentemente con modalità cartacea tradizionale, con la logica cioè della caccia al tesoro, sfruttando indizi cartacei e suggerimenti cartacei durante il percorso di visita.

Tali percorsi mirano a sviluppare nei bambini sviluppo della curiosità, capacità di osservazione ragionata, capacità di lettura e comprensione dei testi, memoria, capacità investigativa, conoscenza, lettura e analisi di opere e testimonianze del patrimonio culturale e, se operato a squadre, spirito di squadra, interazione e comunicazione verbale con gli altri partecipanti. Essi inoltre inducono a scoprire dettagli, svelare misteri e rispondere a domande, quiz e indovinelli, non di rado accompagnati da un personaggio virtuale del passato o da un cicerone a fumetto. Infine, a volte, vengono associati file multimediali (audio/video) relativi a storie e curiosità o percorsi interattivi, per esempio tattili.

L'utilizzo dei codici-QR e dei radiofari BLE bluetooth in ambito museale è allo stato attuale limitato alla lettura di contenuti di approfondimento dei beni e alla notifica di servizi correlati, secondo la logica della *beacon proximity*, ovvero utilizzando

il modello del posizionamento per prossimità; meno spesso è utilizzato a supporto di giochi coinvolgenti per la visita di aree museali.

Nella sperimentazione in corso si vogliono combinare le due tecnologie QR-code e beacon bluetooth per la creazione di percorsi ludici museali, dopo aver validato la tipologia di percorso culturale con esperti di dominio, allo scopo di verificarne, con un pubblico reale, funzionalità, usabilità dell'interfaccia utente dell'applicazione all'uppo sviluppata, ecc.

3. LA TECNOLOGIA UTILIZZATA

La tecnologia utilizzata ha fatto uso di:

- ✓ una webapp su dispositivo (tablet) Android che svela progressivamente 12 sezioni di un documento descrivente la tematica della campagna di caccia;
- ✓ QR-code posizionati in corrispondenza di ogni reperto di potenziale interesse per la tematica esposto in prossimità del percorso di caccia;
- ✓ radiofari BLE bluetooth mimetizzati o inseriti in corrispondenza (o all'interno) di reperti di interesse potenziale per la tematica lungo il percorso di caccia;

Una personalizzazione della caccia è ottenuta selezionando, in modo pseudo casuale, l'associazione degli oggetti di interesse ai rispettivi QR-code o ai rispettivi beacon. Gli oggetti associati a QR-code vanno cercati, gli altri si palesano attraverso messaggi di allerta quando si giunge nelle loro vicinanze.

La scelta di un'applicazione funzionante su mobile (smartphone, tablet) è motivata dal fatto che trattasi di una tecnologia di uso comune e adatta a ogni fascia di età [12,13].

La scelta dei QR-code, già utilizzati nei musei [14] e beacon [17] è stata operata in funzione dell'efficacia, praticità di installazione degli stessi, non essendo invasivi per l'ambiente e non prevedendo interventi strutturali o necessità di cablaggio, non sempre pensabili nelle strutture museali attuali, che sono spesso edifici storici e antichi.

L'app consente di costruire passo dopo passo la storia tematica sotto forma di racconto secondo la logica dello *storytelling* [17,18], sulla base di indizi legati alle caratteristiche degli oggetti e all'ubicazione nel Museo (es. Padiglione dell'Archeologia Subacquea, Padiglione Est, Padiglione Ovest, Padiglione delle Anfore Arcaiche, cortile interno). La storia, suddivisa in tante sezioni quanti sono gli oggetti da trovare, racconta un tema che cita e lega secondo un filo retorico senso tutti i reperti da scoprire. La scoperta di ogni oggetto attiva una sezione del documento stesso e il racconto si va così via via componendo fino a essere per così dire decifrato e svelato nella sua interezza, quindi leggibile.

L'applicazione è strutturata secondo diversi livelli e possibilità di gioco. Essa in particolare prevede:

- l'uso di una mappa con un percorso suggerito;
- la possibilità della sfida a squadre, per giocare in gruppo (non sperimentata negli esperimenti qui descritti e prevista in un prossimo);
- la modalità di rinforzo e verifica delle conoscenze apprese (caratteristiche oggetto, sua contestualizzazione storica, geografica e funzionale, contestualizzazione con l'area archeologica circostante e comprensione del testo riguardante la storia tematica, attraverso quiz a risposta multipla);
- la fruizione dei contenuti testo/audio/video;
- l'acquisizione di punteggi, al fine di indurre alla competizione e alla gratificazione.

L'interfaccia grafica è allo stato attuale ottimizzata al solo sfruttamento ottimale degli spazi sulla finestra grafica disponibile, in quanto scopo del lavoro è la verifica della metodologia e non lo studio di presentazioni particolarmente accattivanti.

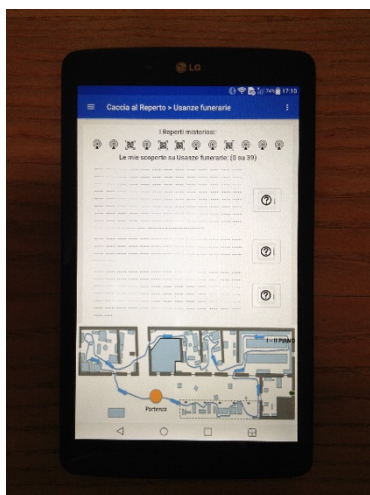


Figura 1. Immagine dell'interfaccia grafica iniziale dell'app con visualizzazione mappa attivata

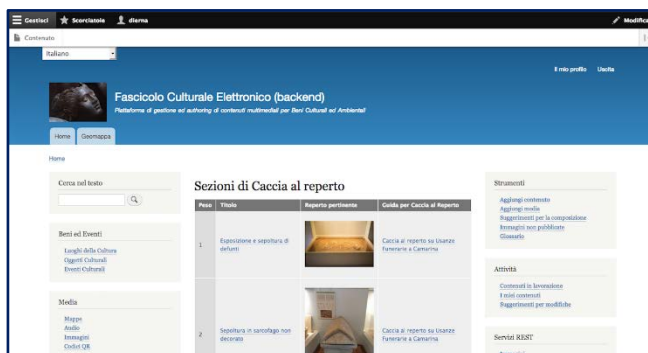


Figura 2. Un'immagine del Fascicolo Culturale Elettronico (backend)

I contenuti dell'app sono stati strutturati e riempiti attraverso un'apposita piattaforma web di gestione e sintesi di contenuti multimediali denominata "Fascicolo Culturale Elettronico (backend)" che ha visto il contributo di redattori, revisori, editori, ontologi e traduttori.

Nel I esperimento descritto nel presente lavoro, è stato sperimentato solo il livello di approfondimento base che prevede solamente le domande di verifica finale sulla comprensione del testo finale, con l'acquisizione del relativo punteggio. Il tablet con l'app è stato consegnato al bambino, dopo un'introduzione sul suo uso, la consegna delle istruzioni su supporto cartaceo e un giro preliminare di perlustrazione del museo.

Nel II esperimento, è stato sperimentato un livello di approfondimento intermedio che prevede le domande di verifica sulla conoscenza del reperto (caratteristiche fisiche e sua contestualizzazione storica, geografica e funzionale) e comprensione del testo finale, con l'acquisizione del relativo punteggio, con l'esclusione della sola lettura delle schede di approfondimento (contenenti anche sezioni audio), destinate in particolare alla rivisitazione didattica dei contenuti prevista nel terzo test.

I QR-code sono stati posti in punti ben visibili e ad un'altezza congrua per i bambini.



Figura 3. Esempio di QR-code associato a un reperto (altezza piedistallo: 147,7 cm)



Figura 4. Esempi di BLE beacon associato a un reperto

I radiofari bluetooth sono stati posizionati in funzione dell'ambiente, alcuni posti nelle pareti a 2 metri dal suolo, altri posti sulle pareti in corrispondenza del reperto, altri ancora nascosti. Essi sono stati configurati in maniera da comunicare all'utente delle notifiche in funzione della prossimità al reperto, prevedendo tre soglie in riferimento alla potenza del segnale percepito (-95 dB, -80 dB e -70 dB nel primo test, valori diversi per ciascun reperto in funzione della posizione del radiofaro e della geometria dell'ambiente e comunque compresi tra -93 dB e -59 dB), calibrate in maniera tale da non creare sovrapposizione fra le aree di prossimità critica (*immediate*).

Le soglie hanno dunque permesso di identificare in tutto il museo, in relazione al singolo reperto, quattro aree;

- ✓ area di non percezione del segnale (potenza del segnale < -95 dB);
- ✓ area *far* (-95 dB < potenza del segnale < 80 dB nel I test, valori dipendenti dal reperto considerato nel II test e comunque compresi tra -93 dB e -86 dB);
- ✓ area *near* (-80 dB < potenza del segnale < 70 dB nel I test, valori dipendenti dal reperto considerato nel II test e comunque compresi tra -85 dB e -66 dB);
- ✓ area *immediate* (potenza del segnale > 70 dB nel I test, valori dipendenti dal reperto considerato nel II test e comunque compresi tra -76 dB e -59 dB).

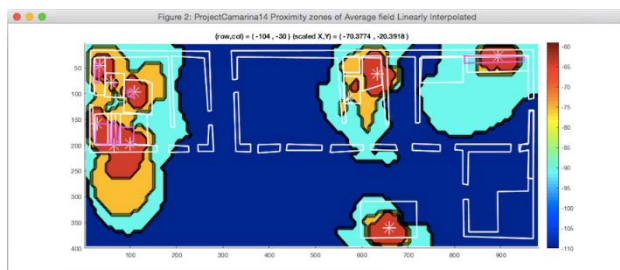


Figura 5. Zone di percezione del segnale bluetooth in corrispondenza dei reperti: zone di prossimità *far* (ciano)–*near* (giallo)–*intermediate* (rosso)



Figura 6. La notifica "far" associata a due reperti

Tali valori nell'interfaccia grafica sono stati associati dei colori di riferimento:

- ✓ area di non percezione del segnale: nessun colore nel I test, colore nero nel II test;
- ✓ area *far*: colore verde;
- ✓ area *near*: colore giallo
- ✓ area *immediate*: colore rosso.

4. L'ESPERIMENTO CON UN SOLO RAGAZZINO E LA SUA FAMIGLIA IN OCCASIONE DELLA GIORNATA NAZIONALE DELLE FAMIGLIE AL MUSEO

Il I esperimento ha visto il coinvolgimento di una famiglia intera ed in particolare di un bambino di 10 anni frequentante la V classe di una Scuola Primaria. Da notare che tale famiglia non aveva mai visitato il Museo di Camarina.

Scopo dell'attività di sperimentazione è stato quello di verificare preliminarmente le modalità di fruizione ludica dei beni culturali, testando contemporaneamente:

- ✓ le ricadute didattiche dell'operazione;
- ✓ l'usabilità dell'applicazione;
- ✓ l'efficacia della metodologia (*gamification*) adottata in termini di coinvolgimento dell'utente;
- ✓ la tecnologia di localizzazione mediante radiofari bluetooth;
- ✓ l'uso ottimale dei QR-code in funzione degli scopi prefissati.

4.1 LE FASI DEL I ESPERIMENTO

L'esperimento si è svolto seguendo le seguenti fasi:

- ✓ introduzione all'uso dell'app, alla sua logica e consegna delle istruzioni cartacee;
- ✓ giro preliminare perlustrativo del museo, al fine di indicare tutte le aree (interne ed esterne) accessibili durante la caccia;
- ✓ la caccia al reperto vera e propria, partendo dai suggerimenti testuali associati ai singoli reperti;
- ✓ lettura del testo della storia tematica e risposta alle domande di comprensione finali;
- ✓ successivo approfondimento mediante lettura pannello esplicativo presente nel museo;
- ✓ saluti finali e foto ricordo della giornata.

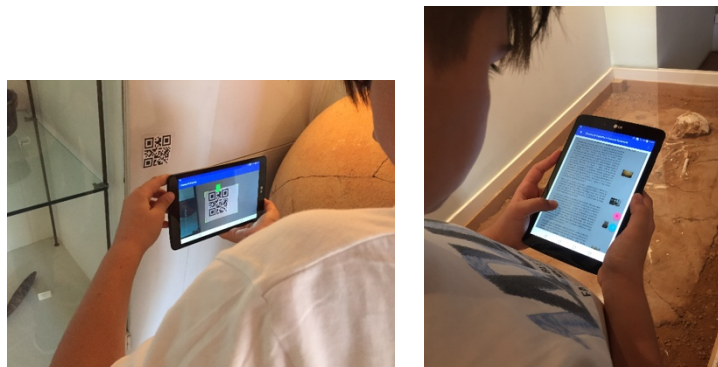


Figura 7. Alcune immagini del I esperimento

4.2 RISULTATI DEL I TEST EFFETTUATO E CONSIDERAZIONI RELATIVE

Il test ha riguardato il livello di approfondimento base prima citato e ha permesso di valutare:

- ✓ l'usabilità dell'interfaccia grafica;
- ✓ il livello di comprensione dei testi;
- ✓ la gestione dei QR-code e dei beacon;
- ✓ la gestione delle soglie di ricezione dei segnali emessi dai beacon;
- ✓ la durata della caccia;
- ✓ il tempo di mantenimento dell'attenzione ed il grado di coinvolgimento emotivo.

Sulla base dei risultati ottenuti, si è potuto constatare che:

- ✓ l'applicazione è risultata di facile uso e adatta alle nuove generazioni che, è noto, sono avvezze e attratte dall'uso del tablet e smartphone nella vita quotidiana, in quanto nativi digitali; a tal proposito occorre aggiungere che le sorelle, di età minore, seppur non accedendo ai contenuti dell'app, hanno seguito, incuriositi, il fratello nel percorso;
- ✓ il testo è risultato efficace e lineare, mostrando in maniera chiara e comprensibile i collegamenti fra reperti e contenuti, sebbene fosse presente qualche tecnicismo da eliminare o calibrare in funzione del target utente (Scuola Primaria II ciclo, Scuola Secondaria di I grado, Scuola Secondaria di II grado, adulti). Auspicabile l'inserimento di personaggi accompagnatori virtuali o di sezioni audio, non ancora attuato e in fase di valutazione in questa fase sperimentale;

- ✓ la tecnologia di localizzazione del reperto associato al beacon in base alla vicinanza graduale dello stesso è risultata idonea allo scopo ma certamente da affinare e perfezionare in base alla geometria dell'ambiente e al luogo in cui è posizionato il beacon, trovandosi spesso l'utente in presenza di un campo irregolare che non consentiva di ricevere le notifiche appropriate nei tempi corretti. E' noto infatti che il campo generato da un radiofaro BLE bluetooth è, fortemente condizionato da elementi riflettenti (vetrine) e materiali assorbenti (pietra, marmo e persone fisiche), presentando di conseguenza una forte anisotropia spaziale. A tal proposito è in fase di studio la creazione di una piattaforma di calibrazione dei beacon che permetta di studiare il miglior posizionamento del radiofaro e la calibrazione dei suoi parametri, analizzando gli effetti del campo in tempo reale, permettendo così di non operare manualmente ed empiricamente, come fatto durante questa sperimentazione;
- ✓ i QR-code associati ai reperti invece sono risultati idonei per il riconoscimento degli oggetti. Il lato negativo del loro uso sta nel fatto che il bambino, per velocizzare la ricerca, specialmente in una caccia a squadre, potrebbe essere indotto a cercare questi e a non focalizzare l'attenzione sugli oggetti. Quest'aspetto viene comunque meno nel caso di numeri elevati di oggetti associati a diverse aree tematiche, risultando conveniente ricercare l'oggetto visivamente qualora nel museo sia presente un numero elevato QR-code, non tutti associati ad oggetti pertinenti il tema della caccia;
- ✓ la durata della caccia, in riferimento al livello di approfondimento scelto, si è attestata intorno a circa un'ora;
- ✓ il bambino è stato positivamente sensibilizzato riguardo a ciò che è esposto presso il museo, in quanto la logica della caccia lo ha spinto non solo a esplorare tutti gli spazi museali ma ad osservare i reperti, orientato dal suggerimento per la ricerca, integrato nell'interfaccia grafica dell'app, e guidato dalle notifiche *far - near - immediate* realizzate nell'applicazione, o dalla vista dei QR-code presenti nelle varie sale espositive; occorre anche aggiungere che tale logica ha coinvolto anche tutta la sua famiglia che lo ha accompagnato nella caccia;
- ✓ la metodologia adottata, che sta alla base della sperimentazione didattica, ha ritrovato positivo riscontro in tutta la famiglia.

5. L'ESPERIMENTO CON UN GRUPPO DI RAGAZZINI

Il II esperimento ha visto il coinvolgimento di un gruppo di ragazzini di tre bambini di 10 anni e uno di 9, frequentanti rispettivamente la V e la IV classe di una Scuola Primaria, con al seguito due genitori e due bambine di 6 e 7 anni, con, in più, il ragazzino di 10 anni del test precedente, che ha ricoperto il ruolo di mentore. Anche in questo caso, i ragazzini e i genitori accompagnatori non avevano mai visitato il Museo di Camarina. Scopo dell'attività di sperimentazione è analoga a quella dell'esperimento precedente. La grafica utilizzata è stata migliorata e ottimizzata rispetto al test precedente.

5.1 LE FASI DEL II ESPERIMENTO

L'esperimento si è svolto seguendo le seguenti fasi:

- ✓ introduzione all'uso dell'app, alla sua logica, con l'ausilio del ragazzino del test precedente in ruolo di mentore;
- ✓ giro preliminare perlustrativo del museo, al fine di indicare tutte le aree (interne ed esterne) accessibili durante la caccia;
- ✓ la caccia al reperto vera e propria, partendo dai suggerimenti testuali associati ai singoli reperti;
- ✓ lettura dei contenuti associati agli oggetti e del testo della storia tematica e risposta alle varie domande di conoscenza/comprendimento intermedie e finali;
- ✓ successivo approfondimento mediante lettura pannello esplicativo presente nel museo;
- ✓ saluti finali e foto ricordo della giornata.

Tale sperimentazione prosegue con singoli gruppi di ragazzini frequentanti la Scuola Primaria "Fratelli Grimm" dell'Istituto Comprensivo Statale "Enrico Berlinguer" di Ragusa.

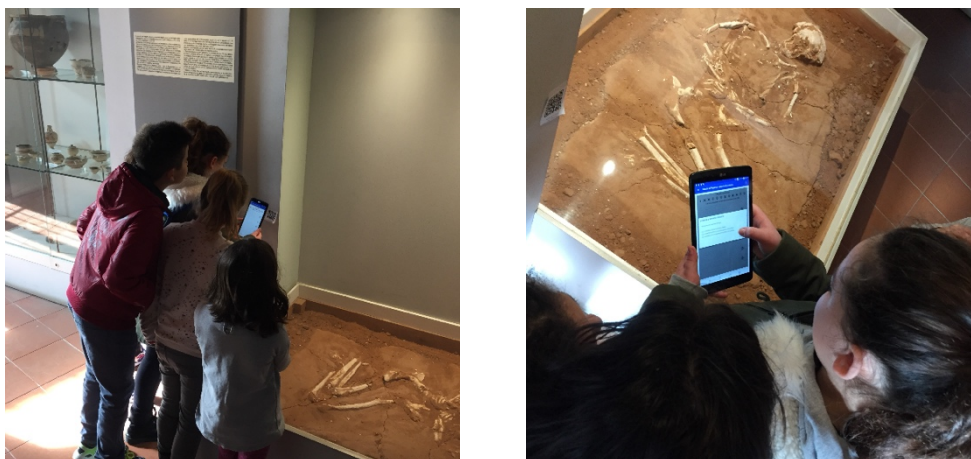


Figura 8. Alcune immagini del II esperimento

5.2 RISULTATI DEL II TEST EFFETTUATO E CONSIDERAZIONI RELATIVE

Il test ha riguardato il livello di approfondimento intermedio prima citato e ha permesso di valutare gli stessi parametri dell'esperimento precedente.

Sono stati confermati i risultati del test precedente. In particolare:

- la metodologia adottata è risultata divertente e coinvolgente per tutto il gruppo, compresi i genitori presenti, con una non indifferente ricaduta in termini di conoscenza del museo e dei reperti esposto;
- è stata confermata la necessità di eliminare i tecnicismi presenti nei testi, in particolare i termini traslitterati in Greco antico, probabilmente poco adatti per studenti di scuola primaria;
- il posizionamento dei radiofari e la scelta delle soglie sono stati ottimizzati grazie alla realizzazione di una piattaforma di calibrazione realizzata in ambiente Matlab®;
- la durata della caccia, in riferimento al livello di approfondimento scelto, si è attestata intorno a circa un'ora e mezza;
- è stato auspicato, da parte del gruppo, il collegamento con l'area archeologica esterna su cui insiste il Museo e da cui provengono parte dei reperti in esso esposti;
- in relazione alle notifiche *far – near – immediate* realizzate nell'applicazione, cui sono rispettivamente associati i colori verde – giallo – rosso, è stato suggerito l'abbinamento inverso (rosso – giallo – verde), ritenuto più naturale perché assimilabile alla logica del semaforo.

6. CONCLUSIONI

L'attività ha mostrato una ricaduta didattica ben precisa:

- ✓ aiutare attraverso un gioco a far appassionare adulti e bambini a conoscere la realtà la storia e alla cultura legata di un sito culturale (in questo caso la cultura greca camarinense);
- ✓ aiutare attraverso un gioco a far suscitare una curiosità verso tutti gli aspetti legati in questo caso alla vita di un popolo antico, gli antichi abitanti della sub-colonia greca di Kamarina.

Si sono inoltre raggiunti i seguenti obiettivi:

- ✓ passaggio dall'osservazione del manufatto alla presa di coscienza dell'uso e funzione dell'oggetto stesso, condotta a diversi livelli di approfondimento;
- ✓ la memorizzazione degli oggetti esposti;
- ✓ l'orientamento nel museo e conoscenza della mappa, tramite la geolocalizzazione;
- ✓ invogliare e motivare gli alunni a prendere confidenza con l'istituzione museo in maniera che diventi per loro un'esperienza gratificante.

7. RINGRAZIAMENTI

Gli autori desiderano ringraziare il Polo regionale di Ragusa per i siti culturali e per i Parchi archeologici di Kamarina e Cava d'Ispica per il contributo fornito alla sperimentazione didattica nonché le famiglie e i ragazzini che hanno collaborato per i test finora effettuati.

8. BIBLIOGRAFIA

- [1] Eike Falk Anderson, Leigh McLoughlin, Fotis Liarokapis, Chrisstopher Peters, Panagiotis Petridis e Sara de Freitas, 2010. *Developing serious games for cultural heritage: a state-of-the-art review*. *Virtual Reality* 14, 4 (2010), 255–275.
- [2] Nikolaos M. Avouris, Nikoleta Yiannoutsou, 2012. *Review of Mobile Location-based Games for Learning across Physical and Virtual Spaces*. *J. UCS* 18, 15 (2012), 2120–2142.
- [3] Markus Montola, Jaakko Stenros, Annika Waern, 2009. *Pervasive Games: Theory and Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [4] Carmelo Ardito, Christos Sintoris, Dimitrios Raptis, Nikoleta Yiannoutsou, Nikolaos Avouris, Maria Francesca Costabile, 2010. *Design Guidelines for Location-based Mobile Games for Learning, International Conference on Social Applications for Lifelong Learning*, Patra, Greece, 4-5 November 2010.
- [5] Nikoleta Yiannoutsou, Nikolaos Avouris, *From information consuming to participating: game-design supporting learning experiences in museums*, 2012/9. C. Karagiannidis, P. Politis & I. Karasavvidis (eds.), *Proceedings of the 8th Pan-Hellenic Conference with International Participation «ICT in Education»* University of Thessaly, pp. 1-6, Volos, Greece, 28-30 September 2012.
- [6] Tsvi Kuflik, Rokeah Oryan, Salman Suzy, 2017. *The Treasure Hunt Game Generator – a system and its application at the Hecht Museum*. MW2014: Museums and the Web 2014. Published January 15, 2014. Consulted April 13, 2017. <http://mw2014.museumsandtheweb.com/paper/the-treasure-hunt-game-generator-a-system-and-its-application-at-the-hecht-museum/>
- [7] Tony Hall, Liam Bannon, 2006. *Designing ubiquitous computing to enhance children's learning in museums*, *Journal of Computer Assisted Learning* 22, pp. 231–243.
- [8] Irene Rubino, Jetmir Xhembulla, Andrea Martina, Andrea Bottino, Giovanni Malnati, 2013. *MusA: Using Indoor Positioning and Navigation to Enhance Cultural Experiences in a museum*. In: *SENSORS*, vol. 13 n. 12, pp. 17445-17471.
- [9] Nikoleta Yiannoutsou, Ioanna Papadimitriou, Vassilis Komis, Nikolaos Avouris, 2009. *“Playing with” Museum Exhibits: Designing Educational Games Mediated by Mobile Technology*, IDC 2009, *Proceedings of the 8th International Conference on Interaction Design and Children*, pp. 230-233, June 3–5, 2009, Como, Italy.

- [10] Johan Scholliers, *EXPLORE A Gaming and Guiding System for Museum and Exhibition Environments*, 2005. VTT Technical Research Center of Finland, D18 - Final Activity Report, 10.12.2005.
- [11] Koji Yatani, Mayumi Onuma, Masanori Sugimoto, Fusako Kusunoki, 2004. *Musex: A System for Supporting Children's Collaborative Learning in a Museum with PDAs*, *Systems and Computers in Japan*, Vol. 35, No. 14, pp. 54-63, 2004, Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J86-D-I, No. 10, October 2003, pp. 773-782, 2004 Wiley Periodicals, Inc.
- [12] Soloway, Cathleen Norris, Phyllis Blumenfeld, Barry Fishman, Joseph Krajcik, Ron Marx, 2001. *Handheld devices are ready-at-hand*. *COMMUNICATIONS OF THE ACM* June 2001/Vol. 44, No. 6, pp. 15-20.
- [13] Michael Curtis, Kathleen Luchini, William Bobrowsky, Chris Quintana, Elliot Soloway. 2002. *Handheld use in K-12: A descriptive account*. Proc. IEEE International Workshop on Wireless and Mobile Technologies (WMTE '02), pp. 23-30, Växjö, Sweden.
- [14] Ugo Biader Cepeidor, Carlo M. Medaglia, Amedeo Perrone, Maria De Marsico, Giorgia di Romano, 2009. *A Museum Mobile Game for Children Using QR-Codes*, In Proceedings of the 8th International Conference on Interaction Design and Children, June 3-5, 2009, Como, Italy, pp. 282-283, ACM, 2009.
- [15] Natalia Georgiadi, Eleni Kokkoli-Papadopoulou, George Kordatos, Konstantinos Parthenadis, Manos Sparakis, Panayiotis Koutsabasis, Spyros Vosinakis, Dimitris Zisis e Modestos Stavrakis, 2016. *A Pervasive Role-Playing Game for Introducing Elementary School Students to Archaeology*, *Mobile-CH 16: Workshop on Mobile access to Cultural Heritage (in conjunction with ACM Mobile HCI 2016)*, 6-9 september 2016, Florence (Tuscany), Italy, *MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile*
- [16] Dierna G. L., Machì A. 2017. *Ludicizzazione didattica per la fruizione dei beni culturali. Percorso di "caccia al reperto" museale*. *Didamatica17: 31st Didamatica Italian Conference on eLearning - 31a Conferenza Nazionale su Didattica e Informatica*, Roma, Italia, 15-16 maggio 2017, (abstract pubblicato sulla rivista "Mondo Digitale", anno XV n. 70, giugno 2017, numero speciale "Tecnologie digitali per rinnovare la didattica l'apprendimento: la sfida per un nuovo modello di sviluppo").
- [17] Irene Rubino, Claudia Barberis, Jetmir Xhembulla, Giovanni Malnati, 2015. *Integrating a Location-Based Mobile Game in the Museum Visit: Evaluating Visitors' Behaviour and Learning*, *ACM Journal on Computing and Cultural Heritage (JOCCH)*, Vol. 8 Issue 3, No. 3, Article 15.
- [18] Michael Danks, Marc Goodchild, Karina Rodriguez-Echavarría, David B. Arnold, Richard Griffiths, 2007. *Interactive Storytelling and Gaming Environments for Museums: The Interactive Storytelling Exhibition Project*, *Edutainment'07 Proceedings of the 2nd international conference on Technologies for e-learning and digital entertainment*, pp. 104-115, K.-c. Hui et al. (Eds.): *Edutainment 2007*, LNCS 4469, pp. 104-115, 2007. Springer-Verlag Berlin Heidelberg 2007.

Labeculæ Vivæ.

Building a reference library of stains found on books

Alberto Campagnolo^{1*}, Erin Connelly^{2*}, Heather Wacha^{3*}

¹ US Library of Congress, Washington (DC), USA - acamp@loc.gov

² University of Pennsylvania, Philadelphia (PA), USA - erincon@upenn.edu

³ University of Wisconsin, Madison (WI), USA - wacha2@wisc.edu

*CLIR/DLF Postdoctoral Fellow in Data Curation in Medieval Studies

ABSTRACT

Stains on manuscripts are signs indicative of their past lives left by time and usage. Reading these signals in concert with conventional information gathered from manuscripts can add to our understanding of the history and use of an object. This project, supported by a microgrant from the Council on Library and Information Resources, and run as a preliminary pilot study, provides an identified, open-access database of a number of commonly found stains in order to help researchers answer questions such as manuscript provenance, transmission, material culture, as well as scientific applications for arts questions and the innovative uses of multi-spectral imaging to acquire new knowledge. This paper will report on the preliminary results of the investigation, our engagement with the public, broader implications, and demonstrate best practices using the database for conservators, archivists and librarians.

KEYWORDS

Stains; Spectral imaging; Manuscripts.

1. INTRODUCTION

Stains on manuscripts are signs indicative of their past lives left by time and usage. Reading these signals in concert with conventional information gathered from manuscripts can add to our understanding of the history and use of an object. There is, however, little pre-existing scholarship on the presence and significance of stains as researchers have traditionally underestimated their significance in old books. This project focuses on those very books that are often overlooked due to heavy soiling and damage, effects that diminish perceived quality and value. We are working on constructing a library of stains by gathering scientific data, drawn from stains found on parchment, paper, and bindings in medieval manuscripts and early printed books. This data will provide a new way for researchers, conservators, librarians, and the public to access information concerning the material makeup of old books, their historical uses, and new approaches for modern studies. To our knowledge, this is the first interdisciplinary attempt to build a library of book stains using the tools of medieval studies, codicology and bibliography, multi-spectral imaging, chemical analysis, and data science.¹ Our presentation will report on the preliminary results of the investigation, our engagement with the public, broader implications, and demonstrate best practices using the database for conservators, archivists and librarians.

2. BUILDING THE LIBRARY OF STAINS

This project, supported by a microgrant from the Council on Library and Information Resources (CLIR), and run as a preliminary pilot study, aims at providing an identified, open-access database of a number of commonly found stains in order to help researchers answer questions such as manuscript provenance, transmission, material culture, as well as scientific applications for arts questions and the innovative uses of multi-spectral imaging to acquire new knowledge of old objects. Identifying stains present in a book and understanding the relationship between the placement of the stain and its surrounding text will contribute to a higher intellectual and material analysis of written culture beyond current analytical approaches to text, illumination, and bibliographical description. Once the results are verified and each type of stain has been identified, others will be able to access the database and verify their own stains against the fixed dataset. The database will hold metadata collected from the multispectral imaging (JSON files), as well as information about the processed image data leading to stain identification. The latter will take the form of tiff files representing images taken at different light wavelengths, ultraviolet and fluorescent settings, in conjunction with associated specific spectral curves, and the relevant data collected from microsampling and XRF/FTIR scanning. The Schoenberg Institute for Manuscript Studies will preserve the database and all files as part of OPenn, an open access repository.

This project is intended for multiple audiences, both scholarly and public. For scholars, conservators and librarians we envision three concrete deliverables. First, our project will model and exemplify a new approach to using scientific technology to research book production and culture. Bringing together multi-spectral imaging experts and humanists offers an opportunity to explore and develop a working model for best practices when engaging in interdisciplinary collaboration. Second, the Library of Stains will be released as an online database that will provide scholars with additional tools for analyzing materiality, provenance, use and preservation; and third, it will document and disseminate a methodological

¹ The multispectral imaging dataset and physical samples developed by Giacometti *et al.* (2015) will be taken into consideration for comparison and as further reference material on staining substances on parchment. See also Giacometti *et al.* 2012; MacDonald *et al.* 2013; Campagnolo *et al.* 2016; Giacometti *et al.* 2016, 2017.

approach for those wanting to work with unknown stains so that researchers can model and replicate our work when faced with an unidentified stain. With data gathered from the multispectral images, it will be possible to create a graphic representation in the form of spectral curves of each identified stain so that when a user seeks to identify a stain in a particular manuscript, an image can be processed and compared to the graphics held in the Library of Stains database. In this way we engage the scholarly community in an on-going collaboration resulting in the continual growth of the Library and in the open access data it creates.

We envision that scholarly audiences will use our data and methodology to advance knowledge into the provenance of manuscripts, their uses within a historical context, their working environment, their transmission, and their circulation. For conservators and librarians in particular new information will help determine proper storage conditions, as well as health and safety issues, in particular the identification of heavy metal contamination, such as mercury residue in alchemical manuscripts or herbaria.² For librarians and archivists, the results of this project will also deliver a heightened awareness of the value of interdisciplinary research and model for future collaborations that can create new content and context for rare book and special collections.

Finally, this project offers a new way for the public to access manuscripts and medieval studies as an academic discipline. There is an enduring interest in medieval themes as a broad concept within the public sphere. Even if these themes are often caricatures or historically inaccurate, this interest in the medieval period in the public imagination offers the perfect opportunity to invite the public in to experience the academic discipline of medieval studies through an engaging and public-facing project. Accessibility to primary sources through an online database like the proposed library of stains juxtaposed with descriptive metadata will contextualize the project, connect with public interest, and provide value in the form of education. Our focus on public engagement is supported through the regular dissemination of information on the project to both public and scholarly communities through a variety of social media platforms, including facebook, twitter (#StainAlive), instagram, flickr, and a blog. With frequent posts across all formats, we hope to engage and excite both academic and public audiences interested in the medieval world and the lived experiences of medieval scribes, scholars, and readers.

3. ACKNOWLEDGEMENTS

The Library of Stains project is supported by a Postdoctoral Fellowship microgrant from the Council on Library and Information Resources (CLIR). The grant is made possible by funding from The Andrew W. Mellon Foundation.

4. REFERENCES

- [1] Campagnolo, Alberto, Alejandro Giacometti, Lindsay MacDonald, Simon Mahony, Stuart Robson, Tim Weyrich, Melissa Terras, and Adam Gibson. 2016. 'Cultural Heritage Destruction: Experiments with Parchment and Multispectral Imaging'. In *Digital Classics Outside the Echo-Chamber*, edited by Gabriel Bodard and Matteo Romanello, 121–46. London: Ubiquity press.
- [2] Giacometti, Alejandro, Alberto Campagnolo, Lindsay Macdonald, Simon Mahony, Stuart Robson, Tim Weyrich, Melissa Terras, and Adam Gibson. 2015. 'UCL Multispectral Processed Images of Parchment Damage Dataset', January.
- [3] Giacometti, Alejandro, Alberto Campagnolo, Lindsay MacDonald, Simon Mahony, Stuart Robson, Tim Weyrich, Melissa Terras, and Adam Gibson. 2016. 'Visualising Macroscopic Degradation of Parchment and Writing via Multispectral Images'. In *Care and Conservation of Manuscripts 15: Proceedings of the Fifteenth International Seminar Held at the University of Copenhagen, 2nd-4th April 2014*, 89–102. Copenhagen: Museum Tusulanum Press; University of Copenhagen and the Royal Library of Denmark.
- [4] Giacometti, Alejandro, Alberto Campagnolo, Lindsay MacDonald, Simon Mahony, Stuart Robson, Tim Weyrich, Melissa Terras, and Adam Gibson. 2017. 'The Value of Critical Destruction: Evaluating Multispectral Image Processing Methods for the Analysis of Primary Historical Texts'. *Digital Scholarship in the Humanities* 32 (1): 101–22. doi:10.1093/llc/fqv036.
- [5] Giacometti, Alejandro, Alberto Campagnolo, Lindsay MacDonald, Simon Mahony, Melissa Terras, Stuart Robson, Tim Weyrich, and Adam Gibson. 2012. 'Cultural Heritage Destruction: Documenting Parchment Degradation via Multispectral Imaging'. In *EVA London 2012: Electronic Visualisation and the Arts*, 301–8. London: BCS, The Chartered Institute for IT.
- [6] MacDonald, Lindsay, Alejandro Giacometti, Adam Gibson, Alberto Campagnolo, Stuart Robson, and Melissa M. Terras. 2013. 'Multispectral Imaging of Degraded Parchment'. In CCIW 2013: IAPR The Fourth Computational Color Imaging Workshop, Chiba, Japan.
- [7] Purewal, Victoria Jane. 2012. 'Novel Detection and Removal of Hazardous Biocide Residues Historically Applied to Herbaria'. Ph.D., University of Lincoln.

² Purewal (2012) has developed a UV-based methodology to identify visually the presence of mercury in herbaria.

Mapping Bucharest from a Geo-Ontological Point of View

Timothy Tambassi

ICUB, University of Bucharest, Romania - timothy.tambassi@gmail.com

ABSTRACT

The Research Institute of the University of Bucharest is financially supporting the project of research *Mapping Bucharest from a Geo-ontological Point of View*, which is aimed at developing the first specific vocabulary for mapping Bucharest from a geo-ontological point of view. The specific goal of this project is to provide a geographical inventory of classes, boundaries, entities and relation, which would describe the city of Bucharest and its historical evolution over time (accordingly, some specific questions that may arise are: may do we need a specific definition of geographical entity? How can we map the boundaries of Bucharest? How can we describe their historical evolution over time? Is the division of Bucharest in six administrative units entirely clear-cut? What is the relationship between such administrative units and the informal *cartiere* such as *Lipscani*, *Pantelimon*, *Titan* and so forth?). Moreover, part of this project will be especially focused on the possible applications of this vocabulary to other fields of research, which might comprehend, among others, public administration, historical studies, cultural heritage, and to the proposal of specific projects in digital humanities, devoted to the study of Bucharest from a geographical perspective.

KEYWORDS

Geo-Ontology, Maps, Bucharest, Boundaries, Digital Humanities, Geographical Entities, Vocabulary, Historical Evolution

1. GEO-ONTOLOGIES IN INFORMATION TECHNOLOGY AREA

Over the past few years, the innovations in the on-line cartographic visualization have created a revolution, and many new applications have dismantled traditional divisions between browsing and searching, thematic layers, web content, spatial processing and geographic datasets. Located at the intersection between geographic computing and web-based information technology, these rapid developments cannot be labelled precisely by any single body of academic literature. Hence, a variety of terms is in use for one or another aspect of this domain: from *web mapping* to *neogeography*, *social cartography*, *geoweb*, *webGIS* or *volunteered geographic information* [6, 12, 13, 27]. In this context, geographical and geospatial ontologies [4, 5, 7, 8, 9, 10, 11, 14, 15, 16, 20, 23, 24, 25, 28] are receiving a strong attention in the information technology area [1, 2, 18, 21, 22], due to four different factors:

- the growing diffusion of Geographic Information Systems (GIS);
- their use in different applications;
- the impulse of Semantic Web [3] in this area of research [17];
- the demand systematization, cataloguing and mapping of geographic information [26].

2. AN ONTOLOGY FOR THE CITY OF BUCHAREST

Given the above, the Research Institute of the University of Bucharest is financially supporting the project of research *Mapping Bucharest from a Geo-ontological Point of View*, which is aimed at developing the first specific vocabulary for mapping Bucharest from a geo-ontological point of view. The specific goal of this project is to provide a geographical inventory of classes, boundaries, entities and relation, which would describe the city of Bucharest and its historical evolution over time (accordingly, some specific questions that may arise are: may do we need a specific definition of geographical entity? How can we map the boundaries of Bucharest? How can we describe their historical evolution over time? Is the division of Bucharest in six administrative units entirely clear-cut? What is the relationship between such administrative units and the informal *cartiere* such as *Lipscani*, *Pantelimon*, *Titan* and so forth?). Moreover, part of this project will be especially focused on the possible applications of this vocabulary to other fields of research, which might comprehend, among others, public administration, historical studies, cultural heritage, and to the proposal of specific projects in digital humanities, devoted to the study of Bucharest from a geographical perspective.

3. DIVISION OF WORK

The development of this project imposes a division of work in four different steps:

1. the analysis of the maps of Bucharest, in order to identify geographical entities, classes, properties and relations, which can be easily translated in two different formal languages for ontology specification: RDF Schema (RDFS) and Web Ontology Language (OWL);
2. the study of the differences between ancient and contemporary geography, in terms of domains, presuppositions, representations and vagueness;
3. a critical review of contemporary geo-ontologies, aimed at identifying common classes and properties, and then specifying eventual missing classes and properties in order to map Bucharest. So, it is possible to establish if this vocabulary must be built *ex novo* or if more simply classes and properties can be selected and imported from other existing ontologies, emphasizing in this way the specific contribution of this inventory to the contemporary debate;

4. a reunification of this information in a geographical inventory for mapping Bucharest, based on common sense classes, properties and relations, and folk conceptualizations. This process is meant to allow the improvement of the usability of this vocabulary, making it more compatible with similar ontologies and conceptualizations¹.

The figure below summarizes the different stages of this project. Specifically, on the left side we can see the conceptual background of geo-ontologies, which might be divided into three areas: philosophy, geography and computer science. Instead, the right side shows some steps that should be completed in order to develop a specific geo-ontology for mapping Bucharest. Moreover, the picture offers the basic contents of this ontology and some possible applications.

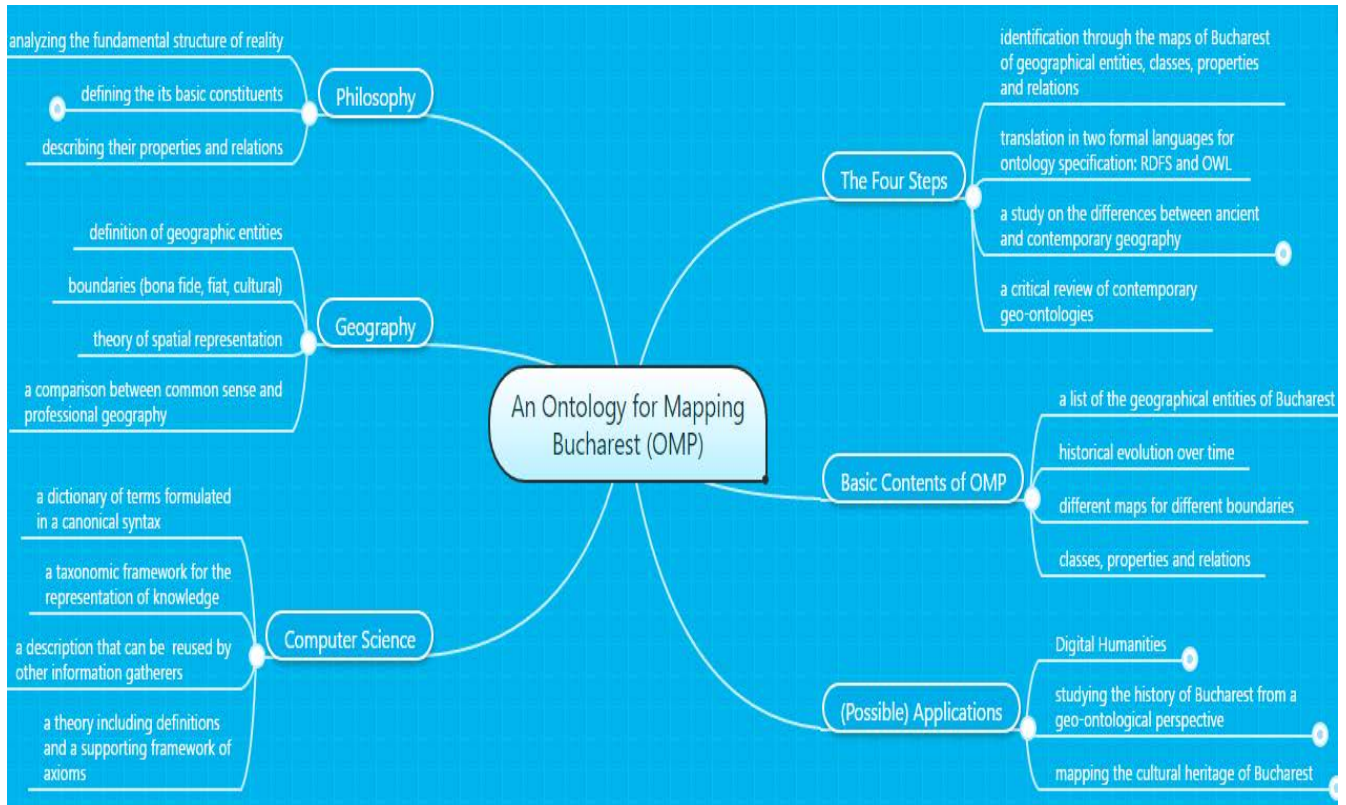


Fig. 1

4. EXPECTED RESULTS

In relation with the four steps here presented, the expected results of this project may be summarized as follows:

- outlining innovative visions about the spatial structure of Bucharest;
- better understanding its geo-historical evolution over time;
- providing possible solutions, at the conceptual level, for terminological incompatibilities among different geographical descriptions and representations of the city;
- defining a shared dictionary of terms for mapping Bucharest that can support the communication between human beings and software systems and between software systems themselves;
- developing a tool, which can be easily reused in other disciplinary contexts – such as projects concerning Digital Humanities (and specifically focused on the geography of the city), the mapping of the cultural heritage of the city and so forth.

5. BIBLIOGRAPHY

- [1] Abdelmoty, A. I., Smart, P. D., and Jones C. B. 2005. “A Critical Evaluation Of Ontology Languages For Geographic Information Retrieval On The Internet”, *Journal of Visual Languages and Computing*, vol. 16, n. 4, pp. 331-358.
- [2] Battle, R., and Kolas, D. 2012. “Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL”, *Semantic Web*, vol. 3, n. 4, pp. 355-370.
- [3] Bernes-Lee, T., Hendler, J., and Lassila, O. 2001. “The Semantic Web”, *Scientific American*, vol. May, pp 29-37.

¹ The steps follow the development of the geo-ontology *GO!* that has been specifically created for the project Geolat (<https://wiki.digitalclassicist.org/GeoLat>). For an analysis of these steps, see [19].

- [4] Bishr, Y., and Kuhn W. 2000. "Ontology-based modelling of geospatial information". In *Proceedings of the 3rd AGILE Conference on Geographic Information Science (Helsinki, 25-27 May, 2000)* edited by A. Ostman, M. Gould and T. Sarjakoski, pp. 24-27.
- [5] Bittner, T., and Winter S. 1999. "On ontology in image analysis in integrated spatial databases". In *Integrated Spatial Databases: Digital Images and GIS, Lecture Notes in Computer Science*, vol. 1737, edited by P. Agouris and A. Stefanidis. Berlin: Springer-Verlag, 168-191.
- [6] Boll, S. *et al.* (Eds.). 2008. *LOCWEB '08: Proceedings of the First International Workshop on Location and the Web*. New York: ACM.
- [7] Buccella, A., Perez, L., and Cechich, A. 2008. "GeoMergeP: Supporting an Ontological Approach to Geographic Information Integration", *International Conference of the Chilean Computer Science Society*, <http://disi.unitn.it/~p2p/RelatedWork/Matching/bucc-perbel-cech08p.pdf>.
- [8] Câmara, G., Monteiro, A., Paiva, J., and Souza, R. 2001. "Action-driven ontologies of the geographical space: Beyond the field-object debate". In *GIScience 2000—Program of the First International Conference on Geographic Information Science (Savannah, 28-31 October 2000)*, pp. 52-54.
- [9] Euzenat, J. and Shvaiko, P. 2007. *Ontology Matching*. Heidelberg: Springer-Verlag
- [10] Frank, A. 1997. "Spatial ontology". In *Spatial and Temporal Reasoning*, edited by O. Stock. Dordrecht: Kluwer Academic, pp. 135-153.
- [11] Frank, A. 2001. "Tiers of ontology and consistency constraints in geographical information systems", *International Journal of Geographical Information Science*, vol. 15, n. 7, pp. 667-678.
- [12] Goodchild, M. 2007. "Citizens as Sensors: The World of Volunteered Geography" http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf.
- [13] Hudson-Smith, A., and Crooks, A. 2008. "The Renaissance of Geographic Information: Neogeography, Gaming and Second Life", *CASA Working Paper 142*, Centre for Advanced Spatial Analysis, London, University College London, <http://www.bartlett.ucl.ac.uk/casa/pdf/paper142.pdf>.
- [14] Janowicz, K. 2006. "Sim-dl: Towards a semantic similarity measurement theory for the description logic cnr in geographic information retrieval", *OTM Workshops*, vol. 2, pp. 1681-1692.
- [15] Kavouras, M., Kokla, M., and Tomai E. 2005. "Comparing categories among geographic ontologies", *Computers and Geosciences*, vol. 31, n. 2, pp. 145-154.
- [16] Kuhn, W. 2001. "Ontologies in support of activities in geographical space", *International Journal of Geographical Information Science*, vol. 15, n. 7, pp. 613-631.
- [17] Khun, W. 2005. "Geospatial Semantics: Why, of What, and How?". In *Journal on Data Semantics*, vol. 3, edited by S. Spaccapietra and E. Zimányi, pp. 1-24.
- [18] Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S., and Koubarakis M. 2014. "GeoTriples: a Tool for Publishing Geospatial Data as RDF Graphs Using R2RML Mappings", *Proceedings of the 13th International Semantic Web Conference*, Poster & Demonstration Track, pp. 393-396.
- [19] Lana, M., Tambassi, T. forthcoming, "The geographical knowledge in classical Latin literature and the GO! ontology", *DHANT Proceedings*.
- [20] Mark, D.M. 1993. "Toward a theoretical framework for geographic entity types". In *Spatial Information Theory, Lectures Notes in Computer Science*, vol. 716, edited by A. Frank and I. Campari. Berlin: Springer-Verlag, pp. 270-283.
- [21] Perry M. and Herring J. (Eds.), *OGC GeoSPARQL – A Geographic Query Language for RDF Data*. OGC® Standard, 2012.
- [22] Ressler, J., Dean, M., and Kolas, D. 2010. "Geospatial Ontology Trade Study". In *Frontiers in Artificial Intelligence and Applications*, edited by L. Obrst, T. Janssen and W. Ceusters. Amsterdam: IOS Press.
- [23] Rodríguez, M., and Egenhofer, M. 2004. "Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure", *International Journal of Geographical Information Science*, vol. 18, n. 3, pp. 229-256.
- [24] Rodríguez, A., Egenhofer, M., and Rugg, R. 1999. "Assessing semantic similarity among geospatial feature class definitions". In *Interoperating Geographic Information Systems (Second International Conference, INTEROP '99, Zurich, 10-12 March 1999)*, *Lecture Notes in Computer Science*, vol. 1580, edited by A. Vckovski, K. Brassel and H.-J. Schek. Berlin: Springer, pp. 1-16.
- [25] Smith, B. and Mark, D.M. 1998. "Ontology and geographic kinds". In *Proceedings of the Eighth International Symposium on Spatial Data Handling (Burnaby, British Columbia, International Geographical Union, 1998)*, edited by T.K. Poiker and N. Chrisman, pp. 308-320.
- [26] Tambassi, T. 2017. *The Philosophy of Geo-Ontologies*. Springer.
- [27] Turner, A.J. 2006. *Introduction to NeoGeography*. Sebastopol (CA): O'Reilly Media Inc. http://pcmlp.socleg.ox.ac.uk/sites/pcmlp.socleg.ox.ac.uk/files/Introduction_to_Neogeography.pdf.
- [28] Visser, U. 2004. "Intelligent Information Integration for the Semantic Web. Lecture Notes", *Computer Science*, vol. 3159. Heidelberg: Springer-Verlag.

Mapping Santiago de Compostela from Gonzalo Torrente Ballester's *Fragmentos de Apocalipsis*: uncertainty and spacial transformation

Alba Rozas Arceo

Universidade de Santiago de Compostela, Spain

ABSTRACT

This paper is part of the interdisciplinary research project, "The projection of place: Compostela and the Geoliterary Imaginary (1844-1926). Geographic Information Systems and Spatial Humanities", carried out by the University of Santiago de Compostela and whose object of research is the modern emergence of the concrete literary and cultural space that constitutes this city. The aim of this work is to analyse the problems of cartographic representation using a specific literary text, *Fragmentos de Apocalipsis* by Torrente Ballester, which was published in 1977. This work has been chosen because it is the fictional testimony of a writer who composes a novel, who, at the same time, makes a diary about the process of writing this novel.

The narrative is constituted by a counterfactual elaboration or alternative space from the real city of Santiago de Compostela, which becomes Villasanta de la Estrella in the novel. The possibilities of representation of the literary space in relation to real space will be explored from the theoretical framework offered by literary cartography. GIS technologies are used to provide, through a storymap and from the point of view of the literary expert, the process of elaborating a literary cartography with the classification of spatial data that allows to visualize in a map differences between real space and literary space attending to the uncertainty problem and space transformation which is registered in the novel. Finally, advantages and limitations for this type of representation and the possibilities of analysis this work gives will be analysed from the cartographic representation point of view in several chapters of the novel.

KEYWORDS

Literary cartography, counterfactual space, Geographic information systems, spatial referentiality.

1. INTRODUCTION

The aim of this work is to analyse the perspectives and the potential use of digital cartography when representing a specific literary text.

Within literary geography studies, literary cartography is a field based on the mapping of content or concrete information of a text or a set of texts related to each other which allows to carry out a statistical, quantitative and qualitative analysis (Piatti and Hurni, 2011: 220).

Several projects, among which excels the *Ein Literarischer Atlas Europas* and other later ones such as *Mapping Emotions* in Victorian London, *Mapping St. Petersburg*, *Mapping the Lakes* or *Vilnius Literature*, are focused in providing a better understanding of fiction in general and of literary space in particular, through cartography.

These maps are used in studies about the logic of fictional spaces, because they are tools that serve the purpose of illustrating the specific geography of literature, relating to the dialectics of the fictional story, as pointed out by Piatti *et al.* (2009: 184). In any case, the mapping of spatial data extracted either from a literary work, or from a set of fictional texts, is conceived as an analysis tool that facilitates alternative ways of interpretation to extract unprecedented conclusions about the production, constitution, transformation and function of literary spaces.

2. SHORT PROPOSAL

Literature offers information about spatial elements that compose it, its features and its functions. Literary mapping is based on the link between fictional spaces and a spatial section of reality, and how so they can be represented on a map through symbols. The cartographic representation of a literary work must serve the general purpose of literary geography, that is "to analyze the interaction between these fictional spaces and their corresponding counterparts in the 'real world' (Reuschel and Hurni, 2011: 294). It is necessary, firstly, to examine separately the geographical spatial content of the proposed fictional text; and secondly, to deconstruct geography into units of place of agreement with a reference typology that determines the degree of approximation to reality.

Considering the novel selected for this research, *Fragmentos de Apocalipsis*, this approach is necessary, since Torrente Ballester creates an alternative space from the real one, Santiago de Compostela, one of the main centres of religious pilgrimage of Christianity, along with Rome and Jerusalem- The concrete fact that facilitates the transition of the story into the alternative pace of Villasanta de la Estrella is the placement of the tomb which contains the remains of a girl and not an apostle in a labyrinth in which, later on, the pulpit of the cathedral will be erected. This event of a strictly fictitious nature with no real basis constitutes the "nexus event" or the "moment in which history shifts to a parallel world" (Piatti and Hurni, 2009: 333). Then, according to the postulates of Hellekson (2001: 12) "alternate histories revolve around the basic premise that some event in the past did not occur as we know it did, and thus the present has changed". Torrente gives birth to his fictional project from a reformed enclave, showing that "settings can impressively support the historical alternative [...]with shifted or even newly drawn political borders and are set in reorganized, remodelled urban and rural spaces" (Piatti and Hurni, 2009: 333). For this reason, it is essential to analyse the concrete features of the space in

Fragmentos de Apocalipsis that support the alternative history and that justify the need of a study from a cartographic perspective.

This methodology is used to visualize the process of construction of Santiago de Compostela as a literary space with special emphasis in transformed spaces in connection with reality and uncertain spaces that do not present a concrete spatial reference in the work. For this, tools of cartographic representation, e.g. ArcGIS, a set of software products with various applications for the capture, edition, analysis, treatment, design, publication and printing of territorial information are used. Specifically, Story Maps allows you to tell the story of a place, an event, a problem or a pattern in a geographical context. It is based on the combination of interactive maps with other rich content, such as text or photos. In relation to the work studied, cartographic representation of several uncertainty and transformed places using an storymap, allows the visualization of the construction, the specific treatment and the characterization of space in the compositional process relating to the previous categories.

3. REFERENCES

- [1] Bulson, E. 2010. *Novels, maps, modernity: the spatial imagination, 1850 – 2000*, New York, London, Routledge.
- [2] Bodenhamer, D. J., Corrigan, J., Harris, M. T. (eds.). 2015. *Deep maps and spatial narratives*, Bloomington, Indiana, Indiana University Press.
- [3] Engberg-Pedersen, A. (ed.). 2017. *Literature and Cartography: Theories, Histories, Genres*, Cambridge, Massachusetts The MIT Press.
- [4] Hellekson, K. 2001. *The Alternate History: Refiguring Historical Time*, Kent State University Press, Kent, OH.
- [5] Piatti, B. and Hurni, L. 2009. "Mapping the Ontologically Unreal – Counterfactual Spaces in Literature and Cartography", *The Cartographic Journal*, 46 (4), pp. 333-342.
- [6] Piatti, B., Reuschel, A. K., Bär, H. R., Cartwright, W. y Hurni, L. (2009). 'Mapping literature towards a geography of fiction'. In W. Cartwright, G. Gartner y A. Lehn (eds.) *Cartography and Art*, Wiesbaden: Springer, pp. 177-192.
- [7] Piatti, B. and Hurni, L. 2011. "Cartographies of Fictional Worlds", *The Cartographic Journal*, 48 (4), pp. 218-223.
- [8] Reuschel, A. K. and Hurni, L. 2011. "Mapping Literature: Visualisation of Spacial Uncertainty in Fiction", *The Cartographic Journal*, 48 (4), pp. 293-308.

Measuring the “Critical Distance”. A Corpus-Based Analysis of Italian Book Reviews

Massimo Salgaro¹, Simone Rebora²

¹ University of Verona, Italy - massimo.salgaro@univr.it

² University of Verona, Italy - simone.rebora@univr.it

ABSTRACT

In this paper, we present a project that aims at testing, questioning, and eventually reshaping the most recent models of reading experience. After a brief survey of the theoretical debate, that shows the necessity of updating Eco’s proposal in light of the latest developments in neuroscience and cognitive studies, we describe an experiment carried out with the LIWC software on a corpus of Italian book reviews. The analysis was focused on the possible contrast/coherence of the cognitive, emotional, and physical aspects of reading experience. We divided the corpus into three subsets: reviews published in a social network (*aNobii*), in a paper magazine (*Il Sole 24 Ore*), and in two literary journals (*Osservatorio Critico della germanistica* and *Between*). This subdivision brought into consideration two main research questions: (1) how the medium shapes the descriptions of reading experiences and (2) if the detachment brought about by the paradigm of “critical distance” is still apt to fully grasp the phenomenon of literary reading. Results suggest a negative answer to the second question, evidencing a consistency in the increase of cognitive endeavors and emotional/physical aspects of language. However, a fine-grained analysis of the vocabulary referable to cognitive processes shows a significant difference in the three subsets: the dominant approach is more related to experiences of the self in the social network, to communicative needs in the magazine, and to analytical processes in the two literary journals. Possible developments of the project are discussed, both in terms of research methods and corpus composition, with the aim of expanding its scope to a more overarching theorization of reading experience and critical interpretation.

KEYWORDS

Reader response, Literary criticism, Neurocognitive studies, Psycholinguistics, LIWC.

1. INTRODUCTION

In light of the most recent developments in various fields of knowledge, Umberto Eco’s celebrated model of reading experience seems to necessitate a substantial updating. The “implicit reader” described in (Eco 1979) appears like an incorporeal being, whose “inferences” are driven by an unemotional process – not by chance, equated to a computing activity. However, neuroscientific research has shown how reading experience directly involves sensorimotor processes (see Gallese and Lakoff 2005; Aziz-Zadeh *et al.* 2008). Also, through the integration of psycholinguistics and cognitive studies, recent models and frameworks suggest the fundamental role of emotion in the act of reading (see Jacobs 2015; Mangen and van der Weel 2016).

With this project, we intend to show how DH tools, together with the increasing availability of reader responses in digitized format, can help testing, questioning, and even reshaping these models.

2. TOOLS, RESOURCES, AND RESEARCH QUESTIONS

With reference to the theoretical debate sketched above, an effective tool for the analysis of reader responses is the software *Linguistic Inquiry and Word Count* (LIWC)¹. The software is based on psycholinguistic research (see Tausczik and Pennebaker 2010) and it measures, among others, the presence of “Affect Words”, “Cognitive Processes”, “Perceptual Processes”, and “Biological Processes” inside the scanned texts. For our purposes, LIWC offers the possibility of ‘measuring’ the cognitive, physical, and affective dimensions of reading, in order to evaluate if they contrast or support each other.

The composition of the corpus, then, suggested at least two other research questions. Limiting the analysis to Italian language, we decided to compare book reviews written in social networks, paper magazines, and literary journals. This comparison brought into consideration the role of the medium in which reader responses are retrieved (i.e. how writing about reading experiences online, on paper, or in a selective environment changes the way we describe those experiences). Also, it implied the possible re-consideration of a dominant paradigm in literary criticism (i.e. if a detached, disembodied analysis of texts can actually lead to a full understanding of literary reading). Well conscious of the extreme delicacy of its implications, we tackled with the ever-open issue of evaluating the “critical distance”.

The construction of a representative corpus of Italian critical writing is currently under development, as it requires an extensive collection and normalization of texts published in multiple formats and with different access policies. For the sake of the present experiment, we limited our choice to a selection of HTML repositories and of open-access journals. To build a sample of “online social reading” (see Faggiolani and Vivarelli 2016), we scraped 5,000 reviews from the website *aNobii*² (all published between 2007 and 2017). As for paper magazines, we downloaded two years (2010-2011) of the

¹ <<http://liwc.wpengine.com/>> (accessed 27-09-2017).

² Reviews are the first 5,000 in the list: <<http://www.anobii.com/groups/01f13b41479ff6d9a0#collezione>> (accessed 15-09-2017).

literary section “Lettere” from the magazine *Il Sole 24 Ore*³. As for literary journals, we downloaded all the reviews published in the first twelve issues of *Osservatorio Critico della germanistica*⁴ (1998-2001) and *Between*⁵ (2011-2016). The analysis of the corpus was carried out through the LIWC Italian dictionary (see Agosti and Rellini 2007).

3. PRELIMINARY RESULTS

Results of the analysis are shown in Table 1.

Subset	Word Count	Affect Words	Perceptual Processes	Biological Processes	Cognitive Processes
aNobii (5,000 reviews)	646,964	4.74	1.2	0.97	5.85
aNobii (1,000 reviews)	149,156	4.18	1.05	0.96	5.65
Il Sole	704,242	3.29	0.83	0.95	4.15
Il Sole (2010)	353,545	3.31	0.85	0.97	4.14
Osservatorio + Between	404,692	2.42	0.54	0.41	4.58
Osservatorio	213,095	2.3	0.57	0.39	4.39
Between	191,597	2.55	0.51	0.44	4.8

Table 1. LIWC analysis of the corpus

As evident, *aNobii* dominates in almost all the considered indices. As for “Affect Words”, it totalizes 44% more than *Il Sole* and 95% more than *Osservatorio* and *Between* (pooled together). As for “Perceptual Processes”, 44% more than *Il Sole* and 122% more than the two journals. As for “Biological Processes”, there is no substantial difference compared to *Il Sole*, but still 136% more than the two journals. The most significant output, however, is about “Cognitive Processes”, where *aNobii* scores 27% more than the two journals and 40% more than *Il Sole*. These results suggest how there is no interference between cognitive and emotive/biological/perceptual processes. Indeed, the last can support the first. Also, results show how, while the “critical distance” is achieved by an inhibition of emotion and physicality (gradually increasing from magazines to journals), it may not enhance cognitive activity. In addition, Table 1 presents the results obtained by segmenting the subsets into smaller portions, to confirm that the scores do not depend on the different size of the subsets. However, a more refined analysis calls for caution in the interpretation of the results. First, *aNobii* shows higher values in the majority of the 102 indices calculated by LIWC. A phenomenon that can be explained through the “collective intelligence” (Lévi 1997) of thousands of reviewers, whose vocabulary is inevitably richer than those of dozens of cultivated critics. Second, a comparative analysis of the vocabulary in the three subsets shows how the cognitive processes individuated by LIWC are significantly differentiated: more related to experiences of the self in the social reading platform, to communicative needs in the magazine, and to analytical processes in the literary journals (see Figure 1).

4. FUTURE PERSPECTIVES

The results here obtained should be validated through a wider analysis, involving more languages and sources. Particularly delicate is the choice of journals, both for technical reasons (few are available in free, plain text/HTML format) and for the consistency of the corpus. Most reviews, in fact, are dedicated to critical studies and not to narrative works (that are dominant in social networks and magazines). We decided to keep this inconsistency in order to preserve the evaluative, engaged language of reviews, but the corpus can be expanded to critical studies, manuals, et al.

Finally, also the adopted methods may be developed further. For example, by choosing more advanced software in sentiment analysis (such as SEANCE, see Crossley *et al.* 2017), or by adopting vocabularies more specifically focused on the description of aesthetic experiences (see Knoop *et al.* 2016).

With this project, we hope to have cast the groundwork for a research that may change the way we study and perceive the experience of reading, together with the act of critical interpretation. Until now, in fact, most of the research in empirical studies and in the esthetic of reception has distinguished professional and common readers based on external information, but not on the texts they produced, while evidence confirms that there is a substantial difference in reading experiences when carried out for pleasure or for the purpose of interpretation (see Fialho *et al.* 2011).

³ <<http://archiviiodomenica.ilssole24ore.com/#>> (accessed 27-09-2017).

⁴ <<http://web.unitn.it/lettere/14961/osservatorio-critico-della-germanistica>> (accessed 27-09-2017).

⁵ <<http://ojs.unica.it/index.php/between/index>> (accessed 27-09-2017).

‘Migrazioni e colonizzazione interna nel Mediterraneo d’età moderna’: un progetto di Umanistica digitale

Giampaolo Salice

Dipartimento di Storia, Beni Culturali e Territorio, Università di Cagliari - giampaolo.salice@unica.it

ABSTRACT

In età moderna numerosi Stati europei progettano l’impianto di colonie di popolamento nei propri territori per promuovere incremento demografico, estensione delle coltivazioni, manifatture e commerci. Tra Sei e Settecento, sono soprattutto forestieri i coloni impiegati in simili intraprese. Gli studi storici hanno ricostruito alcuni episodi di riassetto territoriale, ma manca un lavoro di sintesi complessiva che legga i diversi esperimenti in una cornice globale e comparata.

Questo contributo dà conto del progetto dal titolo ‘*Migrazioni e colonizzazione interna nel Mediterraneo d’età moderna*’ e delle soluzioni digitali che, nell’ambito del medesimo progetto, sono state adottate per facilitare l’organizzazione e lo studio delle fonti relative al coinvolgimento dei forestieri nelle politiche statali di colonizzazione interna nello spazio euro-mediterraneo d’Antico Regime.

PAROLE CHIAVE

Colonizzazione interna, migrazioni, diaspora, Omeka, Neatline, Età moderna

1. LO STATO DELL’ARTE

Le strategie adottate da principati e repubbliche europee per attrarre forestieri sui propri territori sono da tempo al centro di una cospicua produzione scientifica, che ha concentrato la sua attenzione prevalentemente sulle città portuali e sui porti franchi¹; spazi destinati ad attrarre mercanti e agenti capaci di collegare i terminali portuali con le reti commerciali mediterranee ed atlantiche. La rifondazione cinquecentesca di Livorno è figlia di questa tensione generale, che fa della città toscana uno spazio di libertà economica e di relativa tolleranza verso le religioni non cattoliche (Burigana 2015): tra XVI e XVII secolo vi si stanziavano greci ed ebrei di origine italiana, tedesca, portoghese (Galasso 2002; Lehmann 2005; Marcocci 2009; Funis 2007). Ma sono i porti di tutta Europa, inclusi quelli dello Stato della Chiesa (Caffiero 2014) e Spagna (Franch Benavent 2003), che tra Sei e Settecento accolgono forestieri, esuli, migranti, dando loro la possibilità di partecipare da protagonisti alla costruzione dei grandi imperi coloniali e delle reti globali di scambio olandesi (Klooster 2009), britanniche (Panagopoulos 1956; Benady 1992) e russe (Breyfogle, Schrader, and Sunderland 2007).

Se notevole attenzione è stata dedicata dagli storici agli stanziamenti di migranti forestieri negli spazi urbani, lo stesso non si può dire per quelli promossi negli entroterra. La colonizzazione interna d’età moderna è argomento piuttosto frequentato dagli studiosi del secolo scorso e del presente. Si pensi, solo per fare qualche esempio, ai lavori sulla Sicilia (Aymard 1972; Benigno 1986; Verga 1978), sulla Sardegna (Day 1987; Ortu 1996; Ortu 2014; Salice 2010; Salice 2015a) o sulla Spagna (Barrado 2014). Tuttavia, è mancata attenzione specifica ai programmi di riassetto territoriale condotti con coloni. In anni vicini a noi, il tema è tuttavia affiorato grazie a lavori incentrati sulla mobilità umana e le migrazioni nei contesti toscano (Santus 2013), genovese (Nicholas 2005), austriaco (O’Reilly 2003; Alcoberto i Pericay 2011), sardo-piemontese (Salice 2012; Salice 2017a) e minorchino (Salice 2017b). Lo studio della colonizzazione interna della Russia, inoltre, ha aperto un’interessante riflessione sull’impatto che le politiche di riassetto del corpo territoriale dello Stato hanno avuto nella scrittura stereotipata delle identità locali/nazionali (Etkind 2013). A tutt’oggi manca comunque una lettura complessiva del fenomeno alla scala europea e mediterranea, tale da consentirci di collocare il singolo progetto di fondazione dentro un quadro trans-locale. ‘*Migrazioni e colonizzazione interna nel Mediterraneo d’età moderna*’ nasce dal proposito di colmare simile lacuna.

2. LA SCELTA DEL CSM OMEKA

L’impiego di coloni forestieri nei progetti di riassetto del corpo territoriale dello Stato è tema complesso e articolato. Per studiarlo occorre adottare un approccio comparato, una prospettiva transdisciplinare, affiancata dagli strumenti della micro-storia e della *global history*.

I livelli di indagine sono molteplici e le informazioni che scaturiscono dalla ricerca sono eterogenee e disperse. La documentazione archivistica è d’ambito sia secolare che ecclesiastico e può avere carattere istituzionale e/o privato, perché la riforma degli assetti territoriali è l’esito di una politica pubblica, che chiama però in causa e si intreccia con le biografie degli individui che vi si trovano a vario titolo coinvolti: coloni certo, ma anche funzionari di governo, intellettuali, procuratori e mediatori culturali. Della gran parte di questi individui conosciamo a mala pena il nome, ma è possibile imbattersi in figure dal significativo spessore documentale, avendo giocato un ruolo centrale nelle relazioni di potere attivate dalla colonizzazione, alla scala locale, tras-locale e talvolta globale. Anche quando i progetti di colonia interessano un’area circoscritta e coinvolgono un numero limitato di coloni, persino quando falliscono, possono innescare una produzione documentale consistente, che consente di ricostruire sia le fasi dello stanziamento, sia la discussione preparatoria tra le cancellerie e i rappresentanti dei migranti. Gli accordi tra le parti vengono formalmente esplicitati in un

¹ Si veda, ad esempio, il progetto internazionale “A Global History of Free Ports” 2016.

contratto che, generalmente, assume la forma di capitoli. È soprattutto questa tipologia di documento che si punta a digitalizzare, trascrivere e rendere accessibile. I patti di colonizzazioni indicano infatti non solo prescrivono le condizioni e gli impegni che le autorità stabiliscono in cambio della cessione del territorio, ma sono lo specchio dei valori e degli obiettivi che ispirano la politica pubblica. La gestione di documenti e informazioni così eterogenee (dalle bibliografie, ai documenti d'archivio, dalle descrizioni archivistiche, alle mappe cartografiche, alle fotografie dei siti indagati) può essere efficacemente condotta solo se si dispone di un sistema di archiviazione dei dati standardizzato e preventivamente allestito. Seguendo l'esempio di altri progetti tuttora in corso², si è deciso di dotarsi di un sistema di gestione digitale dei materiali della ricerca (Omeka), concepito per facilitare l'organizzazione della documentazione necessaria allo studio. Omeka consente di trasformare ogni documento reperito in un oggetto digitale (*item*) che può essere sia marcato con metadati internazionalmente riconosciuti (ad es. *Dublin Core*) e semplici etichette (*tag*), sia georeferenziato attraverso una mappa. Gli oggetti digitali possono essere organizzati in collezioni e in percorsi tematici, disegnati in base alle esigenze della ricerca. Ciascun *item* è inoltre costantemente implementabile alla luce delle nuove evidenze prodotte dalla ricerca. Soprattutto grazie ad Omeka, i singoli materiali, le collezioni, i percorsi tematici, i quadri locali e quelli complessivi, i contesti territoriali, le notizie biografiche, i dati urbanistici e architettonici, possono essere 'tenuti insieme' in un unico spazio digitale unitario nel quale possono essere richiamati rapidamente e integrati tra loro: cioè è possibile grazie all'integrazione tra Omeka e Neatline.

3. NEATLINE E LE 'COLONIZZAZIONI DIGITALI'

Il *plugin Neatline*, integrato in Omeka, consente di fabbricare digitalmente l'area geografica oggetto dello studio. Una volta allestito, il quadro territoriale digitale può essere 'colonizzato' con le colonie di popolamento prese in esame dalla ricerca. Ogni colonia viene a sua volta 'popolata' con la bibliografia, i documenti digitalizzati (e/o trascritti) e gli altri materiali (fotografie, audio-video, rilievi, carte storiche ecc) reperiti nel corso della ricerca e trasformati in oggetti digitali in Omeka. Collocando ciascun *item* nel territorio al quale è riferito, lo studioso ne rafforza il potere informativo: solo per fare un esempio, i patti di colonizzazione possono essere letti alla luce di dati come uso dei suoli, caratteristiche orografiche, toponomastica, confini statali ecc. Oltre che i patti di colonizzazione, le colonie digitali possono essere popolate con mappe e cartografie storiche, altra fonte di notevole rilevanza, che spesso costituisce l'unica rappresentazione degli spazi concessi ai migranti e delle originarie forme urbanistiche degli insediamenti.

La 'colonizzazione digitale' della cartografia si completa legando ad ogni colonia la bibliografia di riferimento, che viene organizzata col *plugin Zotero*, anch'esso integrato con Omeka. Zotero permette di collezionare le citazioni e di organizzarle sia per area territoriale, che filoni storiografici attinenti: dalla frontiera (Raviola 2007; Pastore 2007), al problematico concetto di 'straniero' in antico regime, passando per i recenti studi su *networks* commerciali³ e diaspora⁴.

4. PRIME CONCLUSIONI

Allo stato attuale del cantiere sono in fase di 'popolamento' gli stanziamenti di coloni manioti nella Toscana Granducale della seconda metà del Seicento, la coeva fondazione della colonia maniotta di Paomia nella Corsica genovese, le colonie progettate dalla monarchia sabauda nella Sardegna del Settecento insieme con la colonia mercantile di ortodossi stabilita nella Minorca britannica del Settecento. Dal lavoro condotto su questi progetti di colonizzazione interna con forestieri è emerso come gli obiettivi di fondo perseguiti dalle diverse cancellerie coinvolte fossero l'aumento demografico e della produttività agricola, il potenziamento dell'ordine pubblico e il presidio militare delle frontiere. In tutti i casi, la condotta spirituale dei coloni è questione di primaria importanza, disciplinata da accordi preventivi con l'autorità religiosa che ha giurisdizione sul territorio concesso. Gli strumenti adottati per dare attuazione alle politiche sono analoghe nei diversi contesti perché, come emerge dalla comparazione tra le diverse esperienze, i governi erano soliti adottare soluzioni, talvolta anche personalee coloni già utilizzati con successo dagli Stati concorrenti.

Grazie alla cartografia digitale sta emergendo come i territori interessati da colonizzazioni con forestieri sono generalmente aree 'periferiche', acquisite di recente, controllate con difficoltà dal potere centrale e contese da nemici esterni. Frontiere in senso geografico dunque, ma anche istituzionale e culturale, abitate da popolazioni che i funzionari statali consideravano 'eccentriche', quando non degradate e devianti, perché use ad un'agricoltura primitiva ed a vivere nei pressi di aree paludose e malariche, infestate da banditi e malviventi. Gli 'stranieri' che approdano in questi territori di frontiera formano un'umanità composita. Mentre in città si incontrano soprattutto mercanti e intellettuali, i governi convogliano nelle aree interne dei rispettivi territori contadini, militari e sacerdoti, ai quali affidano il compito di migliorare le produzioni agricole e presidiare il territorio. In certi casi, lo stesso movimento migratorio può approdare in Paesi diversi: in questo modo, anche i più effimeri tentativi di colonizzazione interna partecipano ad una vicenda che è al contempo locale e policentrica.

Al momento attuale, per rendere la cartografia digitale ancora più ricca di informazioni e sempre più personalizzabile si sta lavorando per integrare Neatline con la cartografia auto-prodotta col software Qgis e basata sulle carte tecniche

² Ci si riferisce ad esempio a "Avisos de Levante" 2017; "Mapping the Republic of Letters" 2017; "Archivo de La Frontera | Comunidad Virtual Para Historiadores. Banco de Recursos Históricos" 2017; "Le « Petit Thalamus » de Montpellier: Édition Critique Numérique Du Manuscrit AA9 Des Archives Municipales: Page d'accueil" 2017.

³ Senza pretese di esaustività citiamo in tema di diaspora commerciali Klooster 2009; Chaudhury; ALFANI and GOURDON 2012; Harlaftis 2005; McCabe, Harlaftis, and Minoglou 2005; Grenet and Gekas 2011; Katsiardi-Hering 2009.

⁴ Corti and Sanfilippo 2009. Sul concetto di 'diaspora' e per la bibliografia su quella greca mi permetto di rinviare all'introduzione in (Salice 2015b).

territoriali, pubblicate con licenza libera dalla gran parte delle amministrazioni di governo locale e nazionale. L'insistenza su una cartografia digitale sempre più performante nasce dalla convinzione che sia necessario attribuire il giusto rilievo alla spazialità e al ruolo che essa ha giocato nei processi di trasformazione delle società umane, incluse quelle d'Antico Regime, come suggerito dalla recente letteratura sul così detto *spatial turn* (Arias 2010).

In conclusione, la decisione di sfruttare le potenzialità di strumenti tipici dell'Umanistica Digitale, per quanto offra vantaggi indiscutibili anche in termini di disseminazione dei risultati della ricerca, è dovuta principalmente alla necessità di effettuare una raccolta ordinata, scientificamente rigorosa ed efficiente delle fonti della ricerca; di collegare informazioni estrapolate da contesti documentali differenti e di consentirne finalmente quella lettura 'a distanza' che sarebbe impossibile in ambiente analogico (Moretti 2013).

5. BIBLIOGRAFIA

- [1] "A Global History of Free Ports." 2016. *University of Helsinki*. <https://www.helsinki.fi/en/researchgroups/a-global-history-of-freeports>.
- [2] Alcobero i Pericay, Agustí. 2011. *La "Nova Barcelona" Del Danubi (1735-1738): La Ciutat Dels Exiliats de La Guerra de Successió*. Barcelona: Rafael Dalmau.
- [3] ALFANI, GUIDO, and VINCENT GOURDON. 2012. "Entrepreneurs, Formalization of Social Ties, and Trustbuilding in Europe (Fourteenth to Twentieth Centuries)." *The Economic History Review* 65 (3): 1005–1028.
- [4] "Archivo de La Frontera | Comunidad Virtual Para Historiadores. Banco de Recursos Históricos." 2017. Accessed December 27. <http://www.archivodelafrontera.com/>.
- [5] Arias, Santa. 2010. "Rethinking Space: An Outsider's View of the Spatial Turn." *GeoJournal* 75 (1): 29–41.
- [6] "Avisos de Levante." 2017. *Avisos de Levante*. Accessed December 27. <https://avisosdelevante.wordpress.com/>.
- [7] Aymard, M. 1972. "Villaggi Abbandonati e Villaggi Di Nuova Colonizzazione in Sicilia." *Problema Dei Villaggi Abbandonati Nel Medioevo*.
- [8] Barrado, José Miguel Delgado. 2014. "Fundación de Nuevas Poblaciones En Los Confines de La Monarquía Hispánica. El Caso Del Reino de Chile (1708-1796)," 135–160.
- [9] Benady, Tito. 1992. "The Role of Jews in the British Colonies of the Western Mediterranean." *Jewish Historical Studies* 33: 45–63.
- [10] Benigno, Francesco. 1986. "Vecchio e Nuovo Nella Sicilia Del Seicento: Il Ruolo Della Colonizzazione Feudale." *Studi Storici*, no. 1: 93–107.
- [11] Breyfogle, Nicholas, Abby Schrader, and Willard Sunderland, eds. 2007. *Peopling the Russian Periphery: Borderland Colonization in Eurasian History*. London, New York: Routledge.
- [12] Burigana, Riccardo. 2015. "'Troppa Tolleranza'? La Ri-Fondazione Della Città Di Livorno (1606)." *Revista de Teologia e Ciências Da Religião Da UNICAP* 5 (1): 121–138.
- [13] Caffiero, Marina. 2014. "Storia Degli Ebrei Nell'Italia Moderna." *Dal Rinascimento Alla Restaurazione*, Città Di Castello (PG), Carocci Editore Roma.
- [14] Chaudhury, Sushil. "Trading Networks in a Traditional Diaspora: Armenians in India, 1600–1800." *Diaspora Entrepreneurial Networks: Four Centuries of History*, 51–72.
- [15] Corti, Paola, and Matteo Sanfilippo, eds. 2009. *Storia d'Italia*. Annali, 24. Migrazioni. Torino: Giulio Einaudi.
- [16] Day, John. 1987. *Uomini e terre nella Sardegna coloniale: XII-XVIII secolo*. Torino: Celid.
- [17] Etkind, Alexander. 2013. *Internal Colonization: Russia's Imperial Experience*. Cambridge: Polity Press.
- [18] Franch Benavent, Ricardo. 2003. "El papel de los extranjeros en las actividades artesanales y comerciales del Mediterráneo español durante la Edad Moderna." In *Los extranjeros en la España moderna: actas del I Coloquio Internacional, celebrado en Málaga del 28 al 30 de noviembre de 2002*, edited by María Begoña Villar García and Pilar Pezzi Cristóbal. Malaga: Ministerio de Ciencia e innovación. <https://dialnet.unirioja.es/servlet/libro?codigo=3567>.
- [19] Funis, Francesca. 2007. "Gli Insediamenti Dei Greci a Livorno Tra Cinquecento e Seicento." *Città e Storia* 2 (1): 61–75.
- [20] Galasso, Cristina. 2002. *Alle origini di una comunità: ebrei ed ebrei a Livorno nel Seicento*. Storia dell'ebraismo in Italia. Firenze: L.S. Olschki. <http://digital.casalini.it/8822251555>.
- [21] Grenet, Mathieu, and Satis Gekas. 2011. "Trade, Politics and City Space(s) in Mediterranean Ports." In *Port Cities: Dynamic Landscapes and Global Networks*, edited by Caroline Hein, 89–103. London, New York: Routledge. https://www.academia.edu/930686/Trade_Politics_and_City_Space_s_in_Mediterranean_Ports_with_Athanasios_Gekas_.
- [22] Harlaftis, Gelina. 2005. "Mapping the Greek Maritime Diaspora from the Early Eighteenth to the Late Twentieth Centuries." In *Diaspora Entrepreneurial Networks: Four Centuries of History*, edited by Ina Baghdiantz McCabe, Gelina Harlaftis, and Ioanna Pepelasis Minoglou, 147–171. Oxford; New York, NY: Berg.
- [23] Katsiardi-Hering, Olga. 2009. "Christian and Jewish Ottoman Subjects: Family, Inheritance and Commercial Networks between East and West (17th-18th C.)." In *Famiglia nell'economia europea, secoli XIII-XVIII: The economic role of the family in the European economy from the 13th to the 18th centuries*. - (Atti; 20), edited by Cavaciocchi, Simonetta. Firenze: Firenze University Press. <http://digital.casalini.it/an/2306668>.
- [24] Klooster, Wim. 2009. "Network of Colonial Entrepreneurs. The Founders of the Jewish Settlements in Dutch America, 1650s and 1660s." In *Atlantic Diasporas: Jews, Conversos, and Crypto-Jews in the Age of Mercantilism, 1500–1800*, edited by Richard L. Kagan and Philip D. Morgan, 33–49. Baltimore: JHU Press.
- [25] "Le « Petit Thalamus » de Montpellier: Édition Critique Numérique Du Manuscrit AA9 Des Archives Municipales: Page d'accueil." 2017. Accessed December 27. <http://thalamus.huma-num.fr/>.
- [26] Lehmann, Matthias B. 2005. "A Livornese 'Port Jew' and the Sephardim of the Ottoman Empire." *Jewish Social Studies* 11 (2): 51–76. doi:10.1353/jss.2005.0017.
- [27] "Mapping the Republic of Letters." 2017. Accessed December 27. <http://republicofletters.stanford.edu/>.
- [28] Marcocci, Giuseppe. 2009. "Itinerari Marrani. I Portoghesi a Livorno Nei Secoli Dell'età Moderna." In *Livorno 1606–1806, Luogo Di Incontro Tra Popoli e Culture*, edited by Adriano Prosperi, 341–353. Torino: Allemandi.

- https://www.academia.edu/8016497/Itinerari_marrani_I_portoghesi_a_Livorno_nei_secoli_dell'et%C3%A0_moderna.
- [29] McCabe, Ina Baghdiantz, Gelina Harlaftis, and Ioanna Pepelasis Minoglou. 2005. *Diaspora Entrepreneurial Networks: Four Centuries of History*. Oxford; New York, NY: Berg.
- [30] Moretti, Franco. 2013. *Distant Reading*. Verso Books.
- [31] Nicholas, Nick. 2005. "A History of the Greek Colony of Corsica." *Journal of the Hellenic Diaspora* 31 (1): 33–78.
- [32] O'Reilly, William. 2003. "Divide et Impera: Race, Ethnicity and Administration in Early 18th-Century Habsburg Hungary." *Racial Discrimination and Ethnicity in European History. Plus Publishing, Pisa*, 77–100.
- [33] Ortu, Gian Giacomo. 1996. *Villaggio e poteri signorili in Sardegna: profilo storico della comunità rurale medievale e moderna*. Laterza.
- [34] Ortu, Gian Giacomo. 2014. *Ager et urbs. Trame di luogo nella Sardegna medievale e moderna*. Cagliari: CUEC Editrice.
- [35] Panagopoulos, E. P. 1956. "The Background of the Greek Settlers in the New Smyrna Colony." *The Florida Historical Quarterly* 35 (2): 95.
- [36] Pastore, Alessandro, ed. 2007. *Confini e frontiere nell'età moderna: un confronto fra discipline*. Franco Angeli.
- [37] Raviola, Blythe Alice, ed. 2007. *Lo spazio sabauda: intersezioni, frontiere e confini in età moderna*. Milano: Franco Angeli.
- [38] Salice, Giampaolo. 2010. "La Gallura nei piani di ripopolamento e nei progetti di bonifica da Rivarolo a Garibaldi." In *Garibaldi: mille volte, mille vite*, edited by Giuseppe Continiello, 91–108. Cagliari: AM&D.
- [39] Salice, Giampaolo. 2012. "L'invenzione della frontiera. Isole, Stato e colonizzazione nel Mediterraneo del Settecento." *Ammantu. Bollettino Storico, Archivistico e Consolare del Mediterraneo (ABSAC)*, no. 2: 93–113.
- [40] Salice, Giampaolo. 2015a. "Culto Dei Santi e Villaggi Di Nuova Fondazione Nella Sardegna Barocca." *Theologica & Historica*, no. XXIV: 83–106.
- [41] Salice, Giampaolo. 2015b. *Colonizzazione sabauda e diaspora greca*. Viterbo: Sette Città.
- [42] Salice, Giampaolo. 2017a. "Popolare con stranieri. Colonizzazione interna nel Settecento sabauda." *ASEI, Archivio Storico dell'Emigrazione italiana* 13: 118–125.
- [43] Salice, Giampaolo. 2017b. "Colonie Di Forestieri Nel Mediterraneo Insulare Del Settecento." In *Isole e Frontiere Nel Mediterraneo Moderno e Contemporaneo*, edited by Arturo Gallia, Lavinia Pinzarrone, and Giannantonio Scaglione, 81–96. Palermo: New Digital Frontiers.
- [44] Santus, Cesare. 2013. "Moreschi in Toscana. Progetti e Tentativi Di Insediamento Tra Livorno e La Maremma (1610-1614)." *Quaderni Storici*, no. 3/2013: 745–778. doi:10.1408/75614.
- [45] Verga, Marcello. 1978. "La 'Sicilia Dei Feudi' o 'Sicilia Dei Grani' Dalle Wüstungen Alla Colonizzazione Interna". *Società e Storia* 3: 563–579.

Patrimoni transdisciplinari.

Memoria e cultura del sensus loci in ambiente digitale

Alessia Scacchi

Università di Roma "Sapienza", Italia - alessia.scacchi@uniroma1.it

ABSTRACT

Il progetto di ricerca che viene illustrato nel presente lavoro si caratterizza per la interdisciplinarietà del patrimonio che intende sistematizzare in modo digitale e per la tipicità dei contenuti selezionati dal lavoro di scavo di geografe, letterate e storiche dell'Università "Sapienza" tra le carte d'archivio e la documentazione informatica relativa al Rione Esquilino di Roma. Ciò potrebbe rappresentare una sorta di modello per lavori che convogliano le competenze di diverse discipline su un medesimo focus, ampliando in tal modo le possibilità significative degli studi umanistici, modificando alla radice la forma stessa della ricerca in questo ambito. Infatti, il problema della confluenza di diverse discipline verso un medesimo argomento può rappresentare una palestra per collaborazioni scientifiche che intendano sviluppare media contents e piattaforme multimediali per l'analisi dei dati e la divulgazione.

PAROLE CHIAVE

Sensus loci, studi interdisciplinari, metodi per la ricerca scientifica

1. PREMESSA

La costituzione di archivi digitali è di frequente una questione che attiene alle singole discipline le quali intendono dotarsi di strumenti elettronici per conservare, valorizzare o studiare corpora tematici o comunque omogenei.

Il progetto di ricerca che viene illustrato nel presente lavoro si caratterizza, al contrario, per la interdisciplinarietà del patrimonio che intende sistematizzare in modo digitale e per la tipicità dei contenuti selezionati dal lavoro di scavo di geografe, letterate e storiche dell'Università di Roma "Sapienza".

Ciò potrebbe rappresentare una sorta di modello per lavori che convogliano le competenze di diverse discipline su un medesimo focus, ampliando in tal modo le possibilità significative degli studi umanistici, modificando alla radice la forma stessa della ricerca in questo ambito. Infatti, il problema della confluenza di diverse discipline verso un medesimo argomento può rappresentare una palestra per collaborazioni scientifiche che intendano sviluppare media contents e piattaforme multimediali per l'analisi dei dati e la divulgazione.

2. RACCOLTA E TRATTAMENTO DATI

Nella fattispecie, la ricerca intendeva studiare e organizzare i risultati sul sensus loci del Rione Esquilino di Roma, facendo confluire in un medesimo strumento differenti competenze e conoscenze disciplinari. Per questo motivo, il Rione è stato osservato da varie angolature.

Le geografe e i geografi hanno raccolto dati statistici e cartografici sulle specificità antropiche e l'organizzazione degli spazi attraverso il tempo. Le letterate hanno cercato ricorrenze lessicali del termine Esquilino e isotopie connesse ai luoghi caratteristici del quartiere, ciò analizzando un corpus testuale che va dall'Ottocento ad oggi. Infine le storiche hanno scandagliato le sedi istituzionali della conservazione archivistica – l'Archivio storico del Comune di Roma e l'Archivio Centrale dello Stato - per ricostruire la storia sociale del Rione, attraversando le fasi più emblematiche del secolo breve, scavando tra le carte dimenticate della storia "minuscola".

Tutto ciò mirava alla costruzione di uno strumento unico, è per questo che periodicamente si rivelava necessario il confronto sulle metodologie utilizzate per le indagini, sulla terminologia scientifica usata – che difficilmente risultava condivisa -, sugli orizzonti d'attesa dei singoli studiosi. Per sistematizzare la mole di dati che si andavano raccogliendo è stato predisposto un data base in cui potessero confluire differenti forme o generi di risorse, cui si potesse attingere in maniera condivisa e interdisciplinare. Parallelamente si lavorava all'ideazione di un ipermedia che mirasse alla divulgazione dei risultati della ricerca in maniera altresì interdisciplinare e multicentrica, poiché l'interesse verso il singolo argomento, il Rione Esquilino, emergesse sia in ottica disciplinare sia in una prospettiva transmediatica e trasversale.

3. VALORIZZAZIONE DI PATRIMONI TRANSDISCIPLINARI

Il risultato della ricerca sono dunque un data base su Fedora e un ipermedia istituzionale, scritto interamente in HTML5, intitolato "Esquilinonarrante". Lo strumento, ospitato dai server del CRILeT, si articola in quattro principali sezioni che al loro interno ospitano una ulteriore partizione per campi semantici extradisciplinari, rimescolando le attese delle singole discipline entro il più ampio e complesso orizzonte del sensus loci.

La Home presenta di fatto l'intero progetto con indicazioni sul gruppo di ricerca, sui presupposti e i metodi utilizzati, sulle possibilità di collaborare al progetto per il tramite di Twitter, Instagram e una mail dedicata. La pagina denominata Storia colleziona al suo interno schede su alcuni degli eventi più significativi del Novecento vissuti nel Rione come segni di cambiamenti epocali – l'urbanizzazione ottocentesca, la Resistenza e la guerra vista da Piazza Vittorio, solo per citare alcuni esempi – e delle specificità artistiche di alcuni luoghi simbolo dell'Esquilino – la Porta magica, l'Acquario romano -. La pagina intitolata Narrazione colleziona schede su testi e questioni relative al sensus loci nelle opere narrative

ottonovecentesche – il quartiere in Gadda, Pasolini, nelle scritture migranti, sui quotidiani e sui social network -. Infine la pagina dedicata al Territorio si occupa di organizzare schede sulle questioni demografiche e sulla geografia umana – il territorio dal punto di vista amministrativo, l'immigrazione, gli indicatori demografici -.

4. UN MODELLO DI DECOSTRUZIONE

L'organizzazione in schede ha consentito una maggiore interazione tra campi disciplinari differenti, stante un comune schema di scrittura che imponeva un numero di caratteri determinato e la necessaria sintesi dei contenuti in un abstract di cinquecento parole. Inoltre l'utilizzo di un data base progettato in team per convogliare contenuti molteplici, sia nella forma sia nella sostanza, rappresenta un ulteriore passo in avanti per lo sviluppo e la cultura stessa del sensus loci che, in tal senso, è stato decostruito dai componenti del gruppo di ricerca e riorganizzato in maniera a-gerarchica proprio grazie al digitale. Seppure difficile, quindi, l'interazione tra sfere disciplinari differenti è possibile ed ha rappresentato, nel progetto descritto, una effettiva prova di scambio e progresso per ambiti culturali diversi che mirino alla costruzione di nuovi patrimoni transdisciplinari¹.

5. BIBLIOGRAFIA

- [1] Anselmi, G. M., and Ruozzi, G. 2009. *Luoghi della letteratura italiana*. Milano: Bruno Mondadori
- [2] Bachelard, G. 1975. *La poetica dello spazio*. Bari: Edizioni Dedalo
- [3] Banini, T. 2013. *Introduzione. Proporre, interpretare, costruire le identità territoriali*. In *Identità territoriali. Metodi, esperienze, prospettive a confronto*. edited by Banini, T. Milano: Franco Angeli, pp. 9-27
- [4] Barrett, E. edited by. 1991. *The society of text: hypertext, hypermedia, and the social construction of information*. Cambridge (MA): MIT Press
- [5] Benigno, F. 2013. *Parole nel tempo. Un lessico per pensare la storia*. Roma: Viella
- [6] Berdoulay, V. 1991. *Parole e luoghi*. Milano: Etas Libri
- [7] Bortolotti, L. 2002. *Storia, città e territorio*. Milano: Franco Angeli
- [8] Creswell, T. 2013. *Place: an introduction*. Oxford: Blackwell
- [9] Foucault, M. 1967. *Le parole e le cose: un'archeologia delle scienze umane*. Milano: Rizzoli
- [10] Fedora Documentation Project. 2012. *Fedora Draft Documentation User Guide: Using Fedora 17 for common desktop computing tasks Edition 17.0.1*. Red Hat
- [11] Hohenberg, P. M. and Lees, L. H. 1990. *La città europea dal Medioevo a oggi*. Roma-Bari: Laterza
- [12] Hunt, L. 2010. *HTML5 Reference: The syntax, vocabulary and APIs of HTML5*. <http://www.w3.org/TR/2010/ED-html5-author-20100809>
- [13] Iacoli, G. 2008. *La percezione narrativa dello spazio: Teorie e rappresentazioni contemporanee*. Roma: Carocci
- [14] Jones, S. 2003. *Encyclopedia of New Media*. London: Sage
- [15] Loda, M. 2012. *Geografia sociale. Storia, teoria e metodi di ricerca*. Roma: Carocci
- [16] Lotman, J. M. and Uspenskij, B. A. 1975. *Tipologia della cultura*. Milano: Bompiani, 1975
- [17] Lotman, J. M. 1985. *La semiosfera*. Venezia: Marsilio
- [18] Massey, D. and Jess, P. 2001. *Luoghi, culture, globalizzazione*. Torino: Utet
- [19] Moretti, F. 1998. *Atlante del romanzo europeo 1880-1900*. Torino: Einaudi
- [20] Mudu, P. 2003. *Gli Esquilini: contributi al dibattito sulle trasformazioni nel Rione Esquilino dagli anni Settanta al Duemila*. In: Morelli, R., Sonnino, E. and Travaglini, C.M. edited by. *I territori di Roma. Storie, popolazioni, geografie*. Roma: Università degli Studi di Roma La Sapienza, Tor Vergata, Roma Tre, pp. 641-680.
- [21] Pasquinelli D'Allegra, D. 2015. *Roma: il senso del luogo*. Roma: Carocci
- [22] Piccioni, L. 2013. *Identità urbane: una riflessione*. In *Lo spazio della storia. Studi per Vittorio Vidotto*, edited by Bartolini, F., Bonomo, B. and Socrate, F. Roma-Bari: Laterza, pp. 179-194.
- [23] Scarpelli, F. 2009. *Il rione incompiuto. Antropologia urbana dell'Esquilino*. Roma: Cisu
- [24] Sorrentino, F. 2010. *Il senso dello spazio. Lo spatial turn nei metodi e nelle teorie letterarie*. Roma: Armando
- [25] Soja, E. W. 1989. *Postmodern geographies. The reassertion of space in critical social theory*. London: Verso.

¹ Banini, T. Piccioni, L. Storini, M. Narrazione, memoria, senso del luogo. Un progetto transdisciplinare per la messa in valore degli spazi urbani, in "Il capitale culturale", 4, 2016.

Pattern della globalizzazione – Trieste e il suo porto, 1859-1910

Gaetano Dato

University of Bristol, Gran Bretagna, g.dato@bristol.ac.uk

PAROLE CHIAVE

data, shipping, insurance, globalization, agent based model

La nostra epoca, dopo aver attraversato un periodo di crescente espansione della globalizzazione e del libero mercato, si trova a dibattere sulla funzione dei confini, sulla fragilità delle identità nazionali, e sul timore di un conflitto mondiale alimentato dalla competizione fra nazioni e tra le élite. Ciò riecheggia quanto avvenuto nelle decadi antecedenti il primo conflitto mondiale.

Il XIX sec., età del positivismo e della seconda rivoluzione industriale, degli imperi europei e del nazionalismo, produsse e archivò una considerevole mole di dati riguardanti numerose attività umane. Cospicui e dettagliati sono i dati sul traffico navale e le attività finanziarie.

Questa ricerca, tutt'ora in corso, analizza i traffici navali del porto di Trieste e i bilanci delle Assicurazioni Generali da metà Ottocento alle soglie della Prima guerra mondiale, e mira a identificare alcuni pattern dello sviluppo della globalizzazione, sia per poterli confrontare altri porti dello stesso periodo, sia per l'implementazione di un modello ad agenti sulla co-evoluzione del commercio e della cultura già elaborato in campo archeologico.

Nato nel corso della Complex Systems Summer School del Santa Fe Institute (NM-USA) nell'estate del 2016, il progetto si è integrato al percorso di ricerca di alcuni dei suoi proponenti¹, nella convinzione che la società umana sia un sistema dinamico e che sia possibile comprendere numerose fenomenologie storiche beneficiano dei recenti sviluppi nel campo della dinamica non lineare e della scienza della complessità [7, 8].

Il porto di Trieste è localizzato all'estremo nord del Mediterraneo e all'estremo sud della Mitteleuropa, in una regione, allora denominata Litorale Austriaco, al crocevia tra mondo latino, germanico e slavo. Trieste era l'hub commerciale dell'impero asburgico, allora il più vasto stato in Europa dopo la Russia, ed era fra i principali porti mondiali. Rappresenta quindi un valido osservatorio per comprendere alcune dinamiche avvenute all'interno del sistema mondo² lungo il percorso che portò alla Prima guerra mondiale.

Questo studio si fonda su alcuni database prodotti da i suoi stessi autori. Due sono le fonti principali: le pubblicazioni dell'ufficio statistico della Camera di Commercio di Trieste [4], conservate presso l'archivio camerale, e i bilanci delle Assicurazioni Generali (fondate a Trieste nel 1831), depositati presso l'Archivio Storico della compagnia.

L'autore di questo testo ha digitalizzato i volumi sul traffico portuale triestino e, utilizzando un software OCR (Optical Character Recognition), ha reso il testo editabile per estrapolarne i dati utili e trasferirli in file .csv. Essi contengono una lista di 58 fra luoghi di partenza e di arrivo delle rotte commerciali internazionali passanti per Trieste dal 1859 al 1909 (Fig. 1), il numero totale di navi a vela e a vapore transitanti su quelle rotte e il tonnellaggio di merci da esse movimentate. Per le fonti di Generali, la compagnia aveva già provveduto a digitalizzare e rendere editabile il testo, per cui si è reso necessario da parte dell'autore soltanto la conversione dei dati in formato .csv. I dati utilizzati dalla ricerca riguardano i ricavi, complessivi e per settore, a cominciare da quello dell'assicurazione marittima, e gli investimenti in obbligazioni e azioni.

Il team di ricerca ha analizzato la distribuzione annuale dei volumi per nave. Ne è emerso il chiaro effetto della costante crescita della capacità di trasporto e della trazione a vapore, in sintonia con i più ampi processi della seconda rivoluzione industriale; crescita però sospesa per circa un decennio dalla crisi mondiale iniziata col crollo della borsa di Vienna del 1873. Si tratta di un pattern che troviamo ripetuto in tutte le serie storiche analizzate da questa ricerca.

¹ In particolar modo Gaetano Dato (storico, University of Bristol), Simon Carrignon (informatico, Barcelona Supercomputing Center - Universitat Pompeu Fabra) e Philip Pika (biogeochimico, University of Bristol – University of Bruxelles). Gli altri membri del gruppo interdisciplinare presenti nel progetto iniziale da cui è scaturita questa ricerca sono: Tucker Ely (Arizona State University), Ben Zhu (Delft University of Technology), Anjali Tarun (Swiss Federal Institute of Technology in Lausanne), Evelyn Strombom (University of Minnesota), Rudi Minxha (Trumid Financial), Brian Ferguson (Singularity University).

² I. Wallerstein, F. Braudel, A. Gunder Frank, G. Arrighi, S. Amin, e altri studiosi svilupparono il quadro teorico del sistema mondo fra gli anni Sessanta e Settanta del Novecento [9, 3, 6, 1].



Figura 1.

Fra le varie analisi sono state inoltre prodotte delle mappe di calore (Fig. 2) rappresentanti i tonnellaggi transitanti sulle 21 principali rotte abbracciate da questo studio, aggregati per incrementi di 30.000 t. Il colore verde indica un alto volume di trasporto, il blu un basso volume, il bianco assenza di dati. Emergono alcuni aspetti, quali l'importanza del traffico verso alcune direttrici, oppure la traccia di grandi eventi, come la guerra civile americana, corrispondente all'affievolirsi di un collegamento di rilievo per lo scalo asburgico.

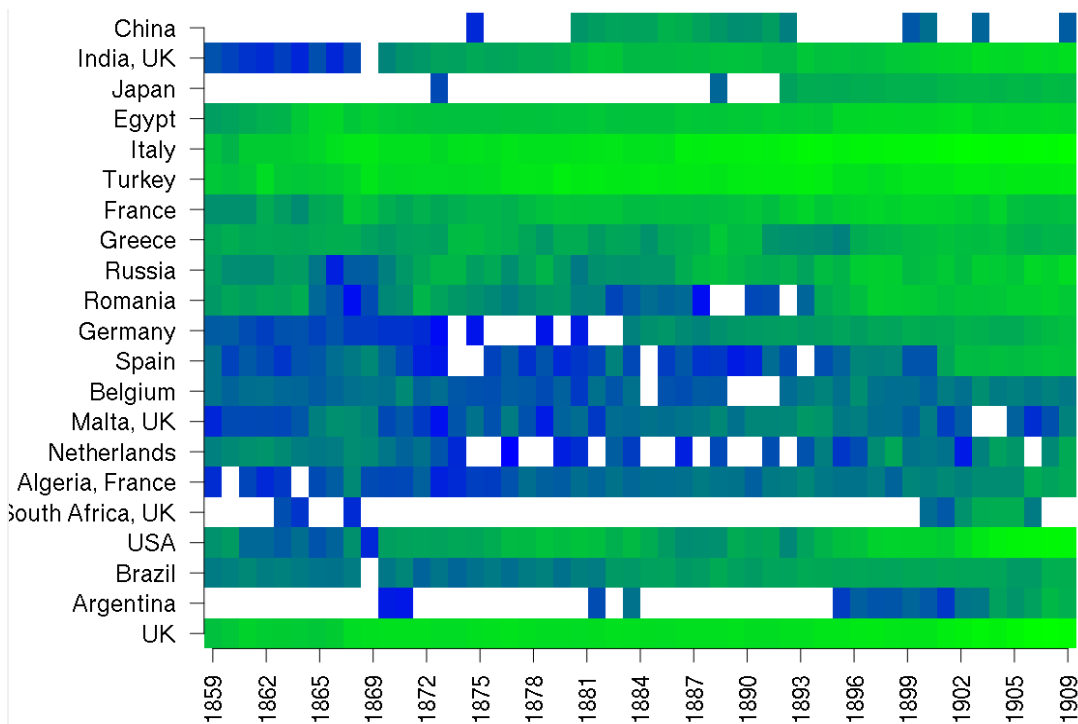


Figura 2

I dati sullo shipping sono stati inoltre incrociati con quelli di Generali (Fig. 3), non soltanto perché Generali è nata da un pool di assicuratori marittimi europei, per poi svilupparsi, come dice il nome stesso, in tutti i rami dell'assicurazione, ma anche perché l'analisi dell'attività assicurativa consente una sintesi della situazione socio-economica del mercato di riferimento [2]. Lo sviluppo dell'assicurazione sulla vita e la riduzione degli introiti da quella marittima, mostra da un lato l'ascesa di una classe media bisognosa di welfare e di nuove forme di investimento, e dall'altro la maggiore sicurezza nei trasporti per mare portata dalla tecnologia e dalle economie di scala.

Il team ha inoltre analizzato i dati utilizzando alcuni approcci derivanti dalla teoria dell'informazione.

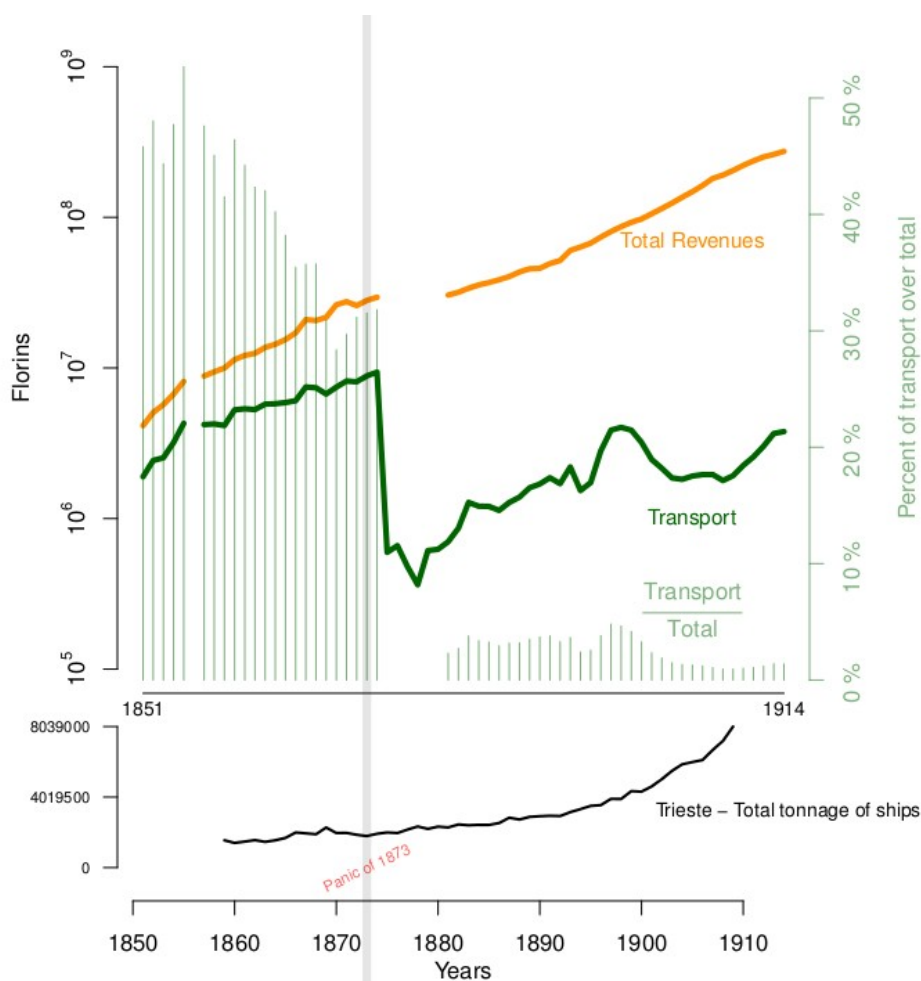


Figura 3. Il grafico superiore è in scala logaritmica e riguarda i ricavi di Generali. “Transport” si riferisce alle assicurazioni marittime.

È ora in corso il completamento di un database di circa 500 merci, transitate per terra e per mare a Trieste (1846-1914 - peso, valore, provenienza, destinazione), e l'implementazione di un modello ad agenti sulla co-evoluzione del commercio e della cultura sviluppato al Barcelona Supercomputing Center per analizzare le reti commerciali e la società dell'impero romano [5], ma efficace anche per lo studio della globalizzazione in Europa fra metà Ottocento e la Prima guerra mondiale.

BIBLIOGRAFIA

- [1] Amin, S., Arrighi, G., Frank, A. G., & Wallerstein, I. (2006). *Transforming the revolution – Social movements and the world-system*, Aakar.
- [2] Borscheid, P., Haueter, N. V., eds. (2012). *World insurance – The evolution of a Global Risk Network*, Oxford University Press.
- [3] Braudel, F. (1973). *Capitalism and material life, 1400–1800*, Harper and Row.
- [4] Camera di Commercio di Trieste (1859-1915). *Navigazione e Commercio di Trieste*, Tip. Morterra.
- [5] Carrignon, S., Montanier, J. M., Rubio-Campillo, X. (2016). “Modelling the co-evolution of trade and culture in past societies”, in *Proceedings of the 2015 Winter Simulation Conference, Huntington Beach, California - December 06-09, 2015*, IEEE Press.
- [6] Frank, A. G. (1990). A theoretical introduction to 5,000 years of world system history, *Review (Fernand Braudel Center)*, 13 (2).
- [7] Turchin, P. (2011). *Toward Clyodinamics*, *Clyodinamics*, 2(1).
- [8] Turchin, P. (2016). *Ages of Discord*, Chaplin, Beresta Books.
- [9] Wallerstein, I. (1974, 1980, 1988). *The modern world-system, (3 vols.)*, New York, Academic Press.

Per un approccio sistemico alla rilevazione, misurazione e valutazione dell'applicazione delle tecnologie digitali alla tutela, gestione e valorizzazione del patrimonio culturale

Paolo Clini¹, Pierluigi Feliciati², Ramona Quattrini³

¹ Università Politecnica delle Marche, Italia - p.clini@univpm.it

² Università di Macerata, Italia - pierluigi.feliciati@unimc.it

³ Università Politecnica delle Marche, Italia - r.quattrini@univpm.it

ABSTRACT

In vista dell'anno europeo del patrimonio culturale sembra più che mai opportuno riflettere in modo sistemico sullo stato dell'arte nell'applicazione delle tecnologie ICT al patrimonio culturale, anche in rapporto agli effetti eventuali di miglioramento della qualità nella tutela e valorizzazione. Data la molteplicità e autonomia delle linee di finanziamento a livello nazionale e locale, allo stato attuale non è possibile monitorare i progetti e la loro efficacia con un'ottica olistica. L'intervento presenterà i passi fatti finora per la costruzione di un osservatorio, di livello regionale, che si candiderebbe ad essere un riferimento condiviso per la ricerca, per la formazione e per l'applicazione delle tecnologie digitali nelle Marche. In particolare l'aspettativa è quella di una valutazione e misurazione di efficacia, evitando sovrapposizioni e dispersioni di risorse pubbliche e favorendo lo scambio di soluzioni nella valorizzazione delle *best practice*.

PAROLE CHIAVE

Digital Cultural Heritage, Osservatorio, Benessere equo e sostenibile, misurazione

1. INTRODUZIONE

In vista dell'anno europeo del patrimonio culturale sembra più che mai opportuno riflettere in modo sistemico sullo stato dell'arte nell'applicazione delle tecnologie ICT al patrimonio culturale, anche in rapporto agli effetti eventuali di miglioramento della qualità nella tutela e valorizzazione. Data la molteplicità e autonomia delle linee di finanziamento a livello nazionale e locale, allo stato attuale non è possibile monitorare i progetti e la loro efficacia con un'ottica olistica. Basta osservare il quadro dei sistemi informativi nei vari settori del *Cultural Heritage* (CH), tutti parziali come copertura informativa, praticamente non interoperabili, con aree del Paese del tutto invisibili alla rete. Il paradigma dei *Linked Open Data* (LOD) tarda ad affermarsi, anzi in alcuni casi viene visto come inutile o lesivo della concettualizzazione dei dati o della loro aggregazione/classificazione in ottica semantica. Più in generale, manca un coordinamento efficace degli approcci di misurazione delle performance del patrimonio culturale, a livello nazionale, nelle regioni e per gli istituti di ricerca e statistica. Un sistema di rilevazione siffatto dovrebbe promuovere l'applicazione di sistemi informativi per la programmazione e valutazione dei servizi relativi al patrimonio culturale in relazione alla sua conservazione, accesso, gestione, condizioni di sicurezza e di fruizione.

2. L'OSSERVATORIO SUL DIGITAL CULTURAL HERITAGE - MARCHE

Le azioni di innovazione e digitalizzazione nell'ambito dei Beni Culturali si stanno facendo sempre meno sporadiche e stanno conoscendo indubbiamente impulso, anche a valle di importanti piani di sviluppo e di ricerca industriale che hanno sottolineato l'esigenza di sviluppare tecnologie abilitanti in grado di generare nuove catene del valore¹. Nonostante questi positivi segnali, è crescente la consapevolezza di una mancanza di coordinamento e sinergia, al fine di garantire una efficacia percepita, ma soprattutto rilevata. In particolare, il design della rilevazione, a partire dalla definizione degli obiettivi di misurazione, riveste un ruolo fondamentale nella valutazione dell'applicazione delle tecnologie digitali ai beni culturali e permette di stimare l'impatto delle politiche di investimento, pubbliche e private, sulla creazione di reddito ed occupazione². Sono in questo senso da espandere e sistematizzare gli standard di misurazione e di valutazione attualmente utilizzati, inclusi quelli dell'ISTAT per il BES, il programma di valutazione sul Benessere Equo e Sostenibile³. Un osservatorio, di livello regionale, si candiderebbe ad essere riferimento condiviso per la ricerca, per la formazione e per l'applicazione delle tecnologie digitali, non escludendo la valutazione di efficacia, evitando sovrapposizioni e dispersioni di risorse pubbliche e favorendo lo scambio di soluzioni, specialmente valorizzando le *best practice*.

Lo scopo di questa presentazione è mostrare la necessità e le possibili linee di sviluppo per un approccio sistemico alla rilevazione, misurazione e valutazione dell'applicazione delle tecnologie digitali alla tutela, gestione e valorizzazione del patrimonio culturale regionale, attraverso uno sguardo davvero multifocale e improntato al controllo della qualità dei servizi. Nelle Marche, si sta costruendo un progetto di Osservatorio Regionale sul Digital Cultural Heritage, con il ruolo

¹ Cfr. Piano Strategico per la Digitalizzazione del Turismo Italiano (2014), Strategia Nazionale di Specializzazione Intelligente (2014-2020)

² Cfr. Succi et al., 2014

³ Cfr. <<http://www.istat.it/it/benessere-e-sostenibilita/C3/A0/misure-del-benessere>>. Purtroppo gli indicatori sul patrimonio culturale (dimensione 9) sono ancora da sviluppare. Un interessante intervento su questo di FAGGIOLANI 2017.

propositivo delle quattro Università marchigiane. L'Università Politecnica delle Marche e l'Università di Macerata stanno svolgendo il ruolo di coordinatori, coinvolgendo naturalmente tutti gli stakeholders: la Regione Marche, gli uffici periferici del Mibact, i comuni, le diocesi, i distretti culturali costituiti e stabilizzati⁴, le fondazioni e le aziende specializzate. Sulla scia degli interessanti risultati del giovane Osservatorio Innovazione Digitale nei Beni e Attività Culturali, nato nel 2016 presso il Politecnico di Milano, è sicuramente auspicabile generare nuovi punti di osservazione più connessi ai territori e a realtà in alcuni casi marginalizzate. In particolare la situazione dei Beni Culturali Digitali in Italia può essere sintetizzata come segue: difficoltà di accesso ai contenuti, mancanza di strategia e pianificazione, sistema di formazione e conseguenti competenze manageriali (che includano un minimo di competenza sul panorama degli strumenti digitali) carenti, rischio di banalizzazione e mancanza di sostenibilità nel breve/medio periodo⁵. L'intervento, che gli autori propongono, evidenzierà come l'Osservatorio sul Digital Cultural Heritage - Marche intenda fungere da *advisor* per le numerose e differenti realtà regionali, monitorandole e garantendo una copertura informativa e un livello di granularità ben maggiori di quanto possano fare organismi con respiro nazionale.

Gli ambiti di applicazione delle azioni dell'Osservatorio saranno:

- Ambito Tecnologico Informatico per la mappatura di strumenti digitali innovativi rivolti alla rilevazione e sistematizzazione dei dati relativi alle performance degli istituti culturali;
- Ambito Acquisizione Digitale per il censimento delle azioni di digitalizzazione, delle campagne di sistematizzazione e catalogazione di dati 3D/4D
- Ambito Economico Statistico per la definizione degli standard di misurazione dei fenomeni, l'identificazione delle metodologie di valutazione delle performance, la traduzione delle tendenze rilevate in azioni di miglioramento
- Ambito Social e gestionale per la conoscenza e gestione del processo partecipativo, comunicativo e di reporting delle attività

Le tipologie di Beni e Attività Culturali coinvolte nelle analisi saranno i Musei e le raccolte, le Biblioteche, le aree archeologiche relative al Polo Museale delle Marche, valutando poi di estendere le azioni di mappatura alle realtà comunali. Questo approccio particolarmente puntuale e esteso intercetta e fa tesoro di precedenti azioni e rapporti volti a rilevare e sistematizzare il patrimonio esistente e attivo nelle Marche ai fini della implementazione del Sistema del Museo Diffuso⁶.

3. CONCLUSIONI E SVILUPPI FUTURI

Ad oggi gran parte del lavoro svolto per la realizzazione dell'Osservatorio riguarda azioni di coordinamento dei quattro Atenei marchigiani, al fine di poter pianificare tempi e linee operative esecutive, anche in collaborazione con i Settori che si occupano di *governance* a livello regionale. Tuttavia non si è ancora in grado di definire con certezza tempi e prime azioni dell'Osservatorio: l'avvio operativo dipende infatti dagli esiti del confronto con la Regione Marche e il Segretariato regionale del MIBACT.

La sfida maggiore che l'Osservatorio si pone è di aggiungere ai dati disponibili l'effettiva qualità dei servizi offerti, composta come è noto, dalla loro efficacia ed efficienza ma anche dalla soddisfazione degli utenti, valutabile con strumenti e sulla base di indicatori diffusi in altri settori⁷. Inoltre, a valle della sintesi dei risultati di indagini osservanti e monitoraggio tramite innovativi sistemi di *Museum Behaviour Analysis*⁸ relativi a installazioni museali ci si attende lo sviluppo di azioni campione. Ad esse, potrà essere affidato il compito di indicare la strada per profilazioni/misurazioni di qualità su altre installazioni digitali e tradizionali nel contesto prescelto.

Oltre a questo si evidenzia come l'Osservatorio possa favorire un miglioramento in termini di prestazioni e sostenibilità dell'ecosistema culturale locale e lo sviluppo di quelle sinergie pubblico-private nelle economie locali in cui esso può fortemente radicarsi, anche grazie al forte ruolo propulsivo che le tecnologie ICT hanno e avranno nel settore turistico. In particolare l'innovazione è stata definita come un principio trasversale anche dal recente Piano Strategico del Turismo 2017-2022: da esso ci si attende una spinta significativa a innovare sistematicamente prodotti, processi, tecnologie e organizzazione dell'attività turistica, a innovare il mercato e le modalità di fruizione, a creare competenze nuove e più avanzate nella valorizzazione dei beni culturali.

4. RINGRAZIAMENTI

Gli autori desiderano ringraziare Rosita Pretaroli e Mauro Oronzo per la preziosa collaborazione nelle azioni di costruzione dell'Osservatorio Beni Culturali Marche.

5. BIBLIOGRAFIA

- [1] AA.VV. 2015, *Distretti culturali: esperienze a confronto. Atti del workshop*. Fermo, 16 maggio 2014. In: Il Capitale Culturale. Studies on the Value of Cultural Heritage. Supplementi (3/2015), ISSN: 2039-2362, <http://riviste.unimc.it/index.php/cap-cult/issue/view/61>
- [2] Agenzia per la coesione territoriale, 2016. *Strategia Nazionale di Specializzazione Intelligente*,

⁴ Sui Distretti culturali Cfr. DISTRETTI CULTURALI 2013.

⁵ Lorenzini, 2017.

⁶ Ci si riferisce al Rapporto di analisi della Regione Marche "VERSO LO SVILUPPO DEL SISTEMA MUSEO DIFFUSO" del 2008.

⁷ Cfr. tra gli altri FELICIATI, ALFIERI, 2017; CLINI, NESPECA, CRINELLI, 2017.

⁸ I primi risultati sono stati mostrati in CLINI, NESPECA, RUGGERI, 2017.

- http://www.agenziacoesione.gov.it/opencms/export/sites/dps/it/documentazione/S3/S3nazionale/all_2_Strategia_Nazionale_di_Specializzazione_Intelligente_Italia.pdf
- [3] CAFFO, R., 2013. *SBN tra presente e futuro*. In "Biblioteche Oggi", pp. 8-13 ISSN: 2421-3810.
- [4] CLINI, P., NESPECA, R. et RUGGERI, L., 2017. *Virtual in Real. Interactive Solutions for Learning and Communication in the National Archaeological Museum of Marche*. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XLII-5/W1, p. 647-654. DOI 10.5194/isprs-archives-XLII-5-W1-647-2017. ISSN: 2194-9034. <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-5-W1/647/2017/>.
- [5] CLINI, P., NESPECA, R., CRINELLI, G., 2017. *Il Distretto Culturale Evoluto Flaminia Nextone. Innovazione digitale nel nuovo museo virtuale della Via Flaminia a Fano*. In: Territori e frontiere della rappresentazione. 39° Convegno Internazionale dei docenti delle discipline della rappresentazione. Napoli: Gangemi editore, p. 821-828.
- [6] FAGGIOLANI, C., 2017. *"Ciò che misuriamo influenza ciò che facciamo"*. *Una nuova stagione per la valutazione in biblioteca pensando al BES*. In: LA BIBLIOTECA APERTA. TECNICHE E STRATEGIE DI CONDIVISIONE. Relazioni del Convegno Stelline. Milano, Editrice Bibliografica
- [7] FELICIATI, P., 2012. *La progettazione di sistemi informativi centrata sugli utenti: presupposti deontologici, metodologici e tecniche di misurazione* in Strumenti di ricerca per gli archivi fra editoria tradizionale, digitale e in rete; Trento, Provincia autonoma di Trento. Soprintendenza per i beni librari, archivistici e archeologici, pp. 37 – 44.
- [8] FELICIATI, P., ALFIERI, A., 2017. *Gli archivi online per gli utenti: premesse per un modello di gestione della qualità*. In: JLIS.IT; 8, 1 pp. 22 – 38. ISSN: 2038-1026
- [9] LORENZINI, E., 2017. *Il contesto: i numeri di mercato e le sfide da affrontare*. In: Osservatorio Innovazione Digitale Nei Beni E Attività Culturali, Beni e attività culturali: l'alba del Rinascimento digitale. Milano: Polimi
- [10] MIBACT, 2014. *Piano Strategico per la Digitalizzazione del Turismo Italiano*, <http://www.beniculturali.it/mibac/multimedia/MiBAC/documents/1460024515998_TD_Lab.pdf>.
- [11] MIBACT, 2017. *Piano Strategico del Turismo 2017-2022*, http://www.pst.beniculturali.it/wp-content/uploads/2017/05/PST_2017_IT_5mag17.pdf
- [12] SOCCI, C., CIASCHINI, M., PRETAROLI, R., SEVERINI, F., 2014. *Effetti macroeconomici dei sussidi alla produzione di cultura nelle Marche*. In: Economia della cultura, ISSN: 1122-7885 n. 2, pp. 213-232.
- [13] Regione Marche, Servizio Cultura, Turismo e Commercio, 2008. *Verso lo sviluppo del Sistema Museo Diffuso: primi esiti del processo di autovalutazione dei musei e delle raccolte delle Marche*, Rapporto di analisi, Osservatorio Cultura Marche.

Semiosi e archeologia del Web. Recuperare eccellenze di fine millennio

Alessia Scacchi

Università di Roma "Sapienza", Italia - alessia.scacchi@uniroma1.it

ABSTRACT

Il patrimonio culturale online è a rischio deperimento. La capacità di conservare e valorizzare le risorse ormai datate è una questione di cui si dibatte da tempo, specie in ambito umanistico. In campi di studio ove predomina l'analisi e l'interpretazione di ogni singola risorsa testuale e/o documentale è fondamentale stabilire metodi e strumenti condivisi e standardizzati, che consentano alla ricerca scientifica di recuperare ed utilizzare materiali di lavoro attraverso il tempo.

In tal senso si propone l'analisi di una sezione specifica di Internet Archive, la biblioteca senza scopo di lucro fondata nel 1996. La "Wayback Machine", infatti, rappresenta una risorsa per quanti intendano provare una sorta di archeologia informatica. Messa a punto e gestita da Mark Graham, essa è uno strumento per catturare, preservare e aiutare le persone a scoprire e utilizzare dati presenti in rete, con più di 1 miliardo di nuove acquisizioni web ogni settimana.

La Wayback Machine può consentire anche di recuperare dati ormai cancellati da server obsoleti o rovinati, come è accaduto a centri di ricerca nati a fine millennio che hanno perduto anni di lavoro di innovazione tecnologica.

Dunque si vuole restituire voce ad alcuni reperti dell'archeologia informatica come esempio per un fattivo e consapevole utilizzo della risorsa messa a punto nell'ambito dell'ambizioso progetto dell'Internet Archive. Ciò può essere utile per il recupero della memoria storica, per la valorizzazione di strumenti informatici creati nel passato e fermi per questioni di obsolescenza o di semplice incuria istituzionale, infine per ricostruire una più completa storia della cultura digitale.

PAROLE CHIAVE

Internet archive, archeologia informatica, storia della cultura digitale

1. PREMESSA

Il patrimonio culturale online è a rischio deperimento. La capacità di conservare e valorizzare le risorse ormai datate è una questione di cui si dibatte da tempo, specie in ambito umanistico. In campi di studio ove predomina l'analisi e l'interpretazione di ogni singola risorsa testuale e/o documentale è fondamentale stabilire metodi e strumenti condivisi e standardizzati, che consentano alla ricerca scientifica di recuperare ed utilizzare materiali di lavoro attraverso il tempo.

Tale necessità va sempre più prendendo corpo nelle Facoltà Umanistiche vista la mole di dati testuali che dalla fine degli anni Ottanta sono stati digitalizzati in formati diversi e con linguaggi di marcatura ormai superati. All'SGML, infatti, è seguito nel tempo l'XML che ha avuto numerose implementazioni, fino ad arrivare alla complessità del web semantico.

Gli umanisti stanno sperimentando l'obsolescenza digitale, specie in questo momento storico, visto che si trovano a dover utilizzare computer dotati di lettore CD-Rom – ormai una chimera nei pc/mac di ultima generazione – oppure mantenere in vita vecchi pc su cui sono montati programmi che non riescono ad essere letti da hardware contemporanei. Il web supplisce in minima parte alle difficoltà di lettura della LIZ, come anche della LIE – in campo letterario – poiché molte delle risorse disponibili rete non sono attendibili dal punto di vista scientifico.

2. METODO: IN DIREZIONE CONTRARIA

In tal senso si propone l'analisi di una sezione specifica di Internet Archive, la biblioteca senza scopo di lucro fondata nel 1996. Un insieme composito di risorse multimediali tra le quali sono compresi strumenti di reperimento e download di testi disponibili in diversi formati: pdf, txt, ePub e molti altri.

Tra le risorse dell'archivio testuale è importante notare uno strumento interessantissimo: la "Wayback Machine". Infatti, essa rappresenta un porto sicuro per l'umanista che voglia scandagliare i fondali del web, ovvero per quanti intendano provare una sorta di archeologia informatica. Messa a punto e gestita da Mark Graham, la Wayback Machine è uno strumento per catturare, preservare e aiutare le persone a scoprire e utilizzare dati presenti in rete fin dal 1996, con più di 1 miliardo di nuove acquisizioni web ogni settimana. L'ottica entro cui nasce e cresce lo strumento è quella della preservazione della cultura digitale conservandone traccia. Come archeologi di una storia dell'informatica umanistica italiana, ad esempio, è possibile rintracciare contenuti ormai superati e recuperare materiali che sono stati messi a disposizione in rete fin dagli anni Novanta del Novecento.

In tal senso, la documentazione riguardante i centri di ricerca che in Italia si sono occupati della commistione tra discipline umanistiche e informatica vede nella Wayback Machine un ottimo strumento di reperimento dei materiali ormai dimenticati.

3. RISULTATI

Come moderni Ulisse alla ricerca di una conoscenza che ci era preclusa, è dunque possibile ricostruire percorsi di ricerca, oltre a tracciare i progressi nelle discipline umanistiche che abbiano utilizzato l'informatica per accrescere le proprie conoscenze. La "Wayback Machine" può consentire anche di recuperare dati ormai cancellati da server obsoleti o rovinati, come è accaduto a centri di ricerca nati a fine millennio che hanno perduto mesi di lavoro e di innovazione tecnologica. In tal senso si illustrerà il caso di un centro di ricerca italiano, il CRILeT di Giuseppe Gigliozzi, i cui server sono andati distrutti all'inizio del XXI secolo, ma di cui resta una traccia nella macchina del tempo gestita da Internet Archive. Un

centro di ricerca all'avanguardia negli anni Novanta rischiava di vedere persi gli esiti di un lavoro decennale. Grazie alla Wayback Machine si può arricchire e in un certo senso anche riscrivere la storia della Scuola Romana di Informatica Umanistica.

4. CONCLUSIONI

Dunque si può restituire voce ad alcuni reperti dell'archeologia informatica come esempio per un fattivo e consapevole utilizzo della risorsa messa a punto nell'ambito dell'ambizioso progetto dell'Internet Archive. Ciò può essere utile per il recupero della memoria storica, per la valorizzazione di strumenti informatici creati nel passato e fermi per questioni di obsolescenza o di semplice incuria istituzionale, infine per ricostruire una più completa storia della cultura digitale.

5. BIBLIOGRAFIA

- [1] Andrews, T. 2015. *Digital Humanities, Living Book*, <http://www.livingbooksabouthistory.ch/fr/book/digital-humanities>
- [2] Bar-Yossef, Z., Broder, A.Z., Kumar, R., Tomkins, A. 2004. *Sic transit gloria telae: towards an understanding of the web's decay*. In: Proceedings of the 13th International Conference on World Wide Web. WWW '04, ACM, pp. 328–337
- [3] Burkhardt, M. 2009. *Is There a Way Back or Can the Internet Remember its Own History?* In A. Maj, & D. Riha (Eds.), *Digital Memories: Exploring Critical Issues*. Oxford: Inter-Disciplinary Press, pp. 129-138.
- [4] Cohen, D. J., & Rosenzweig, R. 2005. *The fragility of digital materials*. In *Digital history: a guide to gathering, preserving, and presenting the past on the web*. <http://chnm.gmu.edu/digitalhistory/preserving/1.php>
- [5] Daquino, M.; Tomasi, F. 2016. *Digital Humanities e Library and Information Science*. In: BIBLIOTHECAE.IT, 5, pp. 130-150
- [6] Derrot, S. Fauduet, L. Oury, C. Peyrard, S. 2012. *Preservation Is Knowledge: A community-driven preservation approach*. 9th International Conference on Preservation of Digital Objects (iPRES), Canada
- [7] Dougherty, M. Meyer, E. Madsen, C. Van Den Heuvel, C. Thomas, A. W Yatt, S. 2010. *Researcher Engagement with Web Archives: State of the Art*, London, JISC. http://repository.jisc.ac.uk/544/1/JISC-REWA_StateoftheArt_August2010.pdf
- [8] Dulong de Rosnay, M. Musiani, F. 2012. *The Preservation of Digital Heritage: Epistemological and Legal Reflections*. In: ESSACHESS – Journal for Communication Studies, 5 (2), pp. 81-94
- [9] Eco, U. 2007. *L'albero e il labirinto*, Milano, Bompiani
- [10] Fabbri, P. 2001. *La svolta semiotica*, Roma/Bari, Laterza
- [11] Giannetto, M. 2009-2010. *Dalla carta alle reti informative al Sistema Archivistico Nazionale*, In: Rassegna degli Archivi di Stato, N.s., V-VI, pp. 231-256
- [12] Graham, B. 2002. *Heritage as Knowledge: Capital or Culture?* In: *Urban Studies*, 33, pp. 1003-1017
- [13] Hale, Scott A., Blank, Grant, and Alexander, Victoria D. 2017. *Live versus archive: Comparing a web archive and to a population of webpages*. In Niels Brügger and Ralph Schroeder (Eds.), *The Web as History*, London: UCL Press, pp. 45-61. <https://www.ucl.ac.uk/ucl-press/browse-books/the-web-as-history>
- [14] Harrison, T.L., Nelson, M.L. 2006. *Just-In-Time Recovery of Missing Web Pages*. In: Proceedings of the 17th Conference on Hypertext and Hypermedia. HYPERTEXT '06, ACM, pp. 145–156
- [15] Hockx-Yu, H. 2012, December 19. *Digital humanities and the study of the web and web archives*. <http://britishlibrary.typepad.co.uk/webarchive/2012/12/digital-humanities-and-the-study-of-the-web.html#sthash.CSWkLtSl.dpuf>
- [16] Information Today. 2005, November 28. *Internet Archive offers new archive service*. <http://newsbreaks.infotoday.com/Digest/Internet-Archive-Offers-New-Archive-Service-16063.asp>
- [17] Internet Archive. 2014, August. *Number of unique IP addresses per day*. <https://archive.org/stats/>
- [18] Internet Archive. *About the Internet Archive*. <https://archive.org/about/>
- [19] Internet Archive. *Internet Archive projects*. <https://archive.org/projects/>
- [20] Lana, M. 2012. *Biblioteche digitali: un'introduzione*. Bologna Bononia University Press
- [21] Lana, M. 2004. *Il testo nel computer: dal web all'analisi dei testi*. Torino Bollati Boringhieri
- [22] Lucatti, E. Treleani, M. 2013. *Fare presente. Per una semiotica dell'archivio*. In: Versus. Quaderni di studi semiotici, 116, Milano, Bompiani, pp. 129-148
- [23] Lusenet, Y. 2007. *Tending the Garden or Harvesting the Fields: Digital Preservation and the UNESCO Charter on the Preservation of the Digital Heritage*. *Library Trends*, 56 (1), pp. 164-182
- [24] Mandic, S. 29/10/2011. *Internet Archive e nuove tipologie di fonti storiche*. In: *Diacronie. Studi di Storia Contemporanea*. Le sembianze di Clío: approcci alla storia. http://www.studistorici.com/2011/10/29/mandic_numero_8/
- [25] Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.
- [26] Masanès, J. 2005. *Web archiving methods and approaches: a comparative study*. In: *Library Trends*, 54(1). http://muse.jhu.edu/journals/library_trends/v054/54.1masanas.html
- [27] Masanès, J. 2006. *Web Archiving*. Berlin Heidelberg, Springer
- [28] Musiani, F. Schafer, V. 2017. *Digital Heritage and Heritagization*. In: RESET [Online], 6. URL: <http://reset.revues.org/806>
- [29] Negulescu, K.C. 2010. *Web Archiving @ the Internet Archive. Presentation at the 2010 Digital Preservation Partners Meeting*. <http://1.usa.gov/XSjDG8>
- [30] Niu, J. 2012. *An overview of web archiving*. *D-Lib Magazine*, 18(3/4). <http://dlib.org/dlib/march12/niu03niu1.html>
- [31] Pennock, M. 2013. *Web archiving*. <http://dx.doi.org/10.7207/twr13-01>
- [32] Rahaman, H. Tan, B.-K. 2011. *Interpreting Digital Heritage: A Conceptual Model With End-Users' Perspective*. In: *International Journal of Architectural Computing*, 9 (1), pp. 99-113
- [33] Rogers, R. 2013. *Digital Methods*. Cambridge, MA, The MIT Press
- [34] Rosenthal, D. S. 2010a. *How green is digital preservation?* <http://www.digitalpreservation.gov/meetings/documents/othermeetings/Rosenthal.pdf>
- [35] Rosenthal, D. S. 2010b. *How are we ensuring the longevity of digital documents?* <https://www.youtube.com/watch?v=h53DMtBUxsk>

- [36] Schafer, V. Musiani, F. Borelli, M. 2016. *Negotiating the Web of the Past*. In: French Journal for Media Research, 6, <http://frenchjournalformediaresearch.com/lodel/index.php?id=952>
- [37] Smith, L. 2006. *Uses of Heritage*. London & New York, Routledge
- [38] Smith, N. 1994. *What can Archives offer the World Wide Web?* Computing Laboratory, University of Kent, Canterbury, UK Ding, W.,
- [39] Thelwall, M., Vaughan, L. 2004. *A fair history of the web? examining country balance in the internet archive*. In: Library & Information Science Research 26(2)
- [40] Tofel, B. 2007. *Wayback for Accessing Web Archives*. In: Proceedings of International Web Archiving Workshop. IAWAW
- [41] Tomasi, F. 2017. *La preservazione del contenuto degli oggetti culturali: formalizzare la provenance*. In: BIBLIOTHECAE.IT, 6, pp. 17-40
- [42] Tomasi, F., *Edizioni o archivi digitali? Knowledge sites e apporti disciplinari*. In: Edizioni Critiche Digitali. Digital Critical Editions. Edizioni a confronto/comparing editions, Roma, Sapienza Università Editrice, 2016, pp. 129-136
- [43] Tomasi, F. Daquino, M. 2015. *Modellare ontologicamente il dominio archivistico in una prospettiva di integrazione disciplinare*. In: JLIS.IT, 6, pp. 13-38
- [44] UNESCO 2003. *Charte sur la conservation du patrimoine numérique*, http://portal.unesco.org/fr/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html
- [45] Van de Sompel, H. 2012. SiteStory transactional web archive software released. D-Lib Magazine, 18(9/10). Retrieved from <http://www.dlib.org/dlib/september12/09inbrief.html>
- [46] Zhuang, Z., Wagle, R., Giles, C. 2005. *What's there and what's not?: focused crawling for missing documents in digital libraries*. In: Proceedings of the 5th ACM/IEEE- CS Joint Conference on Digital Libraries. JCDL '05, pp. 301–310.

The electronic linguistic atlas of the Aegean island of Lesbos (EDAL)

Angela Ralli¹, Vaso Alexelli², Charalambos Tsimpouris³

¹ University of Patras, Greece - ralli@upatras.gr

² University of Patras, Greece - alexelli@upatras.gr

³ University of Patras, Greece - xtsimpouris@upatras.gr

ABSTRACT

The aim of the paper is to present the electronic linguistic atlas of the Aegean island of Lesbos (EDAL), the first and only electronic linguistic atlas in Greece, the implementation of which has been carried out within the framework of a research program funded by the *General Secretariat for the Aegean and Island Policy*. EDAL deals with the charting of the dialectal variety of Lesbos, an island with a big range of linguistic phenomena that do not only diverge from those of Standard Modern Greek (SMG), but also vary from village to village. Thus, it contributes to the preservation of a rich linguistic heritage which is threatened with extinction. EDAL is available in Greek and English, has been completed by 90%, and users have free access to it from the link <http://lesvos.lmgd.philology.upatras.gr>. It comprises 53 maps corresponding to 53 linguistic phenomena and dialectal words, which are divided into four levels of linguistic analysis, *phonology, morphology, morphosyntax, vocabulary*. The phenomena and dialectal words are depicted on 68 villages with special conventional signs/symbols, depending on the incidence. The material under investigation was drawn from both written sources and oral testimonies based on specifically built questionnaires. EDAL provides information on the number and type of informants who have responded to the questionnaires by area. It is built on Drupal 7, an open source content management system (CMS) that is actively supported by thousands of programmers, is interactive and is based on Google maps. Each entry has its own web page, which gives a description and an example of usage, so that it is accessible to both dialectal and non-dialectal users. It is also accompanied by an audio file, which shows the pronunciation of the material under examination.

KEYWORDS

Linguistic atlas, dialect of Lesbos, morphology, morphosyntax, phonology, vocabulary

1. EDAL, AN ELECTRONIC ATLAS FOR A GREEK DIALECT

Since the end of the 19th century, many countries have created *linguistic atlases* (see, among others, [4]), most of which are in printed form (the first atlas was produced for French by Gilliéron & Edmond 1902-1912), while recently, an electronic atlas has been built at the Max Planck Institute (Leipzig), the so-called WALS, containing structures and lexical material of 2676 languages of the world ([7]).

With respect to the Greek language, there is no linguistic atlas, in either printed or electronic form, with the exception of a sketchy one which exists on paper only for the Cretan dialect ([5]). In this context, the Electronic Dialectal Atlas of Lesbos (EDAL) - an island with many linguistic phenomena that do not only diverge from Standard Modern Greek (SMG) but also vary from region/village to region/village - is unique. EDAL consists of 53 maps, depicting the variation of 53 linguistic phenomena and dialectal words, which are distributed into four linguistic levels: phonology, morphology, morphosyntax, vocabulary. It excludes phenomena and words of SMG as well as those widespread in all areas of the Lesbian territory (68 villages). For example, *raising* of mid vowels /e/ and /o/ to /i/ and /u/, respectively, in unstressed position (e.g., *ci'ri*' > *ce'ri*' 'candle', *'laθus* > *'laθos* 'mistake'), and *deletion* of the high vowels /i/ and /u/ (e.g., *'çer* > *'çeri* 'hand', *vno* > *vu'no* 'mountain'), also in unstressed position, are not noted on EDAL, since they characterize all Lesbian varieties, Lesbian being part of the broader group of Northern Greek Dialects ([8], [6], [12]) which are distinct from that of the Southern ones with respect to these two phenomena. Moreover, due to lack of homogeneity, EDAL does not cover the speech of the inhabitants of Mytilene, the capital of the island, which is nowadays populated by people originating from several areas of Lesbos as well as from other parts of Greece.

EDAL is available in both Greek and English and the 53 phenomena and dialectal words were drawn from two sources: (a) the oral corpora (primary sources) of the *Laboratory of Modern Greek Dialects* (LMGD) of the University of Patras (www.lmgd.philology.upatras.gr), consisting of recordings of 60-hour narratives, the collection of which was made according to ethnographic methods ([11]); (b) written data (secondary sources) regarding the dialect of Lesbos, including dictionaries, grammars and dialectal texts (see, among others, [1], [2], [9], [13], [10], [12])

All 53 entries were tested by specifically built questionnaires, distributed in the 68 areas under investigation and responded by at least 6 informants per area. The questionnaires contained 3-4 sentences for every phenomenon or dialectal word, which native speakers were asked to recognize or reject acoustically. For instance, for the phonological phenomenon of [t] conversion into [c] before a front high vowel, interviewees of all areas listened to sentences such as: *'Kopsi tu ci'ri na 'fami ku'mat* 'Cut the cheese to eat a little', where [c] of the word *ci'ri* 'cheese' comes from [t] (SMG corresponding word *ti'ri*). All informants were over sixty years old, men and women, with low educational skills and permanent residents of the area under examination, in that they had never left their village to live in another area in or out of Lesbos.

¹ The examples are transcribed according to the characters of the *International Phonetic Alphabet* (IPA, [3]).

Each map has its own web page, which gives the description of the phenomenon or the dialectal word, an illustrative example of usage, and an audio file demonstrating the dialectal pronunciation. The same page also shows the results collected through the questionnaires, which were previously calculated in per cent (%). Percentages are shown on EDAL via the following conventional signs /symbols (Table 1) in a way that the visitor can be informed about the existence and the percent occurrence of each phenomenon by region.

Symbol	Percentage (%)
◆	0% - 24%
▲	25% - 49%
■	50% - 74%
●	75% - 100%

Table 1: Symbols reflecting the occurrence rates of the phenomena and dialectal words per region.

As can be seen in Figure 1 below, the use of the first person singular ending *-um* of the imperfect tense (e.g. *plal-um* ‘I was running’ of the dialectal verb *plalo* ‘run’) appears to exhibit high rates of occurrence (75% - 100%) mainly in southern Lesbos, while its absence or the lowest percentage (blue symbol) are recorded in regions that have accepted many refugees from the opposite coast of Asia Minor (e.g. villages such as Thermi, Nees Kydonies, Filia, etc.) in 1922, following the end of the war between Greece and Turkey, and the exchange of populations between Muslim and Christian Orthodox people impelled by the Lausanne treaty in July 1923.

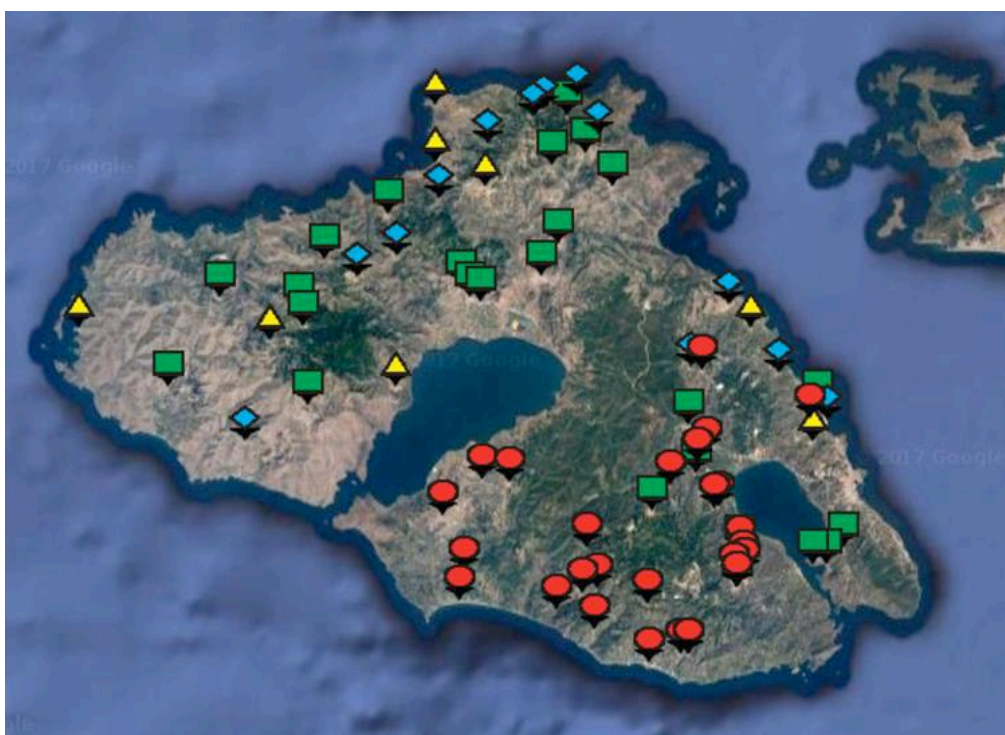


Figure 1: Percentage by region of the occurrence of the *-um* ending of the imperfect.

At the same time, EDAL enables the user to seek out information about the number of speakers who have answered the questionnaire by area. As shown in Figure 2, six people (three men and three women) were surveyed in the village of Pelopi (north-eastern Lesbos) about the use of the *-um* ending, where the percent occurrence of the phenomenon ranges from 50% to 74% (see the green square symbol).

Admittedly, EDAL displays a very narrow language coverage compared to WALS. However, it prevails in the following points: (a) it focuses on the dialectal variation of a region, while WALS provides a very limited dialectal information; (b) it draws information from both written and oral sources (recordings of narratives), not only from written ones, as is the case of WALS; (c) phenomena which have synchronically disappeared are not charted on EDAL, since the dialectal variety has been tested and confirmed by native speakers; (d) possible errors are minimized since all researchers conducting the field work are also native speakers of the dialect.

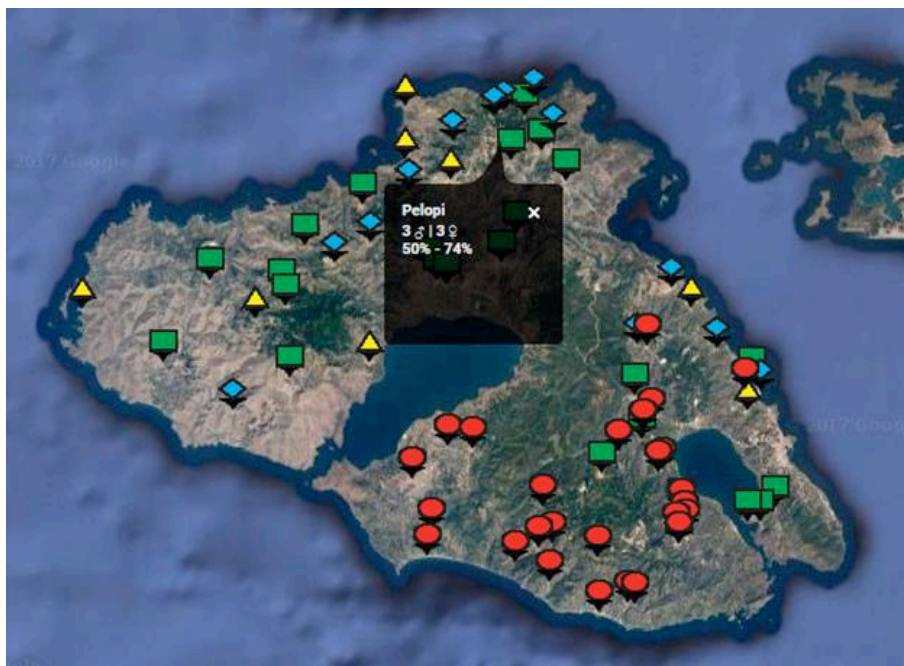


Figure 2: Information regarding the north-western village of Pelopi, its informants and the percent occurrence of the *-um* ending of the imperfect.

2. TECHNICAL DETAILS

EDAL has been built on Drupal 7, an open source content management system (CMS) that is actively supported by a live community of thousands of programmers. The Page is fully dynamic and is split in two main parts: frontend, publicly available to all users, and backend, employed by researchers to update existing data, or add new ones. Frontend includes a dynamic webpage for each entry that provides an interactive Google-based map, along with an example of usage so that it is easily comprehended by both dialectal and non-dialectal users. It is also accompanied by an audio file, which shows the pronunciation of the material under examination. Furthermore, for each entry a table is produced for an overall overview of all data regarding a specific entry, along with the number of participating informants. Backend is accessible to researchers and administrators. Researchers can log in using their own username and password and update existing data or add new entries. Data involve available regions on the map, dialectal variations categorized in groups, descriptions and accompanied examples in audio format for each entry. Moreover, administrators can add new users (researchers), if needed, and track extra information within the system's logging diary of events.

Last but not least, it should be stated that the site is SEO friendly and can be crawled by any search engine, like Google and Bing.

3. ACKNOWLEDGMENTS

EDAL has been built within the framework of the research project “Documentation and cartography of the dialectal varieties of Lesbos”. The authors wish to express their sincerest thanks to the *General Secretariat for the Aegean and Island Policy* of the *Hellenic Ministry of Mercantile Marine and Island Policy* whose generous financial support has made possible the realization of EDAL.

4. REFERENCES

- [1] Anagnostopoulou, M. 2013. *Λεσβιακοί ιδιοματισμοί* [Lesbian idiomatisms]. Mytilene: Aiolida.
- [2] Anagnostou, S. 1903. *Λεσβιακά, ήτοι συλλογή λαογραφικών περί Λέσβου πραγματειών* [Lesbian, a folklore collection about Lesbos]. Athens.
- [3] Atkins, B. T. S. and Rundell M. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- [4] Chambers, J. K. and Trudgill, P. 1998. *Dialectology*. Cambridge University Press.
- [5] Contosopoulos, N. 1975. Ο γλωσσικός άτλας της Κρήτης [The dialectal Atlas of Crete]. *Proceedings of the 3rd International Conference of Cretan Studies*. Rethymno. 145-152.
- [6] Contosopoulos, N. G. 2008. *Διάλεκτοι και Ιδιώματα της Νέας Ελληνικής* [Dialects and varieties of Modern Greek.]. Athens: Grigoris.
- [7] Dryer, M. S. & Haspelmath, M. (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- [8] Hadzidakis, G. 1905. *Μεσαιωνικά και Νέα Ελληνικά* [Medieval and Modern Greek]. Athens: Sakelariou.
- [9] Kretschmer, P. 1905. *Der Heutige Lesbische Dialekt*. Wien: Alfred Holder.
- [10] Papanis, D. and Papanis, I. 2000. *Λεξικό της Αγιασώτικης Διαλέκτου: Ερμηνευτικό-Ετυμολογικό* [Dictionary of the Agiasos Dialect: Interpretative-Etymological]. Mytilene: Community of Agiasos.

- [11] Papazachariou, D. and Archakis A. 2003. Εθνογραφικός προσδιορισμός δεδομένων από νεανικές συνομιλίες [Ethnographic identification of data from youth conversations]. *Studies on Greek Linguistics* 23: 289-300.
- [12] Ralli, A. 2017. *Λεξικό διαλεκτικής ποικιλίας Κυδωνιών, Μοσχονησίων και Βορειοανατολικής Λέσβου* [Dictionary of the dialectal variety of Kydonies, Moschonisia and North-East Lesvos]. Athens: Hellenic History Foundation.
- [13] Sakaris, G. 1940. Περί της διαλέκτου των Κυδωνιέων εν συγκρίσει προς τας Λεσβιακάς [On the dialect of Kydonies compared to Lesbian]. *Asia Minor Chronicles* 3. Athens.

The “Shape of Monuments” project: traditional methods and new technologies in the Basilica of San Lorenzo Fuori le Mura, Rome

Corrado Alvaro¹, Simone Amici², Jade Bajeot³, Valeria Danesi⁴,
Gian Michele Gerogiannis⁵, Chiara La Marca⁶, Giovanna Liberotti⁷,
Daniele Moscone⁸, Antonella Pansini⁹, Enrico Pizzoli¹⁰, Martina Zinni¹¹

¹ Department of Ancient World Studies-Sapienza University of Rome - corrado.alvaro@uniroma1.it

² Department of Ancient World Studies-Sapienza University of Rome - simone.amici@libero.it

³ PhD, Department of Ancient World Studies -Sapienza University of Rome - jdebajeot@gmail.com

⁴ PhD student, Department of Art History and Performing Arts -Sapienza University of Rome - valeria.danesi@uniroma1.it

⁵ PhD student, Department of Ancient World Studies -Sapienza University of Rome - gerogiannis.g@gmail.com

⁶ PhD, Department of Ancient World Studies -Sapienza University of Rome - chiara.lamarca@uniroma1.it

⁷ Italian Archaeological Expedition in Eastern Anatolia -Sapienza University of Rome - giovanna.liberotti@gmail.com

⁸ PhD student, Department of Ancient World Studies -Sapienza University of Rome - daniele.moscone@uniroma1.it

⁹ PhD student, Department of Ancient World Studies -Sapienza University of Rome - antonella.pansini@uniroma1.it

¹⁰ Department of Art History and Performing Arts -Sapienza University of Rome - enrico.pizzoli@uniroma1.it

¹¹ PhD student, Department of Ancient World Studies -Sapienza University of Rome - martina.zinni@uniroma1.it

ABSTRACT

In 2015-2016, the Sapienza University of Rome carried out “The Shape of Monuments” project in cooperation with Leica Geosystems to integrate new technologies with traditional methodologies and test their potential for studying ancient remains. The Department of Ancient World Studies and that of Art History and Performing Arts explored problems concerning the three-dimensional representation of monuments and archaeological artefacts by using a terrestrial laser scanner P20 and a Romer Absolute Arm to scan mobile artefacts as lithics, pottery and epigraphs, as well as entire buildings. Here we report the study case of the Basilica of San Lorenzo Fuori le Mura in Rome, an example of a particularly complex religious architecture with a remarkable construction history. The complexity is due to the juxtaposition and the interaction of various architectural structures like the catacombs, the cloister and the church itself with its irregular planovolumetric development. The main problem we had to face concerned the survey of the catacombs, which consist of several underground chambers and corridors physically difficult to reach out.

We reviewed the literature on the most recent use of laser scanning survey to record and document large architectural complexes and we decided to investigate different aspects of the buildings through a multidisciplinary approach. We considered the different phases of the church from the late antiquity to the modern period, we focused on the phenomenon of architectural reuse, we analysed capitals, columns and trabeations from the stylistic point of view and we discovered inscriptions that were unknown until now since they were located at inaccessible height.

The use of new technologies complemented with traditional methods typical of the historical-artistic and archaeological research enabled us to analyse the church from new points of view, to get brand new data and to open new research perspectives on the San Lorenzo area.

KEYWORDS

Laser scanning survey, multidisciplinary approach, church.

1. INTRODUCTION

This paper aims at presenting part of the results of “the Shape of Monuments” project, which was carried out by the Sapienza University of Rome in cooperation with Leica Geosystems. Between 2015 and 2016, researchers and PhD students from different research milieu such as archaeology (prehistory, classical studies, Near East and Etruscology) and art history (medieval) have worked together to explore problems concerning the three-dimensional representation of monuments and archaeological artefacts. They used a terrestrial laser scanner P20 and a Romer Absolute Arm, with their related software, to scan mobile artefacts (lithics, pottery, metal objects, statues and epigraphs), as well as entire buildings (churches and earthen structures). The aim of the project was to investigate, through a multidisciplinary approach, the use of new technologies complemented with the traditional methods and test their potential for studying ancient remains from different perspectives.

In this paper, the study case of the Basilica of San Lorenzo Fuori le Mura in Rome (Italy) (Fig. 1), an example of particularly complex religious architecture with a remarkable construction history, is presented. The complexity is due to the juxtaposition and interaction among various architectural structures such as the catacomb complex, the cloister and the church itself with its irregular planovolumetric development (Fig. 2).



Figure 1: 3D model of the complex with the workstations

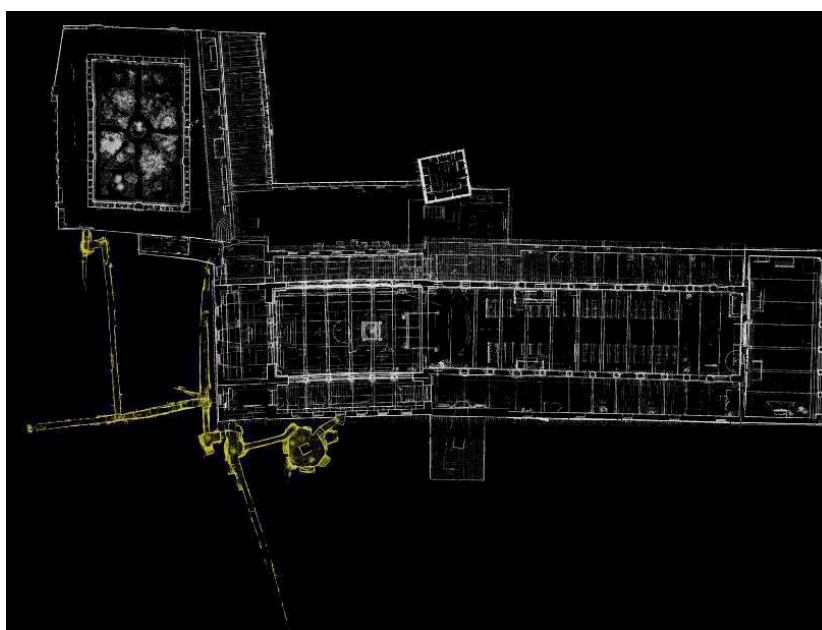


Figure 2: Plan of the complex

2. HISTORY OF THE COMPLEX AND STATE OF THE ART

The area of the actual Basilica was used as a cemetery since the first century BCE. Emperor Constantine built the first Basilica in the first half of the 4th century near the tomb of Saint Lawrence. Close by the Basilica, Pope Pelagius II (579-590) built a second three-nave church with a west-oriented apse eroding the catacomb to place the new church above the tomb of the Saint. This arrangement enabled a direct access to the catacombs from the church. After the 7th century, the structure was gradually abandoned, until Pope Leo IV (847-855) built two different monasteries nearby the Basilica. Later on, the Cluniac monks reorganised the abbey. They erected a cloister and included the tomb of the Saint within the Basilica through the construction of a new apse, which Pope Honorius III (1216-1227) destroyed in the occasion of a radical renovation of the building. Eventually, between 1855 and 1864 architect Virginio Vespignani restored the complex on behalf of Pope Pius IX (1846-1878). In 1943 the Basilica was bombed during the Second World War.

In 1933, Da Bra published initial studies about the Basilica, resumed by Munoz in 1944 and by Krautheimer in 1962. Since then, various scholars discussed on the analysis of the complex in a sectoral way. In 2014, Fabio Remondino and Stefano Campana Among published the results of a laser scanner survey carried out only on the church. The complexity of this centuries-old building made it particularly suited to test this technology and the collaboration of researchers with such a different background. It was thus possible to use the laser scanning survey to deepen multiple aspects of the history of the monument.

3. METHODOLOGIES AND RESULTS OBTAINED

The study of a complex building like the Basilica of San Lorenzo Fuori le Mura placed in the first place the question of what is the most appropriate methodological approach to meet diverse and multidisciplinary needs. We identified it in the laser scanner survey, which would have provided us with the greatest amount of data, both metric and stylistic. Given that each of us aimed at achieving specific goals according to the expertise of each (archaeology, topography, epigraphy, history of art), we believe that the potential of laser scanning survey is a tool to meet our research needs rather than the objective of our research.

In particular, we have scanned the church (inside and outside), the cloister and a portion of the catacomb through a P20 laser scanner with a built-in high-resolution digital camera. We performed a total number of 57 scans with a scan density of 6.3 mm at 10 m (medium quality level). The resulting product of each scan was a point cloud where each point yields the spectral colour information for the represented material. We downloaded and processed all the data through Cyclone software provided by Leica Geosystems. We produced plans, profiles and the basic elements needed to analyse the building. After the three-dimensional reconstruction was done, we virtually separated the Pelagian building from the Honorian one (Fig. 3) according to the previous literature review.

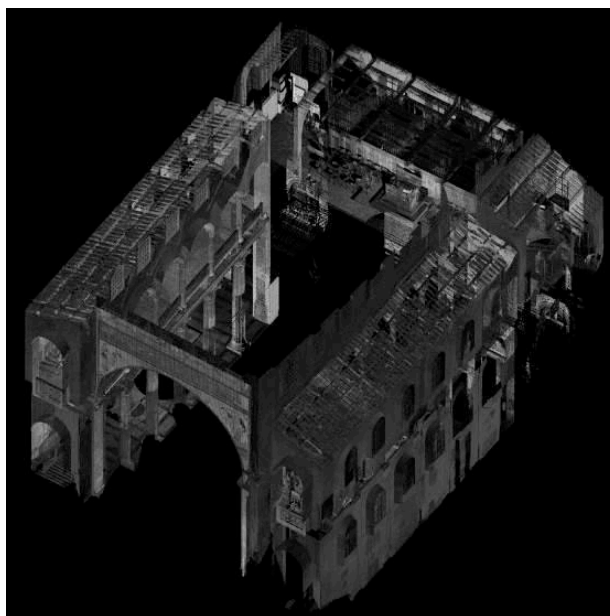


Figure 3: Axonometry of the pelagian basilica.

We examined the construction techniques used in the different sectors of the Basilica like the use of mats in the vaults of the cloister or the extended reuse of roman and medieval architectural elements. In particular, we focused on a roman pillar reused in the trabeation of the Pelagian Basilica and on a high-medieval marble inscription reused as a window screen in the Honorian Basilica (Fig. 4).



Figure 4: Scan of the roman pillar.

To perform in-depth architectural as well as stylistic analysis, we documented these elements also through Cyclone high-density scans (3.1 mm at 10 m) and high definition photos, and by additional. We imported portions of the point clouds into Hexagon 3D-Reshaper software to create a mesh on which high-resolution photos were applied obtaining 3D models with fine textures. By carefully studying the practice of reusing previous architectural elements, we were able to identify, record and relocate an early 11th century transenna of the cloister, today walled in the south-eastern wall, on a 3D model of the Pelagian church in a CAD environment (Fig. 5).

The preliminary results of the catacomb survey, which included part of the Pelagian Basilica and the cloister, allowed documenting the interaction between the catacomb and the related buildings, the several levels of the catacomb and some second thoughts about the execution of the galleries as witnessed by blind tunnels.



Figure 5: 3D mesh of the XIth Century transenna.

One of the most important results reached so far by this project is that it involved participants from various humanistic research areas, encouraging them to communicate with each other, overcome methodological and theoretical differences among disciplines and make an effort in order to find mutual research tools. This partnership, integrated with the use of the laser scanner technology, allowed to reconsider some architectural and historical issues related to the complex and multi-layered building of San Lorenzo Fuori le Mura by seeing them from new points of view.

4. BIBLIOGRAPHY

- [1] Bagordo G.M. 2013, *Evoluzione del sito e della Basilica*, in Cundari et al. 2013, pp. 15-24.
- [2] Broccoli U. (ed.) 1981: Broccoli (ed.), *Corpus della scultura altomedievale. La diocesi di Roma*, 5. Il Suburbio, 1, Spoleto 1981.
- [3] Ciranna S. 2000, *Spolia e caratteristiche del reimpiego nella Basilica di San Lorenzo fuori le mura a Roma*, Roma 2000.
- [4] Cundari C., Bagordo G.M., La Mantia M., Lanfranchi F. 2013, *S. Lorenzo fuori le mura*, Roma 2013.
- [5] Da Bra P.G. 1931, *Le iscrizioni latine della Basilica di S. Lorenzo Fuori Le Mura, del chiostro e delle catacombe di S. Ciriaca*, Roma 1931.
- [6] Da Bra G. 1952, *San Lorenzo fuori le mura*, Roma 1952.
- [7] Fancelli P. 2000, *Restauro e spolia*, in Ciranna 2000, Roma 2000, pp. 9-20.
- [8] Krautheimer R. 1962, *Corpus Basilicarum Christianarum Romae. Le Basiliche Paleocristiane di Roma (IV-IX sec.)*, II, Citt. del Vaticano-New York 1962.
- [9] Matthiae G. 1966, *S. Lorenzo fuori le Mura*, Roma 1966.
- [10] Muñoz A. 1944, *La Basilica di S. Lorenzo fuori le mura*, Roma 1944.
- [11] Pani Ermini L. (ed.) 2000, *La Visita alle "Sette Chiese"*, Roma 2000.
- [12] Passarella R. (ed.) 2015, *Il culto di San Lorenzo tra Roma e Milano: dalle origini al Medioevo*, Roma 2015.
- [13] Remondino F., Campana S. (eds.) 2014, *3D recording and modeling in archaeology and cultural heritage: theory and best practices (BAR International Series, 2598)*, Oxford 2014.
- [14] Romano A. 2013, *Contributi alla restituzione delle basiliche Pelagiana e Onoriana di San Lorenzo fuori le Mura in Roma*, Roma 2013.
- [15] Serra S. 2000, *San Lorenzo fuori le mura*, in Pani Ermini 2000, pp. 101-111.
- [16] Serra S. 2015, *Le fonti e l'archeologia. Alle origini del culto di San Lorenzo a Roma*, in Passarella 2015, pp. 29-53.

Tracing Showrunners' Impact

Joanna Byszuk

Institute of Polish Language – Polish Academy of Sciences, Poland - joanna.byszuk@ijp.pan.pl

ABSTRACT

Relatively little attention has so far been given to stylometry of audiovisual works, and combining distant reading (or rather: watching) methods with traditional analysis with most notable works concerning translation [5, 6] and idiolects [8]. Proposed paper builds on the results obtained in a research project which has now been developing for almost a year, and which examines the use of stylometric methods in studying showrunners' stylistic fingerprint and impact on writers subordinate to them. Showrunner, a function especially important in author-oriented television writing [7], is a person who combines roles of executive producer and head writer, and as such sets the tone of the show.

Proposed paper presents the ongoing project dedicated to examining the influence of the showrunner - a person who combines roles of lead writer and executive producer – on the style of individual writers using computer text analysis methods, in particular stylometry. The paper briefly presents the results of the close study conducted on the corpus of Doctor Who, which served as a model for this study and allowed for creating first hypotheses: the distinct impact of showrunner on other writers and usefulness of applying stylometric methods traditionally used in the analysis of literary texts in the stylometric analysis of texts written for television.

The paper includes an overview of particular appliances of specific methods, among others, network analysis in discovering topics the interest in which distinguishes showrunners and in finding both well-known and less obvious relations between given writers within the series and Rolling Delta (Eder 2016), machine learning based classifier in examining shifts in authorship in individual texts included in the corpus. The main part of the paper is an introduction to a new stage of the project, namely verification of these hypotheses on a larger corpus of seven showrunners and television productions of various genres. The aim of this stage is to see if observed patterns are specific to the previously examined duo of Doctor Who showrunners Davies and Moffat or if they may be a feature characteristic to showrunners in a more general perspective independent of the genre or observable only in science-fiction.

KEYWORDS

stylometry, corpus linguistics, television studies, showrunner

1. EXTENDED ABSTRACT

Despite obvious benefits of applying distant reading methods to large collections of texts, well-known from literature studies in Digital Humanities, such as discovering unknown relations between considered works and certain trends and patterns of genre development, relatively little attention has so far been given to stylometry of audiovisual works, and combining distant reading (or rather: watching) methods with traditional analysis, with most notable works concerning translation [5, 6] and idiolects [8]. Proposed paper builds on the results obtained in a research project which has now been developing for almost a year, and which examines the use of stylometric methods in studying showrunners' stylistic fingerprint and impact on writers subordinate to them. Showrunner, a function especially important in author-oriented television writing [7], is a person who combines roles of executive producer and head writer, and as such sets the tone of the show.

The first part of the project, results of which were presented by the author during April Conference Fourteen (paper: The Voices of Doctor Who), showed distinctiveness of stylometry of Doctor Who showrunners, Russell T Davies and Steven Moffat. Network analysis revealed a very clear stylistic division of the series at the time of takeover of the position, especially considering general tone of the series rather than that of individual characters. When compared with the observations made via close watching, in each examined aspect the distinction between showrunners appeared to be more important than other considered factors – extracted characters' lines, topics of particular episodes or full dialogue lists.

Noticeable was also a relative distinctiveness of several writers. Whereas the style of the showrunners frequently appeared to be significantly different than that of other writers, only a few writers seemed to have a well-visible authorial signal and to be able to navigate their style between eras, writing more similarly to the showrunner of the time. Coincidentally, these were also some of the most prolific and experienced of Doctor Who creators. Interestingly, in Moffat's era these writers seemed to be a second driving force next to the showrunner's influence. While it is impossible to determine that without conducting interviews with the writing team, this could be a sign of either a more democratic or a less coherent process of writing the show, which is one of the issues further explored in the project.

The methods applied to the study included a variety of multidimensional machine-learning methods, both exploratory and supervised. The example of the first were bootstrap consensus networks (a variant of hierarchical clustering), and of the latter – Support Vector Machines used sequentially. The methods were used in such a form as implemented in the stylo package [3]. Network analysis performed with stylo() function of the package helped in measuring the similarity of the texts in the corpus, calculating distances between the texts based on the distribution of most frequent words and allowed for unraveling the division into two showrunner clusters (see fig. 1). Oppose(), function for performing contrastive analysis based on Craig's Zeta, pointed out different thematic interests of them, confirming scholars' close watching observations. Rolling.delta(), which is a machine learning algorithm useful for detecting shifts in authorship of one text [2], produced

interesting results for the texts which are known to have been heavily edited by the showrunner, suggesting that the method could be further applied to examine showrunner's influence on individual works (see fig. 2).

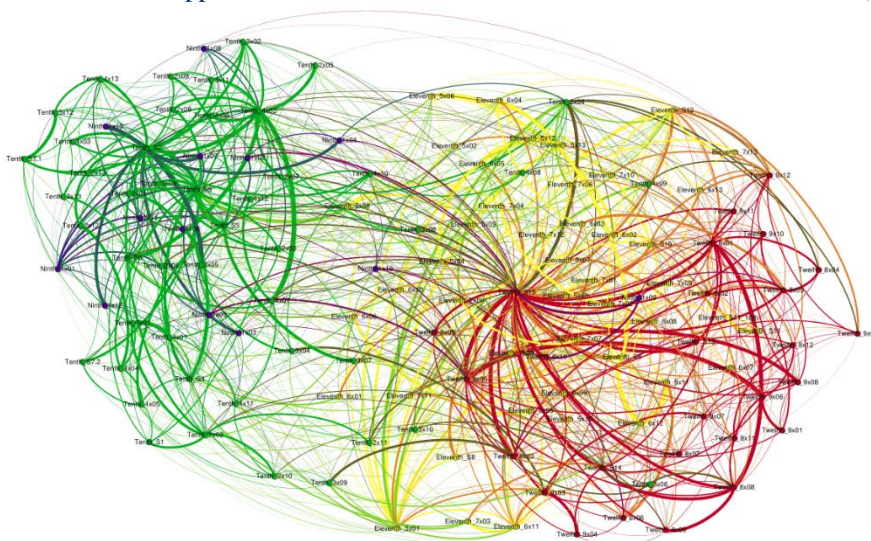


Fig.1 Visualisation of the bootstrap consensus network of Doctor Who episodes (new series 2015-2016). Green and violet marks episodes from Davies's tenure, and yellow and red – from Moffat's. Each color represents one of the Doctor's regenerations. Observable are two main hubs, one for each showrunner, which reveals showrunner's impact as more significant than that of Doctor's regeneration.

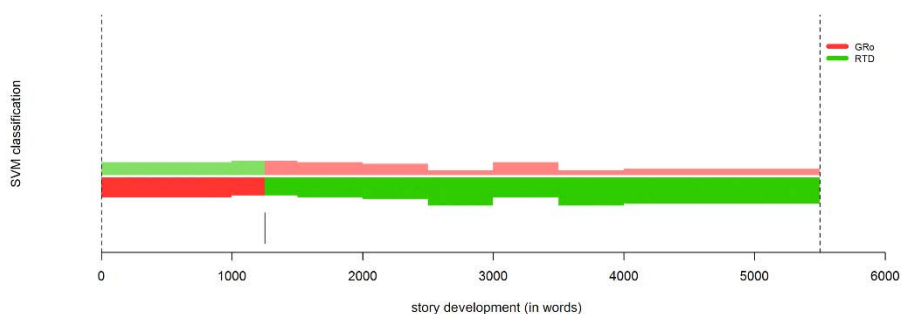


Fig.2 “The Planet of the Dead” (Doctor Who) contrasted sequentially against R.T. Davies (green) and G. Roberts (red); 100 most frequent words. The lower bar shows the stronger writer's influence at a given part of the story and, through its thickness, its intensity. Noticable is the dominance of R.T. Davies's style, especially towards the end of the story.

These early promising results sparked further interest in studying stylometry of prominent showrunners, prompting the questions of whether such distinct influence of Doctor Who creators was a unique event or a feature characteristic only to the genre of science-fiction, seemingly often less restraining and more author-oriented, or a more general observation. In order to investigate these issues in a wider perspective, at the current stage of the project the corpus of considered series – in the form of files containing dialogue lists of given show episodes – was significantly extended, now including samples of other science-fiction series as well as samples of shows belonging to other genre. The series were chosen based on two conditions: a) they had to have a prominent showrunner who b) also contributed to the series as writer of at least 10% of the episodes (thus enabling detection of their stylometric signal). Whenever possible, more than one series by a given showrunner was included in order to examine if showrunner's style would be detectable in between different series after excluding words characteristic to only one of them. Apart from Davies and Moffat (Doctor Who, Sherlock), the examined showrunners are Joss Whedon (Buffy the Vampire Slayer, Firefly), Terence Winter (Boardwalk Empire, The Sopranos), Shonda Rhimes (Grey's Anatomy, Scandal) and Tina Fey (Unbreakable Kimmy Schmidt, 30 Rock). Additionally considered was Steven S. DeKnight, working as a script editor and writer for Buffy and a decade later running his own show Spartacus.

While examining the showrunners credited as writers-creators in every episode of the show (e.g. J.J. Abrams – Lost, Fringe) might be interesting, they were insofar excluded as with no reliable information about how much they actually contributed to writing, extracting their stylome – stylistic voice – would be rather impossible to achieve, or at least, to produce reliable results.

In the proposed paper presented will be the results of the analysis conducted on this extended corpus in the context of the mentioned observations made at the earlier stage of the project. Considering a larger and more diverse set of data will facilitate verifying earlier hypotheses made after Doctor Who analysis, and searching for common features of author-oriented television productions giving more insight into the specifics of the mysterious art of running a show.

2. REFERENCES

- [1] Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, 2009.
- [2] Eder, M. (2016). Rolling stylometry. *Digital Scholarship in the Humanities*, 31(3): 457-469.
- [3] Eder, M., Kestemont, M. and Rybicki, J. Stylometry with R: A package for computational text analysis. *R Journal*, 16(1), 2016: 107-121.
- [4] Eder, M. Visualization in stylometry: cluster analysis using networks. *Digital Scholarship in the Humanities*, 32(1), 2017: 50-64.
- [5] Hołobut, Rybicki, Woźniak, *Stylometry on the Silver Screen: Authorial and Translatorial Signals in Film Dialogue*, 2016.
- [6] Hołobut, Rybicki, *Pride and Prejudice and Programming: The Novel and its Adaptations in Stylometric Analysis*, 2017.
- [7] Mittel, J. *Complex TV: The Poetics of Contemporary Television Storytelling*. NYU Press, 2015, as accessed: <http://mcpress.media-commons.org/complextelevision/authorship/>
- [8] Van Zyl, Botha, Stylometry and characterisation in *The Big Bang Theory*, 27. *Literator* 37(2), a1282. <http://dx.doi.org/10.4102/lit.v37i2.1282>, 2016.

Una proposta di annotazione semantica di concetti musicali nel *de institutione musica* di Boezio

Francesca Michelone

Università del Piemonte Orientale, Italia - francesca.michelone@gmail.com

ABSTRACT

Questo contributo presenta una proposta di annotazione semantica di concetti musicali nel primo libro del *de institutione musica* di Boezio. Si tratta di un'interpretazione formalizzata che viene inserita manualmente all'interno del testo secondo lo standard internazionale XML/TEI. È quindi un lavoro che si concentra non sulla struttura del testo, ma sul suo contenuto. Il valore dell'annotazione semantica di concetti risiede nel fatto che permette di fare ricerche nel testo servendosi dei concetti e non delle parole.

Per poter annotare un testo attraverso i concetti in esso presenti in primo luogo è stato necessario creare una tassonomia, che è stata definita attraverso la lettura degli autori enciclopedici tardi che nelle loro opere parlano di musica antica, in particolare: Censorino, Marziano Capella, Cassiodoro, Boezio, Isidoro di Siviglia. I concetti individuati riguardano gli strumenti musicali antichi, le teorie sul suono, lo studio della melodia, la notazione e il ritmo. Tra questi sono stati scelti per la tassonomia: le corde (*chorda*), il tetracordo (*tetrachordum*) con i suoi generi diatonico, cromatico ed enarmonico, gli intervalli (*intervallum*), specificati in tono e semitono, il suono (*sonum*) e la consonanza (*consonantia*).

Questi argomenti sono i primi ad essere proposti perché compaiono nel primo libro del *De institutione musica*, scelto come punto di partenza perché è l'unico testo latino tardo completamente dedicato a questo tema, ma è possibile ampliare questa tassonomia con l'aggiunta di altri concetti. L'interpretazione così proposta nel primo libro del trattato di Boezio può essere inserita anche negli altri testi degli autori tardi prima citati o in altre opere non letterarie appartenenti al patrimonio culturale che rientrano in questa prospettiva. Annotando l'insieme delle opere che riguardano la musica antica è possibile collegare tra loro i passi messi in luce, restituendo così una visione complessiva degli argomenti annotati.

PAROLE CHIAVE

Annotazione semantica; XML/TEI; musica; concetti; ricerca dei concetti nel testo

1. L'ANNOTAZIONE SEMANTICA DI CONCETTI

Il campo dell'annotazione semantica di concetti è innovativo e ha molto da dare allo studio del testo: una volta create le biblioteche digitali è necessario farne uso in modi che siano realmente innovativi nello studio dei testi. L'annotazione semantica permette di studiare i testi partendo dall'individuazione dei concetti e quindi consente un approccio al testo più fine e complesso di quello semplicemente basato sulle occorrenze delle parole, generalmente cercate in quanto segnali della presenza di un concetto. Inoltre, formalizzare appropriatamente un'interpretazione ed esprimerla per mezzo di annotazioni consente di costruire elementari interrogazioni ("in quali autori del periodo X il concetto A è specificato dal concetto B?") che aprono prospettive realmente innovative per lo studio dei testi.

Questo contributo presenta in forma di prototipo una proposta di annotazione semantica di concetti e argomenti musicali presenti nel primo libro del *De institutione musica* di Boezio (nell'edizione [1] disponibile nella biblioteca digitale del latino tardo digilibLT [2], www.digiliblt.uniupo.it), che ha come filo conduttore la musica antica ([3], [8], [13]). Si tratta di un'interpretazione formalizzata che viene inserita manualmente nel testo utilizzando lo standard internazionale XML/TEI [5]; è quindi un lavoro che non si concentra sulla marcatura e descrizione della struttura formale dell'opera, ma sul contenuto e sulla sua interpretazione. L'annotazione semantica è usualmente annotazione di *named entities*, [6], [14] mentre in questo caso l'annotazione riguarda l'interpretazione di specifici passi di testo.

Per poter effettuare l'annotazione è necessaria (e quindi occorre definire) una struttura concettuale che descriva lo specifico ambito di conoscenza di proprio interesse: nella sua forma più semplice è una tassonomia, nella sua forma più complessa questa struttura è un'ontologia (una web ontology se si vuole favorirne l'utilizzo operativo). In ogni caso si deve partire da un'esplorazione del dominio concettuale e quindi il lavoro è iniziato con la lettura dei testi latini tardi che trattano in tutto o in parte di musica, in particolare il *De institutione musica* di Boezio, il *De Die Natali* di Censorino con la sua epitome, il *Fragmentum Censorini*, il *De nuptiis Mercurii et Philologiae* di Marziano Capella, le *Institutiones* di Cassiodoro e le *Etymologiae* di Isidoro di Siviglia. La lettura di queste opere ha permesso di individuare e definire i concetti di base: gli strumenti musicali, le teorie riguardanti il suono, le basi teoriche della composizione musicale antica, la notazione e il ritmo. Una volta determinati i concetti di base si è passati alla scelta dei termini più appropriati per indicarli e sono stati creati i marcatori corrispondenti. Gli argomenti sviluppati in questo contributo sono quelli che trovano riscontro nel primo libro del trattato di Boezio, quindi in particolare le definizioni di suono, gli strumenti a corda e i principi fondamentali delle teorie musicali antiche. L'ultima fase del lavoro è stata tornare sul testo di Boezio per inserire l'annotazione.

Due osservazioni importanti: trattandosi di uno studio di fattibilità si è proceduto allo sviluppo non di un'ontologia ma di una semplice tassonomia di concetti:

- le corde (*chorda*)
- il tetracordo (*tetrachordum*) con i suoi generi diatonico (*diatonum*), cromatico (*chroma*) ed enarmonico (*enarmonium*)

- gli intervalli (*intervallum*), specificati in tono (*tonus*) e semitono (*semitonium*)
- il suono (*sonum*)
- la consonanza (*consonantia*).

Inoltre, poiché i concetti individuati derivano dalla lettura complessiva dei passi riguardanti la musica in tali opere, la tassonomia individuata anche se in questo studio è stata utilizzata per annotare passi nel *De institutione musica* di Boezio (scelto perché è l'unico testo latino tardo che si occupa in modo specifico di musica per tutta la sua estensione; è una *summa* delle teorie musicali antiche e le trasmette dall'antichità al medioevo; presenta una buona coerenza lessicale interna per quanto riguarda i termini tecnici musicali) in realtà può essere utilizzata per l'annotazione semantica di concetti in tutti i testi latini tardi prima citati da cui è iniziata la creazione della tassonomia.

Dopo aver individuato i concetti è stato scelto il marcatore TEI da usare nell'annotazione, e si è deciso di utilizzare *interp* (in associazione con *interpGrp*). Il marcatore che definisce il concetto di tetracordo con le sue sottocategorie, ad esempio, è stato strutturato nel TEI-header come segue:

```
<interpGrp type="tetrach">
  <interp xml:id="tetrach">tetrachordum</interp>
  <interp xml:id="diaton">diatonum</interp>
  <interp xml:id="chrom">chroma</interp>
  <interp xml:id="enarm">enarmonium</interp>
</interpGrp>
```

Infine i marcatori sono stati usati per l'annotazione manuale di alcuni passi esemplificativi del trattato di Boezio, si veda per esempio I, 21 *De generibus cantilenae*:

```
<p><seg ana="#tetrach">His igitur expeditis dicendum de generibus melorum. Sunt autem tria: diatonum, chroma, enarmonium</seg>. <seg ana="#diaton">Et diatonum quidem aliquanto durius et naturalius</seg>, <seg ana="#chrom">chroma vero iam quasi ab illa naturali intentione discedens et in mollius decidens</seg>, <seg ana="#enarm">enarmonium vero optime atque apte coniunctum.[...]</seg> </p>
```

Il testo predisposto per essere annotato per il momento è quello del primo libro del *De institutione musica*, ma sarebbe interessante in un futuro prossimo portare a termine l'annotazione ed estenderla agli altri testi dove è possibile inserirla. L'alternativa tra annotazione semantica automatica o manuale (cfr. [7]) nel caso di testi in lingue antiche, come il latino, non si pone [8] e peraltro l'annotazione manuale è sostanzialmente indispensabile quando il contenuto che si annota deriva da un'attività interpretativa.

Il valore dell'annotazione semantica di concetti risiede nel fatto che permette di fare ricerche nel testo "utilizzando" i concetti e non solo le parole (peraltro c'è chi [12] obietta che il solo denominare i concetti non è sufficiente a comunicare il contenuto di pensiero della fonte testuale: ma se l'annotazione dei concetti è concepita come un modo per accedere ai passi corrispondenti, come in questo lavoro, il problema non si pone): è possibile inserire l'annotazione per un concetto anche dove il termine di riferimento del marcatore non è presente, perché sostituito da un sinonimo, da una perifrasi o sottinteso; e non si annotano i passi dove la parola è presente ma ha un'altra accezione. È questo, ad esempio, il caso dell'annotazione del concetto di consonanza: il termine *consonantia* è utilizzato in modo sistematico per la prima volta nella trattatistica musicale latina da Boezio, mentre i suoi predecessori utilizzavano un'ampia gamma di sinonimi per identificare questo fenomeno, come *symphonia*, *convenientia*, *concordia* o *concentus* [10]. Grazie a questo tipo di annotazione semantica è possibile marcare tutti i passi dove si parla di consonanza anche se sono presenti i suoi sinonimi. Il suddetto utilizzo dell'annotazione può essere ben esemplificato dai seguenti passi dove sono presenti i sinonimi *symphonia* e *concordia* ed è assente il termine *consonantia*:

```
<seg ana="#conson"> Est autem symphonia duarum vocum disparium inter se iunctarum dulcis concentus. </seg>
```

Censorino, *De Die Natali*, X, 6 [4]

```
<seg ana="#conson"> Tetrachordum quippe est quattuor sonorum in ordinem positorum congruens fidaque concordia. </seg>
```

Marziano Capella, *De nuptiis Philologiae et Mercurii*, IX, 961 [9]

Si apre così la strada alla possibilità di analisi intertestuali, che possano agevolare il confronto tra più testi attraverso i concetti che li fondano e non solo attraverso le concordanze delle parole.

In questo approccio interpretativo al testo, che è specifico dell'annotazione semantica di concetti, si riversa all'interno del testo la conoscenza che noi ne possediamo e poiché annotare un testo comporta scelte che talvolta possono essere evidenti, altre volte meno, è importante in uno sviluppo futuro individuare se e come nell'annotazione si possano motivare e argomentare i criteri che hanno portato ad una determinata interpretazione.

I testi annotati entrano come tali nella biblioteca digitale, quindi non sono utili soltanto per chi ha inserito l'annotazione, ma anche per coloro che sono interessati ad un determinato testo o argomento. Lavorare sui beni culturali digitalizzati in questo modo può favorire la condivisione delle ricerche e la collaborazione tra studiosi.

2. BIBLIOGRAFIA

- [1] Boezio, Anicio Manlio Severino, *De Institutione Arithmetica libri duo, De Institutione Musica libri quinque*, a cura di Gottfried Friedlein, Leipzig, Teubner, 1867
- [2] Boezio, Anicio Manlio Severino, *De Institutione musica*, a cura della biblioteca digitale del latino tardo DigilibLT, URL: <http://digiliblt.lett.unipmn.it/opera.php>
- [3] Boezio, Anicio Manlio Severino, *De Institutione musica*, a cura di Giovanni Marzi, Roma, Istituto italiano per la storia della musica, 1990
- [4] Censorino, *De die natali liber ad Q. Caerellium; accedit anonymi cuiusdam epitoma disciplinarum (Fragmentum Censorini)*, a cura di Nicolaus Sallmann, Leipzig 1983, (Bibliotheca Scriptorum Graecorum et Romanorum Teubneriana)
- [5] Burnard, Lou; Bauman, Syd, (a cura di) *TEI Guidelines Version 3.1.0. Last updated on 15th December 2016*, revision d3f5e70, <http://www.tei-c.org/Guidelines/P5/>
- [6] Codice Pelavicino, edizione digitale a cura di E. Salvatori, URL: <http://pelavicino.labcd.unipi.it/>
- [7] De Maio, C., Fenza, G., Gallo, M. et al., *Formal and relational concept analysis for fuzzy-based automatic semantic annotation*, in "Applied Intelligence" (2014) 40: 154. <https://doi.org/10.1007/s10489-013-0451-7>.
- [8] Mainoldi, Ernesto, *Ars musica: La concezione della musica del Medioevo*, Milano, Rugginenti, 2001
- [9] Marziano Capella, *Martiani Capellae, De nuptiis Philologiae et Mercurii: liber 9*, a cura di J. Willis, Leipzig, 1983
- [10] Petretto, M. Alessandra (2003), *Per uno studio sul lessico latino della Harmonica Disciplina*, in: "Sandalion" Vol. 23-25 (2000-2002), p. 65-93
- [11] Piao, C., Dallachy, F., Baron, A., et al., *A time-sensitive historical thesaurus-based semantic tagger for deep semantic annotation*, in "Computer Speech & Language", Volume 46, 2017, Pages 113-135, ISSN 0885-2308, <https://doi.org/10.1016/j.csl.2017.04.010>.
- [12] Prasad, A.R.D., Guha, N., *Concept naming vs concept categorisation: a faceted approach to semantic annotation*, in "Online Information Review", Vol. 32 Issue: 4, 2008, pp.500-510, <https://doi.org/10.1108/14684520810897377>.
- [13] Sachs, Curt, *La musica nel mondo antico: Oriente e occidente*, Milano, Rusconi, 1992
- [14] Vespasiano da Bisticci, *Lettere*, a cura di F. Tomasi, URL: <http://vespasianodabisticciletters.unibo.it/>

Voci della Grande Guerra

Preserving the Digital Memory of World War I

Alessandro Lenci¹, Nicola Labanca³, Claudio Marazzini⁴, Simonetta Montemagni²,
Federico Boschetti², Irene De Felice¹, Stefano Dei Rossi⁵, Felice Dell’Orletta²,
Michele Di Giorgio^{1,2,3}, Lucia Passaro¹, Giulia Venturi²

¹ Università di Pisa, Laboratorio di Linguistica Computazionale (CoLing Lab), Pisa, Italia

² Istituto di Linguistica Computazionale - CNR, Italia

³ Università di Siena, Dipartimento di Scienze Storiche e Beni Culturali

⁴ Università di Siena, Dipartimento di Scienze Storiche e Beni Culturali

⁵ WebSoup

ABSTRACT

Voci della Grande Guerra is a scientific and cultural initiative with the aim of preserving and promoting the memory of Italy in World War I through the creation of an annotated corpus of digital texts selected by historians and linguists in order to be representative of the different ways to experience and describe the Italian war by its protagonists.

KEYWORDS

Computational linguistics, Digital history, World War I, annotated corpus

1. INTRODUCTION

The Great War is the first war of mass death, but it is also the first war of mass text production. An important part of such texts are also first-hand accounts by people endeavouring the experience of writing for the first time, to make sense of the dramatic and disruptive events they witnessed. Increasing amounts of such sources are available in digital form, but many historical documents still need to be digitized. *Voci della Grande Guerra* (Voices of the Great War: <http://www.vocidellagrandeguerra.it/>) is a scientific and cultural initiative with the aim of preserving and promoting the memory of Italy in World War I (WWI) through the creation of an annotated corpus of digital texts selected by historians and linguists in order to be representative of the different ways to experience and describe the Italian war by its protagonists. With the help of advanced techniques of computational linguistics, the digitized historical materials will be explored with an online interface to enable easy but effective and innovative search modalities. These will allow experts as well as non-experts to become acquainted with and appreciate the “linguistic polyphony” of Italy during WWI: the official voices of the propaganda and the voices of soldiers, the voices of newspapers and the voices of the letters, the intellectual elite’s and the people’s voices, the voices of consensus to war and the voices of dissensus. The project *Voci della Grande Guerra* is part of the official initiatives for the Celebrations of the 100th Anniversary of WWI: it started in May 2016 and will end in September 2018.

2. ASSEMBLING THE VOICES OF WWI

The project *Voci della Grande Guerra* is extremely relevant both from the historical and the linguistic point of view. If WWI as a factual event is quite well-known, much less known are the different narrative and experiential perspectives on this war. The texts produced (with different purposes) in that period have had a crucial role in shaping the images of war before, during and after the conflict. Linguists have always ascribed a very important function to the Great War as a decisive time in the process leading to the linguistic unification of Italy [2], because imposing masses of men from different regions of the peninsula were forced to live together for months in the trenches and behind the lines, and were forced to use the national language as the main communicative medium, in contact with more educated officers possessing a higher level of Italian.

Voci della Grande Guerra aims at creating an archive of digital texts, most of which never digitized before. Given the practical impossibility of keeping track of all the varieties of the Italian of a century ago, the corpus consists of a selection of texts representative of the most relevant communicative situations to characterize the language of the time:

- the official military language: books of military strategy, analysis war conduct, senior officers’ memoirs and diaries; propaganda texts and court martial records;
- the language of the middle class: samples of low rank officers’ diaries and memoirs, most likely written in a high-level Italian inspired to major literary examples of the time;
- the popular language: examples of letters, diaries and memoirs from soldiers (and from the common people);
- the language of the political class: samples of parliamentary proceedings, official speeches and secret sessions of the parliament;
- the language of the intellectual elite: samples of pamphlets, literary journals, etc.;
- the standard language of public opinion: samples of newspaper articles, magazines, news reports from the front, etc.

The final corpus will be balanced along various dimensions: textual genres, author type, time, education, etc. The corpus will include texts from the 1913 up to the early ‘20s, in order to cover not only the years of the war, but also the cultural

and social environment leading to the war and the aftermath of the Great War. We will also balance the texts with respect to the various war years, in order to investigate empirically the immediate impact of the war and of its different phases (e.g., before and after Caporetto) on language and communicative styles.

In the great majority of cases, when not available in digital form, the texts selected for *Voices of the Great War* are digitized with high resolution scanners and then analyzed with optical character recognition software (OCR). OCR performance is closely dependent on the quality of the text to be scanned. For this reason, the most advanced techniques of multiple OCR output alignment are used with a “voting” system, already experienced in previous work on historical documents, to increase the accuracy of character recognition. The final output is checked and corrected manually, and later codified in the TEI-XML standard format. Collaborative proof-reading involves not only the members of the project, but also volunteers, through the WikiSource platform, in collaboration with the Biblioteca del Comune di Trento.

The digitized texts undergo the following computational processing [1]:

- automatic analysis of the linguistic structure of the texts - Lemmatization, morphological and syntactic analysis, etc.;
- semantic information extraction from texts - Extraction of simple terms (e.g., *irredentismo* “irredentism”) and complex terms (e.g., *terre irredente* “unredeemed land”, *gas asfissianti* “poisonous gas”, etc.) significantly associated with different texts types, and named entity recognition (recognized named entities will include person names like *Luigi Cadorna*, location names like *Ortigara*, military units like *9 Reggimento Bersaglieri*, etc). The locations mentioned in the texts are also normalized with respect to spelling variations and associated with their geographic coordinates. Named entities will be linked to existing knowledge bases such as DBpedia;

The linguistic annotations of about 1 million tokens will also be checked manually, thereby representing a sort of “gold” subset of the whole corpus. The remnant of the corpus will instead be annotated automatically with a random checking for errors. The pilot project *Memorie di Guerra* (War Memories: <http://www.memoriediguerra.it>) [1] shows in action a part of the tools and of the analyses that will be applied to the new, extended corpus.

The project will also develop a software platform to assist researchers during the phases of corpus building, and to provide various search functionalities. The tool will consist in a back-end module to support the correction of digitized and automatically annotated texts, and a front-end module for the exploration of the corpus with advanced forms of information visualization and query, to perform both “close” and “distant” readings of the texts [3].

3. ACKNOWLEDGEMENTS

The project *Voci della Grande Guerra* is funded by a two-year grant from the Special Mission for the Celebrations of the 100th Anniversary of WWI at the Presidenza del Consiglio dei Ministri of the Italian Government.

4. REFERENCES

- [1] Boschetti, Federico, Andrea Cimino, Felice Dell'Orletta, Gianluca E. Lebani, Lucia Passaro, Paolo Picchi, Giulia Venturi, Simonetta Montemagni, and Alessandro Lenci. 2014. “Computational analysis of historical documents: An application to Italian war bulletins in WWI and WWII”. In *Proceedings of the LREC 2014 Workshop on “Language resources and technologies for processing and linking historical documents and archives – Deploying Linked Open Data in Cultural Heritage”* (LRT4HDA 2014), Reykjavik.
- [2] De Mauro, Tullio. 1963. *Storia linguistica dell'Italia unita*. Laterza, Bari.
- [3] Moretti, Franco. 2013. *Distant Reading*. Verso, London.

AIUCD 2018
Poster papers

Aristotele sull' *holodeck*

Luca Bandirali

Università del Salento - luca.bandirali@unisalento.it

Da diversi anni ci si interroga sulle molteplici direzioni che i media audiovisivi hanno intrapreso dopo l'avvento delle tecnologie digitali, e su quelle che potranno intraprendere in futuro. Le ipotesi principali sono state sintetizzate da due grandi archetipi finzionali: 1 - l'*holodeck* di *Star Trek Next Generation*, un ambiente di simulazione olografica che immerge l'utente in situazioni legate alla storia e alla letteratura (ipotesi *narrativa*); 2 - le rappresentazioni non lineari sui monitor di *Fahrenheit 451* di Ray Bradbury e il cinema sensoriale de *Il nuovo mondo* di Aldous Huxley (ipotesi *antinarrativa*). Cosa hanno in comune oggi prodotti audiovisivi come i videogiochi, gli *unscripted format* come i reality e i talent show, una web serie e un film in 3D? Sono generati e gestiti a vari livelli da un unico medium, il computer; raccontano storie, e le storie che raccontano condividono le medesime strutture narrative profonde, anche quando ci si offrono come pure "riprese" del reale (Chalaby 2016). Tra la tendenza degli integrati di *Star Trek* e quella degli apocalittici di Huxley, sembra prevalere attualmente la prima. L'atto narrativo funziona ancora come nella *Poetica* di Aristotele: è un racconto in tre parti (inizio, mezzo e fine) basato sul conflitto e destinato a descrivere la trasformazione di uno o più caratteri. Le storie possono svilupparsi su media differenti (*transmedia storytelling*), ma tendono a costruire un mondo coerente che si modifica drammaturgicamente in termini di causalità: il Marvel Cinematic Universe, costruito da comics, film, serie tv, web series, short movies per il web, prevede che ogni segmento narrativo porti avanti una sola grande storia. Il modello precedente di questa grande architettura narrativa, in fondo, è la mitologia: un mondo narrativo complesso con personaggi ricorrenti che sono protagonisti di un gran numero di storie. Anche la narrazione mitica, per inciso, obbediva a una struttura profonda tripartita e trasformativa. Raccontare in epoca transmediale, dunque, è ancora la stessa cosa rispetto a 2.500 anni fa, e al tempo stesso non lo è. Alcuni aspetti specifici del nuovo medium hanno una forte influenza sulle nuove forme di racconto; se ci soffermiamo sulle quattro proprietà fondamentali dell'ambiente digitale individuate da Janet H. Murray (1998 e 2011), cioè l'essere procedurale, partecipativo, spaziale ed enciclopedico, ci rendiamo conto di come alcune di queste proprietà abbiano introdotto delle discontinuità rispetto al modo di narrare precedente. In particolare, se pensiamo alle narrazioni seriali, l'assetto attuale di questo prodotto audiovisivo sembra aver incrementato iperbolicamente la propria dimensione enciclopedica (una sorta di *epos*, il contenitore di tutto il passato della storia raccontata), e di aver procrastinato indefinitamente il concetto di chiusura. La difficoltà di chiusura introduce una criticità nel sistema narrativo transmediale nella misura in cui l'ampio movimento orizzontale di situazioni e azioni fatica a trovare una convergenza causale; Noël Carroll ricorda che esistono narrazioni strutturalmente prive di chiusura, come le *soap operas* e le *histories of nations* (Carroll 2007) ma sottolinea anche che il primo filosofo narratologo, Aristotele, aveva descritto la tragedia come l'imitazione di un'azione sino alle sue conseguenze finali e considerava il genere tragico superiore a quello epico proprio per la capacità di abbracciare la narrazione con un unico sguardo, dall'inizio alla fine. Le storie concepite nel sistema transmediale tendono a una pratica intensa di *world building*, la costruzione di ambienti ed esistenti dettagliati e complessi, che si sviluppano nel tempo; ma se la conclusione è da più parti indicata come il luogo del senso di un'opera narrativa (Ryan 2006), il rischio della procrastinazione o dell'interruzione è la carenza di senso. Ci troviamo dunque di fronte a universi narrativi molto stratificati che richiedono una competenza in merito al nuovo assetto dei media, a partire da quello che Janet H. Murray definisce il "single medium" (il computer); ma allo stesso tempo esigono una conoscenza delle strutture narrative profonde, quelle stesse che hanno presieduto alla creazione di tutte le opere della storia dei media narrativi ben prima dell'introduzione delle tecnologie digitali. In questo senso, possiamo dire che la cultura digitale al banco di prova del racconto si rivela strutturalmente umanistica, come un Aristotele sull'*holodeck*.

BIBLIOGRAFIA

- [1] Aristotele (2008), *Poetica*, Torino, Einaudi.
- [2] Bandirali L., Terrone E. (2009), *Il sistema sceneggiatura*, Torino, Lindau.
- [3] Carroll N. (2007), «Narrative closure», *Philosophical Studies* 135 (1):1 – 15.
- [4] Chalaby J. K. (2016), *The Format Age*, Cambridge, UK, Polity Press.
- [5] Jenkins H. (2007), *Cultura convergente*, Milano, Apogeo, 2007.
- [6] Murray J. H. (1997), *Hamlet on the Holodeck*, Cambridge, Mass., MIT Press, 1997.
- [7] Murray J. H. (2011), *Inventing the Medium: Principles of Interaction Design as a Cultural Practice*, Cambridge, Mass., MIT Press.
- [8] Negri E. (2015), *La rivoluzione transmediale*, Torino, Lindau.
- [9] Ryan M. L. (2006), *Avatars of story*, Minneapolis, MN, University of Minnesota Press.
- [10] Zecca F. (2013), *Cinema e intermedialità*, Udine, Forum Editrice.



Fig. 10. Ronta (FI) – Cartolina anni Cinquanta



Fig. 11. Stessa cartolina colorata a mano BN



Fig. 12. Cartolina BN colorata con Zhang 2016



Fig. 13. Cartolina BN colorata con Iizuka 2016

Sopra abbiamo una cartolina di Ronta (FI) stampata in bianco e nero con accanto la versione colorata a mano. Sotto la cartolina in bianco e nero è processata tramite due diversi algoritmi di colorazione automatica, senza nessun intervento successivo di correzione. Gli algoritmi tendono ad essere meno invasivi dell'artista, che poteva avere come riscontro la realtà. La colorazione automatica è più possibilista, valutando ad esempio per terreni forse incolti colori più bruciati, anche se si riscontra in entrambi gli algoritmi una scarsità di giallo che riduce il verde, come si vede negli esempi successivi, dati dal confronto di due foto scattate in modo nativo a colori e in bianco e nero con il processamento poi di quella senza colore⁹.

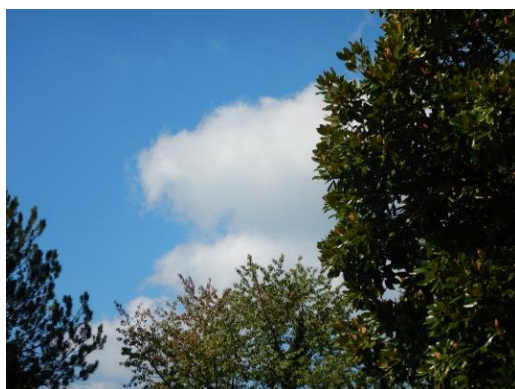


Fig. 14. Nikon Coolpix AW130 – Immagine nativa a colori – Foto dell'autore, ottobre 2017



Fig. 15. Nikon Coolpix AW130 – Immagine nativa in bianco e nero - Foto dell'autore, ottobre 2017

⁹ Ho verificato come le immagini scattate in bianco e nero nativo da macchine digitali non hanno al loro interno dati sul colore. In ogni caso nell'immagine usata sono stati cancellati i dati Exif.

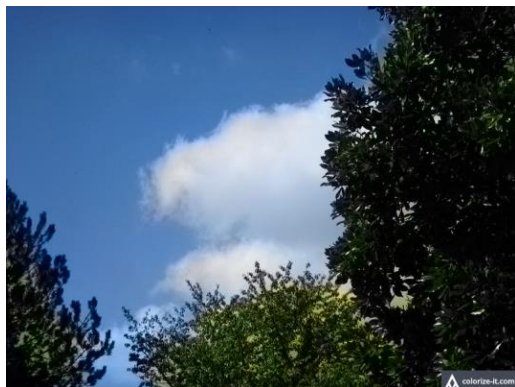


Fig. 16. Nikon Coolpix AW130 BN colorata con Zhang 2016



Fig. 17. Nikon Coolpix AW130 BN colorata con Iizuka 2016

3. ALBANIA

Gli esempi e le premesse teoriche hanno come fine quello di approfondire l'analisi di due serie di immagini in bianco e nero attinenti all'Albania che sto esaminando nell'ambito di un progetto più articolato sulla percezione italiana di quello stato, in uno scarto temporale ampio (1743-1943), tramite l'uso di fonti narrative e iconografiche.

L'esame delle fonti narrative (resoconti di viaggi, diari, articoli) tramite liste di frequenza e *word cloud* può offrire una rappresentazione delle mutazioni che gli italiani introducono nella descrizione dell'Albania. Ogni testo ha le sue caratteristiche e peculiarità linguistiche, tuttavia è possibile far emergere singoli termini o gruppi che vanno a definire con maggior precisione la trasformazione del punto di vista.

Questo aspetto della ricerca è ancora acerbo e solo un'interpretazione generale ci mostra come si passi da una visione "geografica" nella prima metà dell'Ottocento ad un interesse militare e coloniale degli anni successivi.

Le immagini sono due nuclei compatti ed omogenei: primo foto di una missione cartografica italiana in Albania nel 1928 (164 fotografie), il secondo foto di un ignoto militare nel 1941 (85 fotografie). La possibilità di calibrare il risultato su modelli più vicini alla realtà, anche con le prove di avvicinamento a colori originali, permette in senso generale una miglior comprensione del dato finale, che è di analisi sociale, dove il colore è una maggiore informazione tout court, ad esempio sull'uso del suolo in un preciso momento storico, tema che potrebbe essere prioritario in progetti di rimboschimento, agricoltura avanzata, salvaguardia del suolo e restauro architettonico per quanto riguarda l'emergere del colore sugli edifici.



Fig. 18. Albania – 1941 – Modello di colorazione Satoshi Iizuka 2016



Fig. 19. Albania – 1928 – Modello di colorazione Satoshi Iizuka 2016



Fig. 20. Albania – 1928 – Modello di colorazione Iizuka 2016

4. BIBLIOGRAFIA

- [1] *La scienza in cucina e l'Arte di mangiar bene. Manuale pratico per le famiglie compilato da Pellegrino Artusi. 790 ricette. In appendice "La Cucina per gli stomachi deboli". Un ritratto dell'Autore*, Firenze, Landi, 1911.
- [2] R. Dahl. *Automatic Colorization*, Jan 2016, <http://tinyclouds.org/colorize/>
- [3] Deshpande, J. Rock and D. Forsyth. Learning Large-Scale Automatic Image Colorization. In ICCV, Dec 2015., <http://vision.cs.illinois.edu/projects/lscolor/>
- [4] J. Frazer, *The Golden Bough. A Study in Magic and Religion*, London, Macmillan and Co., 1922.
- [5] J. Frazer, *Il ramo d'oro. Storia del pensiero primitivo: magia e religione*, traduzione di Lauro de Bosis, Roma, A. Stock, 1925.
- [6] S. Freud, *Opere di Sigmund Freud*. Edizione diretta da C. Musatti, Torino, Boringhieri, 1966-1980.
- [7] S. Iizuka and E. Simo-Serra and H. Ishikawa, *Let there be Color! Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification*, "ACM Transactions on Graphics (SIGGRAPH)", 2016, V. 35, n. 4, <http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization/en/>
- [8] J. Joyce, *Finnegans Wake*, London, Faber & Faber, 1939.
- [9] J. Harris, *Word clouds considered harmful*, <http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/>
- [10] G. Larsson, M. Maire, and G. Shakhnarovich. *Learning Representations for Automatic Colorization*. In ECCV 2016, <http://people.cs.uchicago.edu/~larsson/colorization/>
- [11] Olivetti, *Ai lavoratori. Discorsi agli operai di Pozzuoli e Ivrea*, Roma, Edizioni di comunità, 2012
- [12] V. Pareto, *Trattato di sociologia generale*, Firenze, Barbera, 1916.
- [13] V. Pareto, *Compendio del Trattato di sociologia generale per cura di Giulio Farina*, Firenze, Barbera, 1920.
- [14] Zhang R., Isola P., Efros A.A., *Colorful Image Colorization*. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. "Lecture Notes in Computer Science", vol 9907. Springer, Cham, <https://arxiv.org/abs/1603.08511>

Colette, Willy and the *Claudine* book series: exploitation or collaboration? Authorship attribution in the novel *Claudine à l'école* (*Claudine at school*, 1900)

Marie Puren

Inria - marie.puren@inria.fr

ABSTRACT

Has the French writer Colette been exploited by her husband Willy? That is what Colette implied when she evoked the *Claudine* book series, published between 1900 and 1903 under her husband's name only. At the time of the publication, Colette was not yet the wellknown author of *Gigi*. In 1893, the young Sidonie-Gabrielle Colette - she was twenty years old - married Willy, "nom-de-plume" of Henry Gauthier-Villars. Fourteen years older than his wife, Willy was a best-selling novelist, well introduced into the intellectual and artistic parisian circles. From 1900 to 1903, four *Claudine* novels were published and signed by Willy: *Claudine à l'école* (*Claudine at school*, 1900), *Claudine à Paris* (*Claudine in Paris*, 1901), *Claudine en ménage* (*Claudine married*, 1903), *Claudine s'en va* (*Claudine and Annie*, 1903). But until her death in 1954, Colette has constantly claimed that she was the true and only author of the book series. Even if it is now clear that Willy has played a role in the writing of these texts, it is still difficult to precisely determine in which extent Willy has influenced and reviewed, or even partly written, the texts produced by his young wife.

To better understand this literary collaboration, I propose to analyze *Claudine à l'école*, first novel of the *Claudine* book series. I will use the possibilities offered by text mining, and more specifically by authorship attribution methods. These methods can offer crucial elements to determine how the two writers influenced each other, and which texts (or parts of texts) have been written by Colette or Willy. This study will allow to better understand the literary relationship developed by Colette and Willy, to shed a new perspective on the history of this intellectual collaboration, and to precisely describe how Colette made her first steps in literature.

KEYWORDS

Text mining, authorship attribution, French literature, Colette, Willy

1. PROPOSAL

Has the French writer Colette been exploited by her husband Willy? That is what Colette implied when she evoked the *Claudine* book series, published between 1900 and 1903 under her husband's name only. At the time of the publication, Colette was not yet the well-known author of *Gigi*. In 1893, the young Sidonie-Gabrielle Colette - she was twenty years old - married Willy, "nom-de-plume" of Henry Gauthier-Villars. Fourteen years older than his wife, Willy was a best-selling novelist, well introduced into the intellectual and artistic parisian circles and known for his several love affairs.

From 1900 to 1903, four *Claudine* novels were published and signed by Willy: *Claudine à l'école* (*Claudine at school*, 1900)¹, *Claudine à Paris* (*Claudine in Paris*, 1901)², *Claudine en ménage* (*Claudine married*, 1903)³, *Claudine s'en va* (*Claudine and Annie*, 1903)⁴. In her autobiography entitled *Mes Apprentissages* (1936), Colette put these words in Catulle Mendès' mouth: according to the French poet and writer, she is the only « auteur des *Claudine* » [5]. As outlined by Stéphanie Michineau in *L'autofiction chez Colette*, Colette doesn't use the word "collaboration", and presents herself as the true author of the book series [5]. From 1948 to her death in 1954, Colette have continuously minimized the role played by Willy in the writing of the *Claudine* book series, until her name replaces the name of her first husband on the novels' cover. And when Colette was asked whether Willy had helped her to write these texts, she answered in 1949: « Plutôt par des indications, mais ça ne peut s'appeler une aide...⁵ ».

However it seems necessary to reconsider this position. First, Willy has never hidden the fact that Colette had been involved in the writing of these books, as she recognized it in *Mes Apprentissages*⁶. Besides, in an interview she gave in 1903, Colette explained that Willy « crie si obstinément ma participation [aux *Claudine*] que je dois m'incliner⁷ ». The use of the name "Willy" could also be explained by commercial reasons: Willy was a famous and successful author, and his name was a guarantee of success. Moreover, Willy could also have signed these books for social reasons: as a woman in 1900, Colette could have encountered many difficulties to publish a novel under her own name [5]. Not to mention that the methods of "literary production" used by Willy were perfectly known by his parisian colleagues and friends. In 1900, it is

¹ Willy, *Claudine à l'école*, Paris: P. Ollendorff, 1900.

² Willy, *Claudine à Paris*, Paris: P. Ollendorff, 1901.

³ Willy, *Claudine en ménage*, Paris: "Mercure de France", 1902

⁴ Willy, *Claudine s'en va*, Paris: P. Ollendorff, 1903

⁵ André Parinaud, *Colette, mes vérités, Entretiens avec André Parinaud*, Paris, Ecriture, 1996, p.71.

⁶ Colette, *Mes apprentissages, Oeuvres complètes*, Paris: Gallimard, 199, tome III, p.1026

⁷ Déclaration de Colette dans La Dépêche Républicaine de France-Comté en 1903, citée par Stéphanie Michineau dans *L'autofiction chez Colette...*, p.315.

well known that Willy employs many “secrétaires” or “collaborateurs” to write “his” books. Colette described their parisian apartment as a “literary workshop” [1], where several ghostwriters were writing novels, meticulously reviewed, rewritten and commented by Willy. During his literary career, it is estimated that around fifty collaborators have been employed by Willy [2]. According to François Caradec, the draft manuscript of *Claudine à l'école* has also been “treated” by this workshop; and after being reviewed by Willy, it would have been reduced by half [2].

Nevertheless we don't know exactly which part of the text has been written by Colette and which part has been rewritten (or even written) by Willy. If Colette, at the end of her life, alleged that she was the true and only author of these novels, Willy claimed that he has been fully involved in the writing of these books. By citing Willy, his third wife, Madeleine de Swarte, seems to directly speak to Colette when she writes about Willy's ghostwriters: “Leur orgueil est si grand que rien ne l'humilie. Si l'œuvre dont ils ont fourni l'esquisse – d'après mon plan – connaît le succès, chacun de ses praticiens répand le bruit qu'il en est l'unique auteur. Il arrive à le croire⁸”.

For researchers (and readers), it is difficult to unravel true from false. Therefore reading the *Claudine* books series under this perspective could open up new possibilities for the critical analysis of Colette's work. But some questions should be addressed first:

- In which extent has Willy influenced the first literary texts written by Colette?
- Which role has he really played in the writing of these novels?

To answer these questions, I will use the possibilities offered by text mining, and more specifically by authorship attribution methods. These methods can offer crucial elements to determine how the two writers influenced each other, and which texts (or parts of texts) have been written by Colette or Willy [3] [4].

In the framework of this conference, I propose to analyze *Claudine à l'école*, first novel of the book series. Reference corpora will also be used in order to deploy supervised learning methods. I will work on the 200 or 300 (or even 500) terms that are the most frequently employed by the two authors, or on the most commonly employed n-grams (character sequences). I will use R packages to analyze these text corpora [3], in particular with hierarchical clustering and neural network methods [4]. This analysis of the first *Claudine* novel constitutes the first part of a work that will be pursued with the application of the same methods to the three following novels of the series (*Claudine à Paris*, *Claudine en ménage*, *Claudine s'en va*). This study will allow to better understand the literary relationship developed and maintained by Colette and Willy, to shed a new perspective on the history of this intellectual collaboration, and to precisely describe how Colette made her first steps in literature.

2. REFERENCES

- [1] Boustani C. 2011. *Willy: le bonheur de l'imposture* Angers: Presses universitaires de Rennes. In: *L'imposture dans la littérature: Cahier XXXIV*. ISBN: 9782753547858. DOI=<http://books.openedition.org/pur/12172>.
- [2] Caradec F. 2004. *Willy. Le père des Claudine*, Paris: Fayard.
- [3] Eder M., Kestemont M. et Rybicki J. 2016. *Stylometry with R: a suite of tools*. Boston: R Foundation for Statistical Computing In: *The R Journal* Vol. 8/1. ISSN: 2073-4859
- [4] Jockers M. 2014. *Text Analysis with R for Students of Literature*. New York: Springer.
- [5] Michineau S. 2008. *L'autofiction chez Colette*, Paris: Publibook.

⁸ Madeleine de Swarte, *Les Fourberies de Papa*, 1906.

Gustave Roud, « Œuvres complètes ». Edizione digitale e RDF.

Daniel Maggetti, Elena Spadini, Marion Rivoal,
Loïc Jaouen, Julien Burri, Alessio Christen, Bruno Pellegrino

Université de Lausanne, Svizzera

ABSTRACT

Il poster vuole presentare il progetto di edizione, in corso, ‘Gustave Roud, *Œuvres complètes*’ dell’Università di Losanna (Centre de recherche sur les lettres romandes). In particolare, ci si concentrerà sulle scelte tecniche e sull’utilizzo dell’infrastruttura software ‘Knora’ (RDF, OWL, RESTful API).

PAROLE CHIAVE

Edizione digitale, RDF, XML, TEI, scholarly digital edition.

1. INTRODUZIONE

Poeta e attore culturale maggiore della Svizzera francofona, Gustave Roud (1897-1976) è autore di un’opera variegata, di cui si interessa in misura crescente la critica, letta e tradotta internazionalmente. In italiano, si segnalino le traduzioni *Requiem e altre prose poetiche* [1] e *Del camminare in pianura* [2].

Il progetto ‘Gustave Roud, *Œuvres complètes*’, diretto da Daniel Maggetti e sostenuto dal Fondo nazionale svizzero per la ricerca scientifica, prevede l’edizione critica dell’insieme della produzione dell’autore, che comporta diversi volumi di prose poetiche, appunti diaristici, raccolte di traduzioni, testi sulla letteratura e sull’arte. Il progetto prende le mosse dall’analisi di un ricco fondo d’archivio, conservato al *Centre de recherches sur les lettres romandes*, e comprende un’edizione cartacea e una piattaforma digitale.

La piattaforma, in corso di elaborazione, è concepita come accompagnamento genetico all’edizione e come complemento documentario e tematico. Essa è organizzata in due grandi sezioni: l’archivio (descrizione dei documenti conservati nel fondo) e l’edizione testuale, con necessari rimandi dall’una all’altra.

2. EDIZIONE DIGITALE E RDF

L’elemento di maggiore interesse, in questa sede, è l’utilizzo per il progetto dell’infrastruttura software *Knora* (Knowledge Organization, Representation, and Annotation) [3], sviluppata dal Digital Humanities Lab dell’Università di Basilea. In *Knora* i dati sono rappresentati in RDF e immagazzinati in un *triplestore*. Usando ontologie OWL, un’ontologia di livello superiore (base di *Knora*) è progettata per le esigenze generali dei dati di ricerca per le scienze umane, mentre, al livello del progetto, viene concepita un’ontologia che risponda alle caratteristiche specifiche. I dati vengono poi esposti tramite una API RESTful.

Il poster vuole dare conto dello stato di avanzamento del progetto e riflettere sulle soluzioni tecniche adottate. Si presenterà quindi l’ontologia realizzata per rappresentare i dati e le loro relazioni. Si esporranno i punti di maggiore criticità, come la modellazione dei dati che formano l’*avant-texte*; e si tratterà della compatibilità del modello di *Knora* con quello della XML-TEI.

Se il modello delle triple RDF è diffuso per la rappresentazione e gestione dei dati nell’ambito del patrimonio culturale, nel “sotto-ambito” dell’edizione dei testi risulta assai meno utilizzato. Si tratta infatti, in quest’ultimo caso, di rappresentare dati non strutturati e dunque difficilmente organizzabili in triple con valore semantico; per i testi, lo standard più diffuso è quello della TEI, strettamente legato alla tecnologia che gli dà forma, ovvero l’XML. L’integrazione tra RDF e TEI, al momento, passa attraverso due possibili canali, con esiti assai diversi. Il primo prevede la marcatura su più livelli di un testo, ovvero uno “strato” *embedded* di XML-TEI e un altro *stand-off* di RDF, dove quest’ultimo è utilizzato per annotare entità semantiche precipue che compaiono nell’opera; i due strati possono essere connessi, ovvero poggiare l’uno sull’altro. Progetti esemplari in tal senso sono l’edizione delle lettere di Vespasiano da Bisticci [4] e le edizioni che utilizzano il software di annotazione semantica Pundit [5]. La seconda possibilità è quella di separare l’insieme della marcatura — di tipo semantico e strutturale — dal testo, tramite meccanismi di *stand-off* [6][7], in cui le annotazioni (che possono o meno seguire il modello TEI) sono rappresentate in RDF. Quest’ultima opzione, in particolare, è in corso di valutazione nel progetto.

3. CONCLUSIONE

Il poster presenterà il progetto in corso e le soluzioni tecniche adottate, considerando le problematiche relative agli standard per la modellazione e la rappresentazione dei dati nell’ambito dell’edizione digitale.

4. BIBLIOGRAFIA

[1] Roud, G. *Requiem e altre prose poetiche*. Interlinea, Novara 2006.

[2] Roud, G. *Del camminare in pianura*. A. Dadò, Locarno 2014.

[3] Digital Humanities Lab, Università di Basilea. *Knora. Knowledge Organization, Representation, and Annotation*. <http://www.knora.org/>.

- [4] Tomasi, F. (a cura di). *Vespasiano da Bisticci, Lettere*. Edizione digitale 2013. Versione 2, 2015. CRR-MM, Università di Bologna. URI: <http://vespasianodabisticciletters.unibo.it/>. DOI: 10.6092/unibo/vespasianodabisticciletters. ISBN: 9788898010110.
- [5] Netseven. *Pundit*. <http://thepund.it/>
- [6] Schmidt, D. 'Using standoff properties for marking-up historical documents in the humanities'. *Information Technology* 58 (2016) 1.
- [7] Spadini, E. - Turska, M. 'XML-TEI standoff markup. One step beyond'. *Digital Philology: A Journal of medieval cultures* (forthcoming).

LETTERE: LETters Transcription Environment for REsearch

Giovanni Moretti, Rachele Sprugnoli, Sara Tonelli

Digital Humanities Research Group, Fondazione Bruno Kessler
Via Sommarive 18, 38123 Trento, Italy
{moretti,sprugnoli,satonelli}@fbk.eu

1. INTRODUCTION

The digitization of epistolaries is extremely important for the preservation and study of the cultural and historical patrimony of literary correspondence (Hankins, 2015). In recent years, several small and large-scale projects have been carried out: see, among others “Mapping the Republic of Letters”¹, “Vincent Van Gogh Letters” (Jansen et al., 2009) and the letters of Vespasiano da Bisticci (Tomasi, 2013). Many initiatives are based on collaborative work (Price, 2011) adopting a crowdsourcing approach (Estellés-Arolas, 2012) and using web-based transcription interfaces accessible also to laypeople: for example, “Letters of 1916”² uses Scripto (Leon, 2014) within its transcription workflow.

However, some projects need transcriptions made by domain experts so to have a wide and accurate historical apparatus; moreover, web-based interfaces are not usable when it is necessary to transcribe in archives that do not have Internet connections available to users. In this proposal we describe LETTERE (*LETters Transcription Environment for REsearch*) a new standalone, platform-independent application specifically designed for the transcription of correspondence in accordance with requirements defined by history scholars. The development of LETTERE is part of the National Edition of De Gasperi’s correspondence, a project launched in 2017 with the support of the Italian Ministry of Cultural Heritage and Tourism that, for the first time, funded a National Edition in digital format³. In the following, we give a brief overview of this project and we present the functionalities of the system together with future work.

2. PROJECT DESCRIPTION

While “public” documents written by De Gasperi - i.e. official documents, newspaper articles – have been collected, curated and edited (De Gasperi, 2006) also in digital format (Sprugnoli et al., 2016), less attention has been devoted to his correspondence. Published letters tend to cover only exchanges with specific persons (e.g. Sturzo and De Gasperi, 2006), or deal with specific topics (De Gasperi et al., 1970) and circumstances (De Gasperi, 1955). On the contrary, the National Edition project wants to provide an exhaustive analysis of De Gasperi’s correspondence, covering both his public and private life and giving visibility to letters never edited before. A preliminary registry of the letters to be digitized, carried out by a history scholar, lists more than 40 archives all over the world conserving about 2,500 letters written in multiple languages.

The project is coordinated by “Fondazione Trentina Alcide De Gasperi”⁴ and involves scholars from several Italian universities and institutions relying on an interdisciplinary Scientific Committee.

3. TOOL DESCRIPTION

A prerequisite for the acquisition of De Gasperi’s letters in digital format is the availability of a transcription tool that i) allows multiple scholars to transcribe letters in different archives at the same time, ii) includes a quality check step, iii) includes a set of metadata previously defined by the Scientific Committee. Since such a tool was not available, we implemented LETTERE, based on a transcription process that involves transcribers and supervisors: the former visit archives, make digital copies of the letters, upload the photos, compile metadata, transcribe and annotate the text using the tool. Then, as soon as an Internet connection is available, they send the work done in the archive to the central server, where the material will be stored in a database. The supervisors, employing the same environment, check the correctness of the transcription and write the critical apparatus of each letter. Once this review is finished, the work on the letter is considered completed and can be accessed online.

From the technical point of view, the software application is entirely written in Javascript/ECMAScript 6. The interface is made with the ReactJS framework that guarantees greater optimization and reusability of the code. The server-side of the application is written in Java and is implemented in a Tomcat application server.

The set of metadata foreseen by the project is shown in Fig.1: they cover several aspects of the letters (e.g. their physical description, archival condition, subject) in compliance with the Dublin Core standard and the standard for the cataloging of manuscripts in Italian libraries⁵. Some metadata are to be chosen from a drop-down menu, while others have an autocomplete functionality to reduce the risk of mistakes. The title of a letter is automatically created by combining the name and surname of the sender and of the receiver and the sending date: e.g. “*Da [Nome Cognome Mittente] a [Nome Cognome Destinataria], gg/mm/aaaa*”.

¹ <http://republicofletters.stanford.edu/>

² <http://letters1916.maynoothuniversity.ie/>

³ In 2017 also the National Edition in digital format of Aldo Moro’s writings was launched. In this case, however, no transcription of handwritten texts is foreseen.

⁴ <http://www.degasperitn.it/it/fondazione/>

⁵ https://manus.iccu.sbn.it/GUIDAsettembre_2014.pdf

Figure 1. Metadata insertion form.

Given that the transcribers will be historians with no background in digital text encoding, the tool interface has been designed so that an explicit knowledge of tag annotation is not required (see Fig.2). Transcribers have to report the original punctuation, paragraph division and spelling, including possible mistakes. The presence of illegible words, underlined parts, cancellations, mechanical gaps is to be signalled using specific buttons. Similarly, it is possible to add the content written within line-spacing, the correct/current spelling of a single term or a string of text, the expanded forms of abbreviations. Margin notes have to be transcribed as well in a specific field (called “postille”) specifying whether they were written by the author of the letter: the same set of annotation is available also for these notes. All this information, stored in a database, will provide to the final reader of the National Edition different views of the same letter: e.g. the original version, the one with the annotations, the one - more reader-friendly - showing the normalized forms of abbreviations and mistakes.

Figure 2. Transcription and annotation interface.

4. FUTURE WORK

In the next months, LETTERE will be tested with a first run of transcriptions and made available to the project Consortium in the first half of 2018. Then, the tool will be then made publicly available for the research community with a Creative

Commons 4.0 license. In parallel, a navigation system linked to the project website⁶ will be implemented, providing text processing, data visualisation and search functionalities to the transcribed letters. From this latter system, letters will be available for download as raw text, TEI and pdf.

5. BIBLIOGRAPHY

- [1] De Gasperi, Alcide. 1955. *Lettere dalla prigione*. Milano: Mondadori.
- [2] De Gasperi, Alcide. 2006. *Scritti e discorsi politici*. Bologna: Il Mulino.
- [3] De Gasperi, Alcide. 1970. *Lettere sul concordato*. Brescia: Morcelliana. Estellés-Arolas, Enrique, and Fernando González-Ladrón-de-Guevara. 2012. Towards an integrated crowdsourcing definition. *Journal of Information science* 38, no. 2 (2012): 189-200.
- [4] Hankins, Gabriel. 2015. Correspondence: Theory, Practice, and Horizons. In *Literary Studies in the Digital Age*, edited by Kenneth M. Price and Ray Siemens. Available online: <https://dlsanthology.mla.hcommons.org/correspondence-theory-practice-and-horizons/>
- [5] Jansen, Leo, Hans Luijten, Nienke Bakker (eds.). 2009. *Vincent van Gogh - The Letters*. Version: December 2010. Amsterdam & The Hague: Van Gogh Museum & Huygens ING. <http://vangoghletters.org>
- [6] Leon, Sharon M. 2014. Build, Analyse, and Generalise: Community Transcription of the Papers of the War Department and the Development of Scripto. *Crowdsourcing Our Cultural Heritage* (2014): 88-112.
- [7] Price, Kenneth M. 2011. Collaborative Work and the Conditions for American Literary Scholarship in a Digital Age. *The American Literature Scholar in the Digital Age*: 9-26.
- [8] Sprugnoli, Rachele, Giovanni Moretti, Sara Tonelli and Stefano Menini. 2016. Fifty years of European History through the Lens of Computational Linguistics: The De Gasperi Project. *Italian Journal of Computational Linguistics* 2, no. 2 (2016): 89-100.
- [9] Sturzo, Luigi and De Gasperi, Alcide. 2006. *Luigi Sturzo - Alcide De Gasperi: carteggio, 1920-1953* (Vol. 4). Rubbettino Editore.
- [10] Tomasi, Francesca. 2013. L'edizione digitale e la rappresentazione della conoscenza. Un esempio: Vespasiano da Bisticci e le sue lettere. In *ECDOTICA*, 2013, 9, pp. 264 - 286.

⁶ <http://www.epistolariodegasperi.it/>

Modelling between digital and humanities. A project, its research questions and its outcomes

Cristina Marras¹, Arianna Ciula², Øyvind Eide³, Patrick Sahle³

¹ Institute for European Intellectual Lexicon and History of Ideas, National Research Council (Rome)

² King's College Lab (London)

³ University of Cologne

KEYWORDS

modelling, digital humanities, interdisciplinarity, text, languages

1. RATIONALE

By modelling we intend the creative process by which researchers create and manipulate external representations to make sense of the conceptual objects and phenomena they study. This poster presents the rationale behind as well as the work undertaken by the international collaborative project *Modelling between digital and humanities: thinking in practice*, funded by the Volkswagen Foundation (<http://modellingdh.eu/>).

2. WORK HYPOTHESIS/RESEARCH QUESTIONS

Our working hypothesis is that in DH research, implicit and explicit models of cultural objects and phenomena are integrated into external metamodels which often embed both natural languages and formal languages. Two modelling contexts are used as case studies to reflect on modelling in practical terms: Textuality and Events. Moreover, the analysis was guided by the following questions: How are theory and practice blended in the modelling efforts?; What shared terminology can help us gaining an integrative and non-reductive understanding of digital modelling?; Can we describe the methods of digital modelling informed by such an integrative and non-reductive approach?

3. ARGUMENT

In our poster we stress the importance of integrating diverse theories around modelling (i.e. [1], [2], [3], [4], [5]) with a practical dimension. We make use of DH as an interdisciplinary departure to study modelling as anchored both to computer science and to the humanities ([6], [7], [8]), linking scholarly modelling as a formal and informal reasoning strategy across disciplinary boundaries (e.g. from philology to literary theory, from information studies to cultural heritage documentation). Furthermore, building on the novelty of DH research in making explicit and integrating existing diverse models of cultural objects and phenomena the poster, providing some examples, presents: a. diverse possibilities for a new interdisciplinary language around modelling; b. modelling in scholarship as a process of signification.

4. SCOPE

Our scope is to retrieve 1. the polysemic dimension of the concept of model and modelling, 2. To grasp the dynamics of interaction between model as expression (i.e. shared forms of representations of knowledge) and model as content (i.e. shared data models conceptualising and formalising a domain of knowledge) in specific research contexts. We also advocates for a critical DH research offering the instruments to unpack the rhetoric of digital models including data models, so as to act at the core of a new cultural literacy. Presenting a poster at AIUCD 2018 would give the authors the opportunity to discuss the activities and outcomes of the project also with the Italian community with the aim to collect interest from conference delegates possibly also in joining an advisory group for further development of some of the project's ideas.

5. REFERENCES

- [1] McCarty, W. 2005. *Humanities Computing*. Basingstoke [England]; New York: Palgrave Macmillan.
- [2] Mahr, B. 2009. "Information science and the logic of models". *Software & Systems Modeling*, 8: 365–383.
- [3] Morgan, M.S. 2012. *The world in the model: how economists work and think*. Cambridge.
- [4] Kralemann, B., Lattmann, C. 2013. Models as icons: modeling models in the semiotic framework of Peirce's theory of signs". *Synthese* 190.16: 3397–3420.
- [5] Flanders, J., Jannidis, F. 2015. *Knowledge Organization and Data Modeling in the Humanities*. [White Paper]. <https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/11127>
- [6] Sahle, P. 2013. *Digitale Editionsformen - Teil 3: Textbegriffe und Recodierung*. Norderstedt.
- [7] Ciula, A., C. Marras. 2016. "Circling around texts and language: towards "pragmatic modelling". *DHQ*, 10, no. 3, <http://www.digitalhumanities.org/dhq/vol/10/3/000258/000258.html>
- [8] Ciula, A., Eide, O., 2017. "Modelling in digital humanities: Signs in context". *Digital Scholarship in the Humanities*, 32, no. Suppl. 1: i33-i46. doi:10.1093/llc/fqw045.

On Developing a Virtual Museum of Greek Immigration and Language in Canada¹

Angela Ralli¹, Charalambos Tsimpouris², Christos Papanagiotou³, Tonia Tzanavara⁴

¹University of Patras, Greece - ralli@upatras.gr

²University of Patras, Greece - xtsimpouris@upatras.gr

³University of Patras, Greece - cpapanag@yahoo.gr

⁴Municipal Art Gallery of Korinthos - toniatzan@hotmail.com

ABSTRACT

In this paper, we present a digital museum that is being developed in the context of the project ‘Immigration and Language in Canada: Greeks and Greek-Canadians’, sponsored by the Niarchos Foundation. The project involves research teams from four universities, one Greek, University of Patras, and three Canadian, McGill (Montreal), York University (Toronto) and Simon Fraser University (Vancouver). It concerns the first-generation immigrants, mainly of the period 1945-1975, which has seen the bulk of Greek immigration in Canada. The museum will present various aspects of the immigrants’ experiences, laying emphasis on information linked to the reasons for immigrating, the points of departure and arrival, the process of integration in the Canadian environment, the way their mother tongue has been affected, the organization of Greek communities across Canada, labour and working places, their religious life, school education, and lastly the participation in various public events.

For the development of the museum, state-of-the art technologies (HTML5, JavaScript, CSS3, jQuery, AJAX) and a JavaScript 3D-engine on the basis of *babylon.js* are being used in order to create a virtual environment within the client’s browser, thus providing the user with a friendly experience so as to gain access to exhibits via virtual navigation. The material presented is chosen from a platform for the backend database and a repository, which were previously constructed as a prerequisite. It consists of oral data collected from available written sources and via interviews conducted in Canadian provinces which have massively accepted Greek immigrants. Having been constructed with Drupal CMS (v.7), an open source content management system (CMS) that is actively supported by thousands of programmers, the database consists of the following ontologies: *sources*, *field researches*, *linguistic varieties*, *places*, *informants* that all connect to the main ontology of *media files*.

KEYWORDS

Virtual museum, immigration, social history, language, Greek Canadians, database.

1. DEVELOPING A VIRTUAL MUSEUM FOR GREEK IMMIGRATION

The project “Immigration and Language in Canada: Greeks and Greek-Canadians” is the first attempt ever made to create a virtual museum about Greek expatriates, contributing to the general study of Greek transatlantic immigration and the understanding of ethnic diversity in the Canadian society. Research groups from four Universities, one Greek, University of Patras, and three Canadian, McGill (Montreal), York University (Toronto) and Simon Fraser University (Vancouver), collaborate in this newly-undertaken research program (January 2017-December 2018), which has been funded by the Niarchos Foundation and aims at exploring the social history and language of first-generation Greek immigrants in Canada, between the years 1945 and 1975 which have seen the bulk of Greek immigration, for various reasons, political, economic, personal, etc.

The project attempts to move a step beyond the standard practice in linguistic and historical research by producing ground-breaking deliverables in connection with the development of digital humanities that will ensure the sustainability of the results. In this vein, oral (recordings of narratives) and written material are collected in several Canadian provinces, where there are big Greek communities. So far, there are about 210 hours of recordings of narratives, the collection of which has been made by the three Canadian universities, following ethnographic methods [6]. The oral material is complemented by photographs of private and public immigrant life, and written data drawn from the available sources, such as documents (e.g. passports), dissertations (among others, [1], [3], [5]), books describing community life (among others, [1], [2], [4]) and extracts of newspapers. The collected material has been stored in an electronic depository and new technologies are being used for its processing in order to make the data easily accessible to anyone interested in exploring the different historical, social and linguistic aspects of Greek-Canadian communities. The deliverables of the project aspire to raise public interest in GreekCanadian history and language contact between Greek and English, provide a long-lasting point of reference for educational and social purposes, and shed light on the connection of Greek immigrants to the social and cultural history of the host country, that is, Canada.

The main deliverable of the project is a Virtual Museum, which is being designed by the research group of the *Laboratory of Modern Greek Dialects* (LMGD) of the University of Patras (www.lmgd.philology.upatras.gr). It consists of an online portal encompassing the Canadian-Greek immigrant experience from Greece to Canada, the content of which is based on

¹ The authors are much obliged to their colleagues and partners of the project, Dr. Tassos Anastassiadis (McGill), Dr. Sakis Gekas (York) and Dr. Panayotis Pappas (SFU) for their most precious help into providing oral and written material regarding the life and language of firstgeneration Greek immigrants in Canada.

interdisciplinarity among historical, sociological and linguistic research conducted by the four universities. For this purpose, as a prerequisite, a repository and a platform for the backend database were previously constructed by LMGD, which help the researchers of the four collaborating universities to introduce manually the material under investigation and classify a list of their properties concerning data and metadata. Drupal CMS was used, version 7, an open source Content Management System employed by thousands of programmers. The database was developed on the basis of the following ontologies: *sources, field researches, linguistic varieties, places, informants* that all connect to the main ontology of *media files*. In particular, the collected material is categorized as follows:

- (a) old photos and pictures (jpg/pdf);
- (b) raw narratives/interviews (wav/flac);
- (c) fully transcribed interviews in Praat and LaTeX²/pdf;
- (d) translated narratives in English (LaTeX/pdf);
- (e) videos (wav);
- (f) digitized newspapers and/or magazines of the decades under examination (jpg/pdf).

For the construction of the Virtual Museum on Greek Immigration in Canada, thirty-five virtual museums from all over the World have been studied, among which the following have been considered as the most appropriate ones for serving as models to the museum under development:

The *British Migration Museum* (<http://migrationmuseum.org>), the *Australian Immigration Museum* (<https://museumvictoria.com.au/immigrationmuseum/>), the *Virtual Museum of Canada* (<http://www.virtualmuseum.ca/home/>) and the *Virtual Museum of Italian immigration in the Illawarra* (Australia). (<http://vmiii.com.au/>).

In its present version, the Virtual Museum on Greek Immigration in Canada, consists of: (a) an atrium containing a short documentary and two maps, one of Greece and one of Canada with (blinking) towns and villages which represent the points of departure and arrival, respectively; (b) 10 rooms corresponding to 10 major themes that are determined by the interviews and generally the collected material. They comprise: 1) A brief overview of Greece and Canada in the '40s and '50s; 2) Reasons for immigrating - The trip; 3) Arrival and settlement; 4) Labour - Working places; 5) Family life; 6) Community life; 7) Schooling - education; 8) Church - religious life; 9) Language; 10) Leisure - Events.

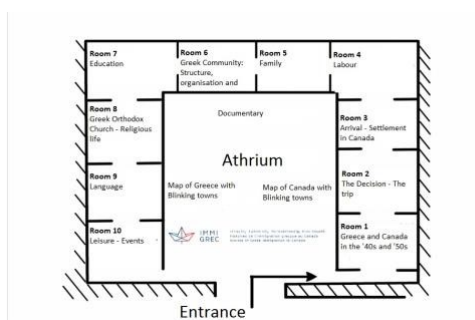


Figure 1: A top view of the virtual museum on Greek immigration in Canada

For its development, state-of-the art technologies are being used (HTML5, JavaScript, CSS3, jQuery³, AJAX) and a JavaScript 3D-engine is employed on the basis of *babylon.js*⁴ in order to create a virtual environment within the client's browser, thus providing the user with a friendly experience so as to gain access to exhibits via virtual navigation.

The pioneering nature of the museum lies in the fact that it is an attempt to portray digitally, yet palpably, the degree of adaptation of Greek immigrants in their new country, the preservation and creative use or rejection of socio-cultural and linguistic identity traits, and how the Greek language and its dialectal variation have evolved in a language-contact situation, where English and/or French is the donor language and Greek the recipient. Interestingly, it constitutes the first attempt to conduct a thorough and systematic research on the history and language contact of Greek immigrants in Canada on the basis of data drawn from their life in an area geographically remote from Greece.

2. REFERENCES

- [1] Aravossitas, Th. 2016. *The hidden schools: Mapping Greek heritage language in Canada*. PhD Thesis, University of Toronto.
- [2] Florakas-Petsalis, S. 2013. *To build the dream. The story of the early Greek immigrants in Montreal*. Montreal: K-Litho Printing (3rd ed.)
- [3] Maniakas, Th. 1990. *The ethnolinguistic reality of Montreal Greeks*. Ph. D. Thesis. University of Montreal.
- [4] Mouratidis, Ch. 2016. *Osa den ezvise o chronos* [What time has not erased]. Toronto: Greek Community.

² For easiness and saving time, LaTeX, a familiar tool to the researchers of all teams was used for the transcription of the interviews, although it is not easily amenable to modification by the user. If necessary, LaTeX will be converted to the Standard Generalized Markup Language (SGML), specifically to the SGML tag set created by the Text Encoding Initiative (TEI), since the structural similarities of the TEI and LaTeX schemes allow for a relatively straightforward conversion.

³ <https://jquery.com/>

⁴ <https://www.babylonjs.com/>

- [5] Dounia, M. 2004. *Your roots will be here, away from your home: migration of Greek women to Montreal 1950-1980*. Ph.D. Thesis. McGill University.
- [6] Papazachariou, D. and Archakis, A. 2003. Ethnografikos prosdiorismos dedomenon apo neanikes sinomilies [Ethnographic determination of data from youth conversations]. *Studies on Greek Linguistics* 23: 289-300.

Physical methods in cuneiform tablet classification - a possibility for a new narrative

Jaroslav Valach¹, Petra Štefcová²

¹Institute of Theoretical and Applied Mechanics of AS CR, Czech Republic - valach@itam.cas.cz

²National Museum, Prague, Czech Republic - petra_stefcova@nm.cz

ABSTRACT

The paper presents a new possibilities of research work based on availability of content rich database combining knowledge from humanities and science domains. The database is developed as one goal of a research project that aims at design and creation of systems of documentation and analysis of artifacts based on the utilization of computer-aided visualization and documentary methods. Some possibilities of new type of queries encompassing physics and humanities are presented and discussed.

KEYWORDS

Digitalization of collections; cuneiform tablets; database system; digital humanities.

1. INTRODUCTION

In contrast to data recorded on CD ROMs, which become obsolete after a few decades of existence, messages written on clay tablets can pass their content to reader at the same level of legibility as millennia ago. Because of durability of the medium, there are hundreds of thousands or even millions of clay tablets preserved in museum collections around the world [1]. Their content, historical and archeological context is often known, but their materiality remains scientifically unexplored or left aside.

This paper demonstrates how a new kind of scientific queries can be posed once synergy of both approaches is achieved via multidomain knowledge system coupling methods of natural sciences and humanities.

2. OBJECTS, METHODS AND DATA

The aim of the research project, started in 2016, is to develop multidisciplinary knowledge system allowing cross-domain synergies. The proposed project of clear interdisciplinary character uses the tools of semantic, visual, spatial and componential documentation of artifacts to obtain maximum of information usable in restoration, conservation and last but not least in research. Its goal is also to demonstrate functionality of these tools on the task of digitalization of thematic groups of national heritage objects. The digital models are supplemented by information acquired by non-destructive methods of analysis (e.g., information on the inner structure, topography and appearance of the surface and its elemental composition), which is stored in a database and annotated with metainformation.

The system is built around the collection of cuneiforms tablets. The collection of cuneiforms tablets originating from early 20th century excavation of prof. Hrozný near Kültepe in Turkey [2, 3] consist of slightly over four hundred tablets. This is the basis for multidisciplinary database integrating existing and developed knowledge based on combined approaches of physics, computer science, linguistics, and philology (including cuneiform paleography). At the present time the collection is being digitalized and multiple tools are tested in the field of digitalization, character recognition, search and retrieval. Every cuneiform tablet and/or envelope is represented by a separate record. The digital record comprise of the three major parts: the first is a paleographic database concentrated on signs, linguistic and historic data containing transcripts, corpus of texts [4], translations and historical data; the 3D digital model [5] (and presenting data in virtual and augmented reality) or CT scan of the tablet are the second part and a third part of the record is based on data from various physical methods like X-ray fluorescence [6] or colorimetry.

3. DISCUSSION AND CONCLUSIONS

The following examples of multidisciplinary queries are presented in order to demonstrate new opportunities for research in digital humanities extended by contributions from other scientific fields in the case of cuneiform tablets collection.

"Is there a relation between tablet treatment (drying only or firing its clay) and importance of its content?" – a question which expects understanding of the script as well as the physical state of the tablet's material assessed by colorimetry in this case.

"Is there any relation between the provenience and the content or, is there any relation between the provenience and a particular scribe?" – involving XRF analysis, grouping tablets by similar composition and handwriting, and searching for correlation in content.

Therefore, "Can special variants of used signs in the writing be attributed to a particular provenience determined by XRF yielding a material composition of the clay?" or "can we recognize individual scribes and/or scribal schools?"

"Is an exception in physical characteristic of the tablet a good indicator also for exceptional textual content?", "Are there any typical features for individual textual genres?", or even "Can some features of the tablets reveal preferences of individual scribes?" – if proved correct, the focused research on the most interesting cuneiform tablets can be applied also on much larger collections, helping to automate the selection.

To conclude, these and similar queries can facilitate some new research possibilities. Rich data records enable to pose and answer queries otherwise impossible to put as the answer assumes simultaneous knowledge from humanities and science domains, which are usually kept separate.

4. ACKNOWLEDGEMENTS

The presented work is carried out within project “Analysis, description and archivation of aggregate information on properties of cultural heritage artifacts and usage of such data in restoration, conservation and research” is supported by the program of applied research and development of national and cultural identity (NAKI) of the Ministry of Culture of the Czech Republic – grant No. DG16P02M022. Authors are also grateful to Dr. J. Mynářová from Faculty of Arts of Charles University in Prague for her invaluable comments.

5. REFERENCES

- [1] Birmingham: A Collaborative Environment for Assisted 3D Reconstruction of Cuneiform Tablets (visited 2017 Feb 10, available from: <http://www.birmingham.ac.uk/research/activity/eese/ise/hit-team/heritage/cuneiform/index.aspx>).
- [2] Klíma, J. 1953. In memoriam Bedřich Hrozný. *Journal of Juristic Papyrology*, VII/VIII, 1953–1954.
- [3] Larsen, M.T. 2015. *Ancient Kanesh: A Merchant Colony in Bronze Age Anatolia*. Cambridge University Press, New York.
- [4] ORACC: Open Richly Annotated Cuneiform Corpus (visited 2017 Feb 10, available from: <http://oracc.museum.upenn.edu/index.html>).
- [5] Valach, J. et al. Digitising 3D surfaces of museum objects using photometric stereo-device. In Bentkowska-Kafel, A.; Murphy, O. (ed.). *COSCH e-Bulletin Proceedings of the colour and space in cultural heritage session at the Denkmäler 3D conference. From lowcost to high-tech. 3D-documentation in archaeology and monument preservation Vol. 1*. Dortmund: The LWL Industrial Museum Zeche Zollern, 2014. ISSN 2409-9503.
- [6] Artax XRF (visited 2017 Feb 10, available from: <https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/microxrf-and-txrf/artax/overview.html>).

Pinte di storia: un progetto di Public History su Youtube

Domenico Matteo Frisone¹, Antonio Iodice², Michele Lacriola³,
Edoardo Nicoletti⁴, Pietro Rubini⁵

¹ Università degli studi di Teramo, PhD student in Storia Medievale, Italia - matteofrisone@gmail.com

² Università degli studi di Napoli "Federico II", PhD student in Storia Moderna, Italia - antonio.iodice@etu.univ-amu.fr

³ Università degli studi di Salerno, PhD student in Storia Contemporanea, Italia - michelelacriola90@gmail.it

⁴ EidoLab a.p.s., Senior Multimedia Manager, Italia – eidolab@eidolab.com

⁵ EidoLab a.p.s., consulente, Italia – piero.rubini346@gmail.com

ABSTRACT

Il web, con tutte le possibilità che contiene e che continua a sviluppare, è il più giovane strumento di divulgazione a disposizione degli storici italiani e non. Il pubblico raggiungibile per mezzo del web è potenzialmente infinito e fornisce vari mezzi per la fruizione del sapere, dalla lettura di siti web alla visione di video. È su quest'ultimo mezzo che ci vogliamo soffermare, che come piattaforma di riferimento ha principalmente YouTube.

In particolare, YouTube Italia è un terreno quasi vergine, contrariamente a quanto avviene nel resto del mondo. Sulla versione italiana della piattaforma, infatti, ad abbondare sono i due estremi della divulgazione storica: canali di diffusione di autorità accademiche, quali università, istituti di cultura e centri studi o, al contrario, canali dedicati alla "controinformazione" storica, che spesso nascondono un revisionismo populista e raramente accompagnato da una logica riflessione sulle fonti. Il panorama estero è, invece, più variegato. Certamente non mancano esempi comparabili a quelli prima citati, ma sono presenti, allo stesso tempo, prodotti destinati al pubblico medio di Youtube (o comunque di social che prevedano l'implementazione video del sito). Questo tipo di contenuti vede affiancate produzioni pubbliche ma anche produzioni private.

Il progetto Pinte di Storia, avviato da un gruppo di laureati e dottorandi in storia (formati in differenti università) si occupa di **divulgazione storica** "per tutti i palati", per **raccontare la Storia attraverso delle storie**. Postosi come obiettivo la semplificazione, senza banalizzazione, delle tematiche storiche, il progetto tenta di attirare il pubblico medio di Youtube usando un linguaggio meno formale e forse più atipico, cercando spunti meno battuti e più particolari. Il comparto tecnico-multimediale vede l'ausilio di un Senior Multimedia Manager, esperto in comunicazione, operante nel settore da oltre trent'anni, grazie al quale l'avvio del progetto è stato possibile.

PAROLE CHIAVE

Public History, YouTube, Divulgazione Storica

1. INTRODUZIONE

Il web, con tutte le possibilità che contiene e che continua a sviluppare, è il più giovane strumento di divulgazione a disposizione degli storici italiani e non. I social network sono ormai parte integrante della vita di tutti i giorni, molto più della televisione. Il pubblico raggiungibile per mezzo del web è potenzialmente infinito e fornisce vari mezzi per la fruizione del sapere, dalla lettura di siti web alla visione di video. È su quest'ultimo mezzo che ci vogliamo soffermare, che come piattaforma di riferimento ha principalmente YouTube.

2. LA STORIA SU YOUTUBE

In particolare, YouTube Italia è un terreno quasi vergine, contrariamente a quanto avviene nel resto del mondo. Sulla versione italiana della piattaforma, infatti, ad abbondare sono i due estremi della divulgazione storica: canali di diffusione di autorità accademiche, quali università, istituti di cultura e centri studi o, al contrario, canali dedicati alla "controinformazione" storica, che spesso nascondono un revisionismo populista e raramente accompagnato da una logica riflessione sulle fonti. Il panorama estero è, invece, più variegato. Certamente non mancano esempi comparabili a quelli prima citati, ma sono presenti, allo stesso tempo, prodotti destinati al pubblico medio di Youtube (o comunque di social che prevedano l'implementazione video del sito). Questo tipo di contenuti vede affiancate produzioni pubbliche (come il canale istituzionale della English Heritage) ma anche produzioni private (come il canadese *Historia Civilis* o il francese *Nota Bene*). All'estero, addirittura, questa è diventata una vera e propria opportunità lavorativa.

3. AD OGNI PIATTAFORMA IL SUO LINGUAGGIO

Tuttavia, come comunicare al meglio in un ambiente come YouTube? Come John Hartley ha evidenziato in un recente contributo sul libro *YouTube. Digital media and society series*, non ci sono regole o vie precise, in quanto su questa piattaforma vige il "learning by doing model", si impara solo facendo. La piattaforma, infatti, non cura "l'alfabetizzazione digitale" dell'utente, che di fatto è lasciato in balia di sé stesso ed è costretto ad apprendere esclusivamente per mezzo dell'imitazione del lavoro altrui. Questo stato di cose ha fatto sì che, per molto tempo, quando si cercava di fare divulgazione storica sfruttando le risorse di questa piattaforma, semplicemente si riproponevano format inadatti, quali la lezione frontale, la diretta video di conferenze o anche la riproposizione di prodotti di qualità, sì, ma sviluppati per la

televisione, che ovviamente non avevano lo stesso impatto. Semplicemente, dunque, non esisteva ancora un format che sfruttasse le diverse modalità di comunicazione di YouTube. Quando, poi, i primi canali dedicati alla divulgazione storica hanno mosso i primi passi, non cercando di riproporre modalità prese da altri contesti, ma sviluppandone una adatta al mezzo adoperato ed al tipo di pubblico cui si rivolgeva, di fatto YouTube ha dimostrato di poter essere un potente strumento di divulgazione (come i casi esteri prima riportati dimostrano) e soprattutto il mondo accademico americano ed anglosassone ha iniziato a studiare le sue possibili applicazioni nella diffusione del sapere, partendo da casi studio di canali già affermati. Contrariamente a quanto avviene con altre forme di divulgazione, su YouTube il rapporto con gli utenti è di primaria importanza. Il feedback, ad esempio, è immediato e consente (o costringe) di cambiare in corso d'opera le modalità utilizzate nel canale. Inoltre, venendo meno la barriera (fisica o emotiva che sia) tra chi parla e chi guarda, spesso si verifica un processo di "bottom up", ovvero è la base, il pubblico cui ci si rivolge, che chiede di conoscere determinate cose o no, il che vuol dire che quest'ultimo è complice della produzione, perché coi suoi commenti indirizza in un senso o nell'altro la produzione video del canale.

4. PINTE DI STORIA

Da tutte queste premesse è nato, quasi un anno fa, il progetto Pinte di Storia, avviato da un gruppo di laureati e dottorandi in storia (formati in differenti università) sotto il motto di *Ogni bella storia inizia con una birra*. Questo detto tedesco ben racchiude lo spirito di questo format multiplatforma che si occupa di **divulgazione storica** "per tutti i palati", per **raccontare la Storia attraverso delle storie**. Postosi come obiettivo la semplificazione, senza banalizzazione, delle tematiche storiche, il progetto tenta di attirare il pubblico medio di Youtube (formato in maggior parte da ragazzi nella fascia d'età 15-30) usando un linguaggio meno formale e forse più atipico, cercando spunti meno battuti e più particolari. Restando metodologicamente fedeli ai propri studi, il gruppo verifica ogni informazione presente nei video, limitandone la durata ad un formato che catturi l'attenzione senza sacrificare, però, il possibile approfondimento: viene infatti consigliata una breve lista bibliografica e/o sitografica alla fine di ogni episodio, così da fornire un primo appiglio alla eventuale curiosità di ogni utente. Il comparto tecnico-multimediale vede l'ausilio di un Senior Multimedia Manager, esperto in comunicazione, operante nel settore da oltre trent'anni, grazie al quale l'avvio del progetto è stato possibile.

5. LEARNING BY DOING

A proposito del "learning by doing model" cui abbiamo fatto cenno, anche il nostro progetto è soggetto alla stessa logica: tenendo costantemente d'occhio le statistiche, i gusti ed i pareri degli utenti, aggiorna costantemente il proprio modo di operare, cercando di creare di volta in volta un prodotto video di qualità e al tempo stesso godibile da parte di un pubblico il più ampio e variegato possibile. Mancando però di una base di utenti preconstituita e non avendo precedenti cui rifarsi, Pinte di Storia è un progetto ben lungi dall'aver raggiunto una forma stabile. Le opzioni che il web offre sono molteplici e probabilmente molti dei tentativi che si proveranno a sviluppare saranno destinati a non incontrare i gusti dell'utenza, eppure solo scendendo nell'arena digitale e traendo insegnamento da fallimenti ed esperienze di divulgazione digitale si possono iniziare a trarre delle prime conclusioni sul modo migliore di operare. La strada è ancora lunga, ma bisogna pure che si inizi a percorrerla.

6. CONCLUSIONI

Quale, quindi, il nostro apporto in una eventuale conferenza? Primariamente, la trasmissione della nostra esperienza. Una base di partenza per entrare in contatto con un mondo in cui lo storico, oltre ad occuparsi del vaglio delle fonti, deve necessariamente trasformarsi in un esperto della piattaforma digitale a cui fa riferimento, ma ancora di più del pubblico che tramite essa può raggiungere. Dopo aver eseguito una disamina di tutte le problematiche del caso, il nostro intervento si propone di descrivere punti di forza e di debolezza della nostra ormai annuale sperimentazione, oltre a mettere in luce i prossimi punti per un rafforzamento del progetto.

7. BIBLIOGRAFIA DI RIFERIMENTO

- [1] Armitage, D., Guldi, J., 2016. *Manifesto per la Storia*, Roma: Donzelli.
- [2] Burgess, J., Green, J., 2009. *YouTube. Digital media and society series*, Cambridge: Polity Press.
- [3] Farnetti, P., B., Bertuccelli, L., Botti, A., 2017. *Public History. Discussioni e pratiche*, Sesto S. Giovanni: Mimesis.
- [4] Godwin, H., T., Khan, M., Yellowless, P., 2017. *The Educational Potential of YouTube*, in: P. Acad Psychiatry. DOI=<https://doi.org/10.1007/s40596-017-0809-y>.
- [5] Kapell, M., W., Elliot, A., B., R., 2013. *Playing with the Past. Digital Games and the simulation of History*, London: Bloomsbury.
- [6] Kean, H., Martin, P., 2013. *The Public History Reader*, Oxon: Routledge.
- [7] Luzzatto, S., 2010. *Prima lezione di metodo storico*, Bari: Laterza.
- [8] Makhortyk, M., 2017. *Remediating the Past: YouTube and the Second World War memory in Ukraine and Russia*, in: Memory Studies, 1-16. DOI=<https://doi.org/10.1177/1750698017730867>.
- [9] Sayer, F., 2016. *Public History. A practical guide*, London: Bloomsbury.

Power of Algorithms for Cultural Heritage Classification: The Case of Slovenian Hayracks

Ajda Pretnar¹, Lan Žagar², Blaž Zupan³, Dan Podjed⁴

¹ Faculty of Arts, Slovenia - ajda.pretnar@fri.uni-lj.si

² Faculty of Computer and Information Science, Slovenia - lan.zagar@fri.uni-lj.si

³ Faculty of Computer and Information Science, Slovenia - blaz.zupan@fri.uni-lj.si

⁴ ZRC SAZU, Slovenia - dan.podjed@zrc-sazu.si

ABSTRACT

Machine learning holds a lot of untapped potential for cultural heritage research. In this paper, we have used predictive modeling and deep models to classify hayracks from images. We have taken two competing typologies and proved that both work well for hayrack classification. However, Hazler's binary typology results in predictive models with more human-like misclassifications and is hence better for practical application. We show that machine learning methods can offer new perspectives to typology delineation and that predictive models can be used for automatic labeling of hayrack images.

KEYWORDS

computational anthropology, machine learning, deep models, hayracks

1. INTRODUCTION

Hayrack is a traditional Slovenian building, which lacks a unanimous typology. Several authors propose different typologies, which often depend on the interpretation of hayrack genesis and locality. These typologies have various levels of complexity, from binary to multi-level hierarchical classification.

We have developed two predictive models for hayrack classification, each inferred from 701 images labeled according to two different typologies. Both models can effectively predict hayrack type, while the study of their misclassification uncovers differences in classification quality. According to our best knowledge, the research presented in this paper is one of the first applications of machine learning for studying material culture.

2. THE DATA

2.1 Hayrack

A hayrack is a free-standing permanent, mainly wooden, vertical, open but roofed building for drying and storing hay and grain [4]. It is found mainly in the ethnic territory of Slovenes and it is a part of Slovenian national cultural heritage.

Between 1948 and 1962 Boris Orel, the then director of Slovenian Ethnographic Museum (SEM), and his team began documenting rural culture in the Slovenian countryside. Several expeditions collected a vast amount of data from around the country. The data consisted of transcriptions, drawings, and photographs, which the museum digitized in the past couple of years and published on their website [7]. A part of the collected data consisted of hayrack images found in different regions of Slovenia.

2.2 Hayrack Typologies

Generally, building typologies may be functional, technical, aesthetic, etc., all of which aim to establish a logical framework to architectural design and construction activities [5]. As hayrack typologies combine all elements into one classification system, several parallel typologies have emerged.

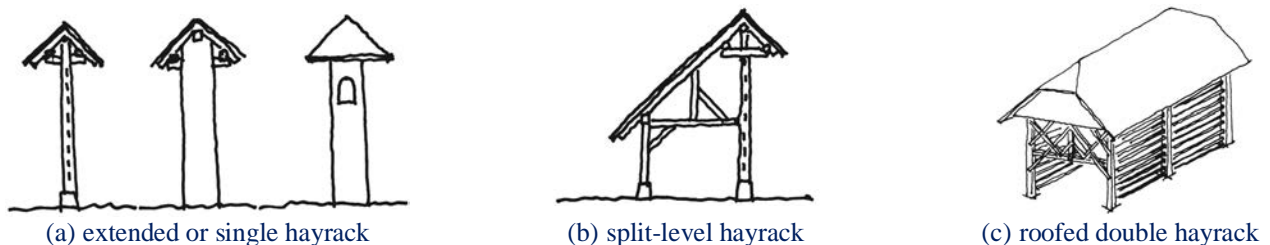


Figure 1: Hayrack variety. Sketches taken from Juvanec [4]

In broad, there are two main hayrack typologies. Melik [6] defines extended, roofed double and split-level hayracks as separate types (Figure 1). Hazler [3], on the other hand, defines only two main types, single and double hayrack. Interpretation depends on whether we see split-level hayrack as a version of a single/double hayrack or an independent type.

2.3 Images

We retrieved 126 black and white images of hayracks with the corresponding metadata from the SEM digital archive. Then, we followed up on previous expeditions by augmenting images of hayracks from Orel’s field team with contemporary images of hayracks from the same region. We visited 4 Slovenian regions and took additional 575 images of 134 hayracks, bringing our final dataset to a total of 701 photographs. Of these, there were 98 extended, 409 roofed double, and 178 split-level hayracks according to Melik’s, and 174 single and 511 double hayracks according to Hazler’s typology.

3. IMAGE CLASSIFICATION

We classified images and evaluated the models in Orange machine learning environment [1]. Hayrack images were converted into their numeric representation with a pre-trained neural network Inception-v3 [8]. We evaluated the performance of logistic regression using 10-fold cross-validation, and manually defined folds to avoid having images of the same hayrack both for training and testing. Figure 2 shows an Orange workflow that was applied to images from both typologies.

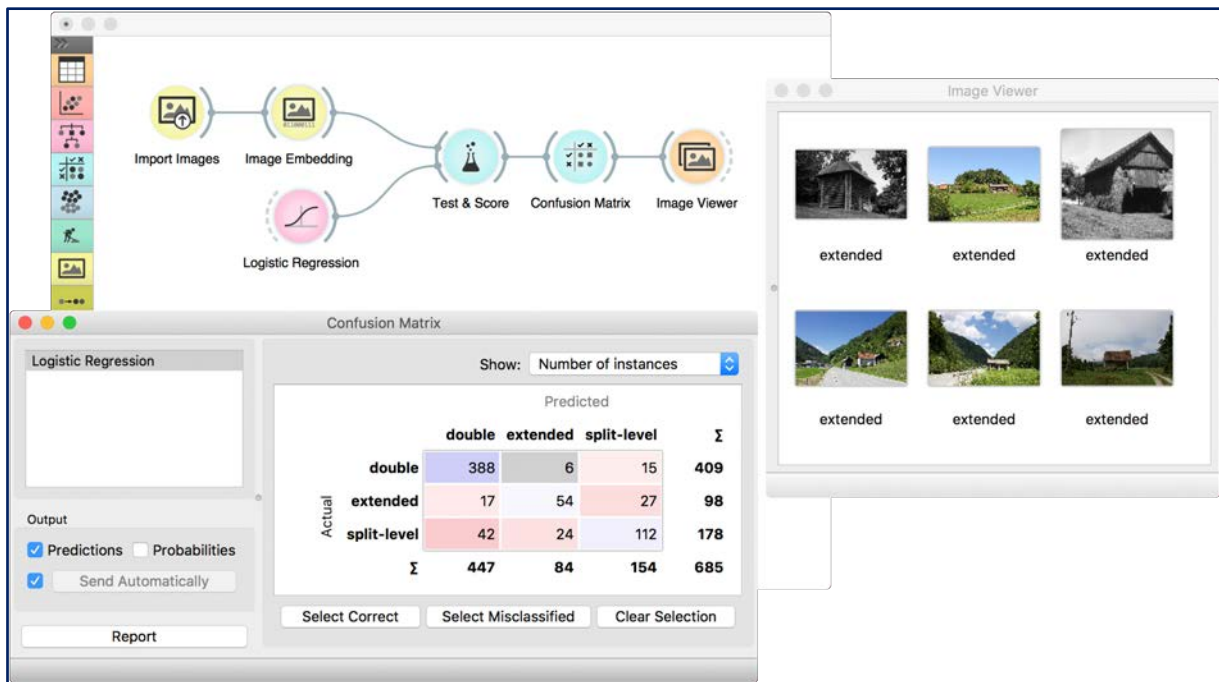


Figure 2: This workflow loads the images, embeds them into a vector-based representation with a deep neural network, and estimates the accuracy of logistic regression using cross-validation. Confusion matrix shows type misclassifications, supporting the user to select select them and present them visually. Here we show the results for Melik’s typology, where six double hayracks were misclassified as extended.

model	AUC	CA	majority
Melik	0.870	0.809	0.597
Hazler	0.930	0.879	0.746

Table 1: Cross-validated accuracy scores (CA: classification accuracy, AUC: area under ROC curve) for logistic regression model on two typologies. The last column reports the proportion of hayracks in majority class.

Accuracy scores in Table 1 are high for both typologies, which means hayrack type can be successfully predicted from images regardless of the typology adopted. However, Hazler’s model seems to perform better across both scoring techniques. To inspect why, we observe misclassifications separately for Melik and for Hazler (Figure 3). We also identified images, where both models fail (column *both*).

It seems Melik’s model is sensitive to occlusion (Figure 3.a) and perspective (Figure 3.e) of the image, while Hazler’s fails on images where classification is often hard even for a professional (delineation between split-level and double hayracks is not precisely defined), (Figure 3.c and f). However, both seem to have some difficulties with distant objects (Figure 3.g, h, and i), which means the model could be further improved with object localization.



Figure 3: Example of misclassifications for both typologies (first column) and for each typology separately (second and third column).

4. DISCUSSION AND CONCLUSION

Image classification can help us uncover and understand typological differences between hayracks. Hazler’s simpler typology works better for prediction and reaches an excellent accuracy of 0.93 (AUC). Moreover, errors of Hazler’s model correspond better to man-made errors. Logistic regression model of Hazler’s typology is hence closer to manual labeling and could be used with a higher precision on unlabeled hayrack images for automatic classification.

Predictive models offer various possibilities for analyzing digitized archives. Hand-annotation of large collections takes a lot of time and resources. Instead, curators can use these models for semi-automatic annotation, where images can be labeled in batches. Moreover, machine learning can not only assist with hayrack typology, but can also be trained to locate hayrack in the image (cf. [2]).

Proposed methodology is scalable and can easily be extended to domains beyond hayracks – other building types, traditional clothing, tools and other elements of material culture. Algorithms and machine learning are not just a domain of computer scientists anymore – researchers and data custodians can use them as tools for data exploration, analysis and management.

5. REFERENCES

- [1] Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A. and Štajdohar, M., 2013. Orange: data mining toolbox in Python. *The Journal of Machine Learning Research*, 14(1), pp.2349-2353.
- [2] Harzallah, H., Jurie, F. and Schmid, C., 2009, September. Combining efficient object localization and image classification. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 237-244). IEEE.
- [3] Hazler, V., Vodovnik, B. and Vuk, R., 1999. *Podreti ali obnoviti?: Zgodovinski razvoj, analiza in model etnološkega konservatorstva na Slovenskem*. Rokus.
- [4] Juvanec, B., Pleština, L. and Bogataj, J., 2000. *Kozolec/Hayrack*. Univerza v Ljubljani, Ljubljana.
- [5] Klein, R., 2014. How to create a building typology?: Typological matrix for mapping 19th century synagogues. *Zbornik radova Građevinskog fakulteta, Subotica*, (24), pp.57-68.
- [6] Melik, A., 1931. *Kozolec na slovenskem*. Znanstvena društva v Ljubljani.
- [7] SEM museum, (2017). [online] Available at: <https://www.etno-muzej.si/en/digitalne-zbirke> [accessed 05/09/2017].
- [8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Scholarly Digital Edition: Manuscripts, Texts and TEI Encoding

un corso online per l'Informatica Umanistica

Marjorie Burghart

CNRS, France

ABSTRACT

Questa relazione propone un nuovo corso online destinato all'insegnamento e all'utilizzo del linguaggio XML e della Text Encoding Initiative al fine di realizzare delle edizioni digitali, con particolare attenzione all'edizione delle fonti manoscritte. Il corso, sottotitolato in diverse lingue, è indirizzato a un pubblico variato e propone un'originale alternativa pedagogica, compatibile con le esperienze dell'insegnamento capovolto

PAROLE CHIAVE

Text Encoding Initiative, Corso online, Codifica dei testi, Edizione, Manoscritti.

1. RELAZIONE

La Text Encoding Initiative (TEI) si è affermata, da tempo, non solo come uno standard essenziale nella realizzazione delle edizioni critiche digitali, ma anche come un mezzo per favorire lo scambio di fonti e di edizioni, in un contesto di crescita dell'Open Science. Agli studiosi desiderosi di formarsi in questo senso, la TEI offre numerosi strumenti pedagogici, a partire dalle Guidelines che servono da principale riferimento, ai corsi svolti da formatori provenienti da diverse istituzioni, fino ai diversi workshops organizzati regolarmente in Europa e Nord America, passando per delle opere di iniziazione come *What is the Text Encoding Initiative?* (Burnard, 2014) e l'eccellente sito web *TEI by Example* (Van den Branden, Terras & Vanhoutte).

Finora, però, è mancato un elemento essenziale a questa offerta pedagogica: un corso online che, associato all'uso del video e dei più recenti sviluppi dell'insegnamento on line, renda la formazione più dinamica.

Il corso *Scholarly Digital Editions: Manuscripts, Texts and TEI Encoding*, tenuto da Marjorie Burghart (CNRS, Francia) ed Elena Pierrazzo (Università di Grenoble – Alpes, Francia) e sostenuto dal programma Erasmus SP+ DEMM, l'ITN DiXiT, dal progetto europeo DARIAH Humanities at Scale, dal consorzio TEI, nonché dalla Région Auvergne-Rhône-Alpes, vuole colmare questa lacuna.

Il corso è destinato principalmente agli studiosi desiderosi di lavorare all'edizione di testi trasmessi attraverso fonti manoscritte, che si tratti di un'edizione critica di testi letterari basati su vari testimoni o di documenti d'archivio come, ad esempio, carte sparse o registri di conti. Il corso, pur essendo principalmente indirizzato ai medievalisti e agli modernisti può in egual modo interessare alcuni contemporaneisti. Vengono, infatti, affrontati i numerosi aspetti relativi alle problematiche legate all'edizione di un testo, in modo da rendere il corso accessibile sia agli studiosi specialisti dell'edizione, sia ad altri professionisti coinvolti in progetti di edizione e desiderosi di acquisire le competenze teoriche necessarie. Il corso, composto da una trentina di brevi filmati, girati in inglese e accompagnati da sottotitoli in italiano, francese ed inglese (con la prospettiva di ampliare l'offerta in altre lingue), è organizzato in cinque sezioni principali che associano i video a degli esercizi autocorrettivi, a dei test e ad altre risorse utili. I primi quattro sezioni del corso sono tenuti da M. Burghart ed E. Pierazzo e costituiscono un filone comune. "Introduction: Encoding and XML" pone le basi teoriche e pratiche concernenti il metodo, le procedure e le tappe da seguire nella creazione di un'edizione digitale in TEI XML; la seconda sezione "Manuscript Transcription" affronta nel dettaglio la descrizione e la codificazione TEI dei fenomeni testuali tipici delle fonti manoscritte (abbreviazioni, addizioni, cancellature etc.). "Critical Editions and the TEI" insegna le differenti strategie di codifica di un apparato critico in TEI, senza dimenticare di presentare e descrivere puntualmente le diverse difficoltà che si possono, talvolta, incontrare mentre si procede alla codificazione di un testo. L'ultima sezione del filone comune, "Critical Editions: Advances Features", fornisce, invece, gli strumenti per la creazione delle indicizzazioni dell'edizione (index hominum e locorum), e per la codificazione della cronologia.

La quinta sezione è composta da moduli tematici opzionali di taglio specialistico e si avvale della collaborazione di insegnanti invitati a partecipare a questo progetto. A Oggi, due moduli sono proposti: il primo, "Annotating Images" (E. Pierazzo), è dedicato alla codificazione delle edizioni di documenti grafici; il secondo, "TEI and XSLT" (Syd Bauman e Martin Holmes), è rivolto all'approfondimento dei rudimenti della trasformazione di files TEI XML con l'ausilio dei fogli di stile XSLT; questa sezione sarà arricchita, in futuro, di nuovi moduli tematici.

Publicato sulla piattaforma europea #dariahTeach in una versione personalizzata di Moodle, che fornisce in open access strumenti riconosciuti da una commissione inter pares per l'insegnamento dell'Informatica Umanistica, questo corso intende rivolgersi a un pubblico largo ed eterogeneo: dagli studiosi specialisti nel campo, a quei professionisti appartenenti alle diverse discipline coinvolte nei progetti di edizione scientifica e desiderosi di comprendere meglio e più fondo gli aspetti teorici della codificazione (informatici, specialisti della documentazione etc.). Il corso è anche indirizzato agli studenti che desiderano formarsi nel campo delle edizioni digitali. Il corso, che si basa sull'apprendimento autonomo, offre così

un'alternativa valida alle giornate di formazione a carattere residenziale a tutti quelli che, per mancanza di tempo o per lontananza geografica, non possono usufruire delle abituali offerte formative. Allo stesso modo, è adatto a esperienze di insegnamento capovolto ("flipped classroom" in inglese), durante le quali gli allievi possono seguire le lezioni online in maniera autonoma prima di partecipare a delle sessioni pratiche in presenza di un insegnante che, a quel punto, può dare tutto l'aiuto e i consigli possibili ai partecipanti, rispondendo a questioni pratiche, con particolare attenzione ai loro progetti. Questo approccio permette quindi di approfittare della flessibilità di una formazione online conservando, al contempo, gli stessi vantaggi dell'insegnamento tradizionale. L'integrazione dell'insegnamento virtuale (online) all'interno delle formazioni universitarie triennali e specialistiche è così facilitata, grazie alle sessioni in presenza del docente: infatti, la questione relativa alla convalida dell'insegnamento può essere risolta durante le sessioni di compresenza docente-allievo.

2. RINGRAZIAMENTI

Ringrazio la dottoressa Giulia Rainis per la sua velocissima traduzione di questa relazione.

Ringrazio anche la Région Auvergne-Rhône-Alpes per sostegno finanziario al progetto "Partager l'expertise en Humanités numériques: valorisation et développement d'un MOOC en anglais (+traduction) sur l'édition électronique" (2018-2020), basato sull corso online *Scholarly Digital Edition: Manuscripts, Texts and TEI Encoding*.

3. BIBLIOGRAFIA

- [1] Burnard, Lou. *What is the Text Encoding Initiative?* Marseille, Open Edition Press, 2014
- [2] Maglioni, Maurizio e Biscaro, Fabio. *La classe capovolta*. Trento, Erickson editore, 2014
- [3] TEI Guidelines: *Linee guida per la codifica e l'interscambio del testo elettronico* [<http://www.tei-c.org/release/doc/tei-p5-doc/it/html/>]
- [4] Van den Branden, Ron, Terras, Melissa & Vanhoutte, Edward. *TEI by Example*, <http://www.teibyexample.org> (accessed on 27 September 2017).

Stemmatology

An R package for the computer-assisted analysis of textual traditions

Jean-Baptiste Camps¹, Florian Cafiero²

¹ Ecole nationale des chartes – Paris Sciences et Lettres Research University - jean-baptiste.camps@enc.sorbonne.fr

² Ecole nationale des chartes – Paris Sciences et Lettres Research University - cafiero@phare.normalesup.org

ABSTRACT

Confronted to the profusion of a text's versions in different manuscripts, researchers can have a hard time understanding which parts are still as wished by the original author, and which have been changed by errors, corrections or personal inspiration of the various scribes. To answer this question, it is necessary to compute the most plausible history of the text – which texts has been copied by which? This complex and tedious task is one of the fundamental purposes of textual criticism and philology. Various computer-assisted procedures and methods have been described since the 1950's, some elaborating on traditional principles (Lachmannian, Quentinian...), some borrowed to fields such as phylogenetics. In this poster, we describe *Stemmatology*, a new package designed for the statistical software *R*, aiming at guiding users through this process. The underlying logic of our software relies on principles derived from the Lachmannian and neo-Lachmannian tradition in general, and particularly on some of Eric Poole's methodological insights and their further developments (Poole, 1974; Poole 1979; Camps & Cafiero, 2015). After a careful encoding of all the significant discrepancies between manuscripts ("variant locations"), the software locates the variant locations to be kept for the final study, and eliminates the "overconflicting" ones. Then, the user is helped to detect two different potential problems. The first one is "contamination", occurring when one manuscript contains passages copied from two or more manuscripts. The second is "polygenesis": different passages could have the same aspect only by chance but in fact derive from different sources. When these problems are either ruled out or addressed, the package computes a simplified *stemma codicum*, then guides the user in building his final version. The poster will also present the projects of new features for this package, such as contamination detection through cardiograms (Watel & Mulken, 1996 ; Den Hollander, 2004).

KEYWORDS

textual criticism, philology, R

1. INTRODUCTION

Before the appearance of the printing press, in the West, the only way of reproducing and spreading a text in written form was manual copying. During this process, both accidents, errors and intentional modifications occurred, progressively modifying the text of each witness. For the philologist interested in the study of a textual tradition or the restoration of the original text, it has been imperative to study the different variants of the witness, to assess their genealogical relations, at least since the beginning of the scientific age of philology in the XIXth century. As a result, the method of common errors (often deemed "Lachmannian") took progressively form during the XIXth and the beginning of the XXth century. Yet, different phenomena such as horizontal transmission (contamination) or the independent appearance of identical variation in different witnesses (polygenesis) cause major difficulties to this method.

Dating back to the experiments of Dom Froger (Froger, 1968), various computational procedures have been used to help assess the genealogy of a textual tradition. These can be roughly divided between methods based on pre-existent textual criticism principles (Lachmannian, Quentinian or other) and methods inspired by fields other than textual criticism, such as phylogenetics or compression-based algorithms (Roos, Heikkilä, 2009).

Amongst these methods, we have decided to start by implementing a method based on traditional philological principles, firstly because these are often less available or lack open-source and user-friendly implementations, secondly because, for now, classical methods (including non computerised ones), still seem to offer very satisfactory results when compared to others. We chose to start with the approach designed by Poole (Poole, 1974; Poole, 1979), extended by Camps & Cafiero (Camps, Cafiero, 2015). It offered us an algorithmic and easy to compute transposition of the common error method, and because we wanted to orient the effort towards some of the major problems of the analysis of textual genealogies, namely contamination and polygenesis. Finally, we wanted to implement methods that value interaction with the expert and allow his or her insights to enhance the results.

2. DATA MODEL

It is not our purpose to decide which data model is better to represent textual variation, or to put constraints on the meaning given to a witness or a variant location. Moreover, the definition of the basic unit of variation, the inclusion or not of some types of variation, are all features that can deeply vary from one context to another or between projects. A variety of models already exists, from the word-based collation table to the TEI encoded apparatus or the graph model (Schmidt, Colomb, 2009). To stay independent from these choices, we adopted a very simple and abstract representation, in which each column

stands for a witness, each line for a variant location, and each variant is given a numeric code (NA for not acquired, 0 for omission, 1...n for variants). The choice to consider omission as readings or not is left to the user.

To keep interoperability at maximum, we offer a configurable and easily customised XSLT stylesheet to transform a TEI-encoded parallel-segmentation apparatus to our data format. We welcome the contribution of other stylesheets to the repository. To facilitate quick experimentation with the package, and reproducibility of procedures, we ship the package with various data-sets, based on artificial or historical traditions.

3. FEATURES OF THE PACKAGE

To implement our package, we chose the open-source environment for statistical analysis R. R makes it very easy to fully document the procedures, include both commands and data-sets, and distribute the package. In its current state, the package is composed of two main sets of features. The first set is dedicated to exploratory analysis of a textual tradition, the second to building a stemma.

In the exploratory set of functions, we have implemented a procedure to cross two by two every variant locations, to identify contradictions in their genealogical configurations. The list of contradictions can then be plotted as a network. This helps the expert assessing which variant locations are the source of the contradictions, through the visualization or through computations (centrality degree). It gives him an intuition on the "conflictuality" between variant locations, and helps him decide which variants generate too many contradictions to be kept for further analysis. If contamination is suspected, various functions are offered to try to distinguish contamination from random polygenetic noise, and to identify contaminated witnesses or localised cases of contamination. Another function allows, for simple yet undecidable cases, to output different databases corresponding to the competing configurations, which can then result in different stemmata.

The second set of functions allows the user to build one or several stemmata, depending on the input. For the construction of the stemma, only one method is implemented as of now, relying on the transformation of the common error method into a disagreement-based algorithm, formulated by Poole (Poole, 1974; Poole, 1979). As such, it is mainly based on disagreements opposing at least two witnesses to at least two others.

4. FURTHER DEVELOPMENTS

The algorithm proceeds step by step, first assessing groups, then reconstructing or identifying their model, and then restarting from step one after eliminating the *codices descripti* from the database. It keeps going until there are less than four witnesses in the database, and, though it can compute a plausible configuration, it lets the expert free in the decision regarding the top of the stemma, the place which is the most delicate to assess (Bedier, 1928). This first version of the package calls for new developments and improvements.

We are in the process of implementing other functions to detect contamination, in particular localised-cases, with procedures such as "cardiograms" (Den Hollander, 2004; Wattel, VanMulken, 1996). In the future, we would like to offer easy access to a large set of existing methods for stemma-building, to facilitate comparisons or benchmarks. The availability of the open source code makes it easy for others to help us complete our scope and contribute to software development.

5. REFERENCES

- [1] Bédier, Joseph, 1928, "La tradition manuscrite du *Lai de l'ombre*: réflexions sur l'art d'édition des anciens textes", *Romania*, 54, p.~161-196 and 321-356.
- [2] Camps Jean-Baptiste and Cafiero Florian, « Genealogical variant locations and simplified stemma: a test case », in *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, dir. Tara Andrews & Caroline Macé, Turnhout, 2015, p. 69-93 (Lectio, 1).
- [3] A.A. Den Hollander, « How shock waves revealed successive contamination. A cardiogram of early sixteenth-century Dutch Bibles. » in *Studies in Stemmatology 2*, dir. P. Van Reenen., A.A. Den Hollander & M.J.P. Van Mulken, Amsterdam, 2004. p. 99-112.
- [4] Froger, Jacques, 1968, *La critique des textes et son automatisation*, Paris (Initiation aux nouveautés de la science)
- [5] Eric Poole, « The Computer in Determining Stemmatic Relationships », *Computers and the Humanities 8-4 (1974)*, p. 207-16; Id., « L'analyse stématique des textes documentaires », in *La pratique des ordinateurs dans la critique des textes*, Paris, 1979, p. 151-161.
- [6] Roos, Teemu and Heikkilä, Tuomas, 2009, "Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets", *Literary and Linguistic Computing*, 24-4, p.~417-433
- [7] Schmidt, Desmond and Colomb, Robert, "A data structure for representing multi-version texts online", *International Journal of Human-Computer Studies*, 67-6, p.~497-514
- [8] E. Wattel & M.J.P. Van Mulken, 1996. « Shock Waves in Text Traditions, Cardiograms of the Medieval Literature », in *Studies in Stemmatology*, P. Van Reenen, M.J.P. Van Mulken, Amsterdam, 1996, p.105-121.

Stories in Situ

Humanities–informed media design for location–based storytelling

Gunnar Liestøl

Dept. of Media & Communication, University of Oslo, Norwa

ABSTRACT

Research in the humanities is often analytical oriented and focused on subject matters post factum. In this proposal we will show how rhetorics can be reemployed as a method for digital design in location–based media. In particular we will focus on how theoretical terms and analytical concepts and categories may serve as rules for construction and composition of new types of digital genres. Our theoretical and methodological probes will be demonstrated by a set of experimental designs with location–based storytelling by means of mobile augmented reality in cultural heritage settings.

KEYWORDS

Augmented reality, situated simulations, storytelling, narrative, cultural heritage, locative storytelling

1. PROPOSAL

The following describes experimental probes into a media design methodology aimed at inventing innovative rhetorical devices and narrative techniques for storytelling and narrative experimentation in situated and sensory media (mobile augmented reality). The paper claims that traditional theoretical–analytical concepts in humanities–informed media studies may serve reverse (inverse) and supplementary purposes in the context of digital media design.

In humanistic media studies theoretical and conceptual constructions are primarily *post factum* endeavors – hindsight reflections always presuming the existence of a media object or phenomenon, the subject matter to be studied, understood and explained. This is a common trait in the humanities in general as well as in humanistic media studies in particular: the subject matter exists prior to the conceptual constellations describing and explaining their properties, and they are primarily produced by actors independent of the analyst’s knowledge and understanding. As a consequence, a situation emerges where, in principle, a concept, given its general characteristics, also may exist prior to a new phenomenon, a new subject matter, which still is pending to come into existence.

In this perspective such concepts will potentially convey meaning and contain explicable value towards new not yet existing objects. In consequence: when such an ‘a priori’ concept might serve as an explanation of certain features in an object, which does not yet exist, that concept may also serve as a rule (schema or topos) for how to design and construct new digital media objects, objects which will thus have features inherent to the given concept [1].

In our presentation we will use this perspective on *inversion* of theoretical and analytical concepts for the purpose of *critical construction* in order to describe a design and a prototype of a potentially new narrative format. This potential genre was early on coined as ‘situated documentary’ [2], a variant of an Indirect Augmented Reality application [3], reconstructing and simulating historical events *in situ* [4]. Augmented reality storytelling has increasingly gained interest over the past decade [5, 6, 7]. Chosen historical venues for experiments with location–based storytelling will in this context will be Via Appia Antica, Ancient Phalasarua, Forum Romanum, and the D–Day assault on Omaha Beach in Normandy.

Based in rhetoric as an architectonic productive art [8] we will show how concepts and categories of narratology [9] and genre theory may serve as topoi for finding (inventing) innovative new narrative features and designs in the development of situated simulations as a form of augmented reality storytelling [10].



Figure1: US soldiers in the morning on D-Day trying to cross the beach under heavy German fire (still from a video showing the ‘Omaha Beach’ mobile augmented reality app in use on location). Link to video demo: <<https://youtu.be/I6LZIFueq1g>>

2. REFERENCES

- [1] Liestøl, G. (2013b) 'Topics of Innovation. Towards a Method of Invention and Innovation in Digital Media Design' in Storsul, T. & Krumsvik, A. H. (Eds.) *Media Innovations. A Multidisciplinary Study of Change*. Gothenburg: NORDICOM, pp. 61–74.
- [2] Hollerer, T., Feiner, S., & Pavlik, J. (1999) 'Situating Documentaries: Embedded Multimedia Presentations in the Real World.' in *Wearable Computers. Digest of Papers. The Third International Symposium on Wearable Computers IEEE*, San Francisco 1999, pp. 39-86. DOI: 10.1109/ISWC. 1999.806664
- [3] Wither, J., Tsay, Y-T., Azuma R (2011): 'Indirect Augmented Reality' in *Computers & Graphics* 35 (2011). 810–822
- [4] Liestøl, G. (2011) 'Situating Simulations between Virtual Reality and Mobile Augmented Reality: Designing a Narrative Space' in B. Furht (Ed.) *Handbook of Augmented Reality*, Springer.
- [5] Papagiannakis, G., Schertenleib, S., O'Kennedy, B. 'Mixing virtual and real scenes in the site of ancient Pompeii.' in *Computer Animation and Virtual Worlds* Vol. 16 (1) John Wiley and Sons Ltd.: Budapest,
- [6] Liestøl, G. (2014) 'Along the Appian Way. Storytelling and Memory across Time and Space in Mobile Augmented Reality', in M. Ioannides et al. (Eds.) *Progress in Cultural Heritage: Documentation, Preservation, and Protection*. Berlin Heidelberg: Springer Verlag, pp 248-257 *Archaeolingua*, 2005: 11- 24.
- [7] Azuma, R. (2015) 'Location–Based, Mixed and Augmented Reality Storytelling' in Barfield, W. (ed.) *Fundamentals of Wearable Computing and Augmented Reality 2nd Edition*, CRC Press 2015, pp. 259–276.
- [8] McKeon, R. (1971/2005) 'The Uses of Rhetoric in a Technological Age: Architectonic Productive Arts' in *The Selected Writings of Richard McKeon: Culture, Education and the Arts*, Volume 2. (edited by Z. K. McKeon and W. G. Swenson, with and introduction by W. C. Booth). Chicago: The University of Chicago Press.
- [9] Genette, G. (1980) *Narrative Discourse. An Essay on Method*. Ithaca: Cornell University Press.
- [10] Liestøl, G. & Morrison, A. (2013) 'Views, Alignment and Incongruity in Indirect Augmented Reality' in *Proceedings of ISMAR 2013, IEEE International Symposium on Mixed and Augmented Reality - Arts, Media and Humanities*. 1-4 Oct. 2013, pp. 23-28.

The ERC Project “Localizing 4000 Years of Cultural History. Texts and Scripts from Elephantine Island in Egypt”¹

Andrea Hasznos¹, Ahmed Kamal Mamdoh²

¹Ägyptisches Museum und Papyrussammlung, Berlin, Germany - hasznosa@gmail.com

²Ägyptisches Museum und Papyrussammlung, Berlin, Germany - abokamal_85@yahoo.com

ABSTRACT

“Localizing 4000 Years of Cultural History. Texts and Scripts from Elephantine Island in Egypt” in the Berlin Papyrus Collection is a project deeply rooted and inspired by Digital Humanities, led by Prof. Dr. Verena Lepper and funded by the European Research Council. The incredibly rich and exciting written material extant from the island of Elephantine contains texts in Ancient Egyptian, Coptic, Aramaic, Greek, sometimes bilingual (sometimes an interesting mixture of loan-constructions), written in hieroglyphs, hieratic, demotic, Coptic, Greek, and Aramaic and spans over a period of four thousand years. Today, the several thousand written pieces (on papyrus, ostrakon, parchment, paper, wood, and inscriptions on walls) are spread throughout more than 60 institutions around the world. That is one of the reasons why digitizing and cooperating also in a modern, virtual and electronic way is inevitable. Not only to preserve an invaluable part of cultural heritage, but also to try and bring together fragments which used to make up one piece and are currently in two or more different parts of the world. For that, a comprehensive database (in File Maker) has been built up containing meta-data, detailed descriptions, and photos of each object, accompanied by Oxygen as a work tool in which the transcription and translation of the texts take place, tagging the texts for all relevant information (persons, peoples, dates, places, lacunae, additions). The scholars of the different languages and scripts work on their own texts and also side by side to share information and insights, as well as regularly contacting and cooperating with the experts in Computer Science, and participating in the periodic plenary sessions where the mathematicians and the physicians share the results of their analyses and research on ink types and the virtual unfolding of papyrus rolls. The online database will thus reunite several thousand objects accompanied by research results and painting a vivid picture of the island of Elephantine and surroundings the culture, literacy, religions, and everyday activities of its inhabitants over the four thousand years that it was a crossroads and melting pot of ancient cultures.

KEYWORDS

database, text encoding, virtual unfolding of papyrus, digitizing, cultural heritage, Egypt, Elephantine, papyrus, ostrakon, parchment, multicultural, multilingual, multi-ethnic, religion, Greek, Coptic, Aramaic, Arabic, hieroglyphs, hieratic, demotic, polytheism, Judaism, Christianity, Islam

1. AIMS OF THE RESEARCH PROJECT

We would like firstly to introduce the Digital Humanities project funded by an ERC grant “Localizing 4000 Years of Cultural History. Texts and Scripts from Elephantine Island in Egypt” led by Prof. Dr. Verena Lepper (Berlin) and running since 2015, and secondly to show how it works in practice and what challenges we as scholars in the Humanities (rather than Computer Science experts) face when digitizing and using databases and text encoding programs. This is a project built upon and thriving on the very fruitful cooperation between scholars in the Humanities from different disciplines, and experts in the fields of Digital Humanities, Computer Science as well as Physics and Mathematics in order to digitize, process and preserve a precious slice of human cultural heritage. The present paper will be held by Andrea Hasznos who is working on the Coptic material, and Ahmed Kamal Mamdoh who is working on the Arabic texts.

Elephantine, this island on the southern border of Egypt, shows an especially vivid multicultural, multi-ethnic, multi-religious, and multilingual scenery documented in an almost unbroken fashion from the Old Kingdom up to the era after the Arab conquest. This lets the observer see and examine processes and habits in everyday life, legal and financial procedures, cultural and religious phenomena, and all that in a very “globalized” context.

Today, several thousand written pieces (on papyrus, ostrakon, parchment, paper, wood, and inscriptions on walls) from Elephantine are spread throughout more than 60 institutions across Europe and beyond. And that in itself shows the importance of digitizing and cooperating also in a modern, virtual and electronic way, making use of the advantages of the digital age. The extant texts are written in various languages and scripts, including hieroglyphs, hieratic, demotic, Aramaic, Greek, Coptic and Arabic, and most of them are unpublished and not yet investigated or evaluated. The great challenge of this project is to process this material with focus on three key topics:

1. Multiculturalism and identity between assimilation and segregation,

¹ Principal Investigator: Prof. Dr. Verena Lepper, Director of the ERC-Starting Grant-Project ELEPHANTINE and Tenured Curator for Egyptian and Oriental Papyri at the Ägyptisches Museum und Papyrussammlung, Berlin
Host Institution: STIFTUNG PREUSSISCHER KULTURBESITZ
VON DER HEYDT STRASSE 16-18
10785 BERLIN
Germany

2. the organization of family and society,
3. the development of the religions (polytheism, Judaism, Christianity and Islam).

2. PRACTICALITIES

The results of the project, description and meta data of each object, their texts and translations will be made accessible in an online database, reuniting several thousand objects (distributed in over 60 collections worldwide) and painting a vivid picture of the island of Elephantine and surroundings, its inhabitants, culture, literacy, religions and their interactions, and the everyday activities of its population over the four thousand years that it was a crossroads and melting pot of ancient cultures. The online database serves one of the main purposes of any such project, namely, digital preservation of material through detailed description as well as high-resolution photographs. Not only digitally, but also physically preserving the material is the responsibility of the conservator(s) working in the project. The online database will also allow the identification of connections between manuscript fragments from different collections, which can then, at least virtually, be "put together". We may hope that through international cooperation, the "papyrus puzzle" can thus be solved.

Cutting-edge new methods not only from the Digital Humanities, but also in Physics and Mathematics are also part of the project, for example for the virtual unfolding of papyri. Because many of the papyri remain rolled or folded and cannot be safely opened to be read, as part of the Elephantine project, virtual unrolling and unfolding by tomography to reveal the text is being developed. For texts written in metal-containing inks, absorption tomography is being pursued. For texts written in carbon-based inks, suitable alternatives are being investigated in a feasibility study.

Using the TEI (Text Encoding Initiative) within the programme called "Oxygen" we encode the texts. So texts will not only appear in their original form and script, with an up-to-date translation, but will also be tagged, which means that persons, dates, places are immediately visible for the reader, and it also shows where lacunae or scholarly additions can be found in the text. The text and translation thus produced will then be incorporated into the database, which has been created in File Maker. A predecessor of that was developed by Daniel Werning for the Rubensohn Bibliothek Projekt already online and accessible on the Berlin Museum's website (<http://elephantine.smb.museum/suche/?lang=en>). Relevant data were then migrated into the new, now much more enormous database further developed by Daniel Werning and Sandro Schwarz – in each case, naturally, the developers worked on the database together with the scholars working on the ancient textual material, most notably Jan Moje, Verena Lepper, James Moore. The database contains already some 5000 datasets at the moment.

Through the use of this database, with its medical, religious, legal, administrative and even literary texts, researchers will be able to study the microhistory of everyday life of the local and global ('glocal') community of Elephantine, in the context of its sociocultural Egyptian background and beyond. This will be coupled with macro-historical topics, for which the newest methods of global history will also be used. In this way, Elephantine can be employed as a case study and model for the past, present and future.

Throughout the project an inevitable and fruitful cooperation between scholars in the Humanities and technology experts has prevailed and will continue to do so, as problems and difficulties arise and therefore recommendations are made in the programs applied to meet user needs and to eventually have a secure, comprehensive and user friendly database.

3. ACKNOWLEDGEMENTS

We are very grateful to the European Research Council for supporting and funding this project.

4. BIBLIOGRAPHY

- [1] Abbott, N. 1941. "Arabic Marriage Contracts among Copts", *Zeitschrift der Deutschen morgenländischen Gesellschaft* 95, 59-81
- [2] Bacot, S. – Heurtel, Ch. 2000. Ostraca coptes d'Élephantine au musée du Louvre, in: N. Bosson (ed.), *Études coptes VII. Neuvième journée d'études, Montpellier 3-4 juin 1999*, Éditions Peeters, 17-45
- [3] Beltz, W. 1984. "Die koptischen Zauberpergamente der Papyrus-Sammlung der Staatlichen Museen zu Berlin", *Archiv für Papyrusforschung und verwandte Gebiete* 30, 83-104
- [4] Beltz, W. 1978. "Katalog der koptischen Handschriften der Papyrussammlung der Staatlichen Museen zu Berlin (Teil I)", *Archiv für Papyrusforschung* 26
- [5] Beltz, W. 1980. "Katalog der koptischen Handschriften der Papyrussammlung der Staatlichen Museen zu Berlin (Teil II)", *Archiv für Papyrusforschung* 27
- [6] Beltz, W. 1980. "Die koptischen Zauberstraka der Papyrus-Sammlung der Staatlichen Museen zu Berlin", *Hallesche Beiträge zur Orientwissenschaft* 2, 59-75
- [7] Beltz, W. 1985. "Die koptischen Zauberpapiere und Zauberstraka der Papyrus-Sammlung der Staatlichen Museen zu Berlin", *Archiv für Papyrusforschung und verwandte Gebiete* 31, 31-41
- [8] Clackson, S. J. 2010. 'Coptic or Greek? Bilingualism in the Papyri' in Papaconstantinou, A., *The Multilingual Experience in Egypt, from the Ptolemies to the Abbasids*, Ashgate Publishing Company, 73-104
- [9] Clackson, S.J. – MacCoull, L.S.B. – Porten, B. – Farber, J.J. 2011. The Coptic Texts. In: Porten, B., *The Elephantine Papyri in English. Three Millennia of Cross-Cultural Continuity and Change*, Atlanta, 569-602
- [10] Clackson, S.J. 1995. "Four Coptic Papyri from the Patermouthis Archive in the British Library", *Bulletin of the American Society of Papyrologists* 32, 97-116
- [11] Crum, W.E. 1921. *Short Texts from Coptic Ostraca and Papyri*, Oxford

- [12] Delange, É. 2012. *Les fouilles françaises d'Éléphantine (Assouan) 1906-1911, les archives Clermont-Ganneau et Clédat*, Mémoires de l'Académie des Inscriptions et Belles-lettres 46, Paris: Académie des Inscriptions et Belles-Lettres, 2 Vols.
- [13] Dietrich, A. 1937. *Arabische Papyri aus der Papyrussammlung der Hamburger Staats- und Universitäts-Bibliothek*, Abhandlungen für die Kunde des Morgenlandes 22, 3, Leipzig
- [14] Dietrich, A. 1955. *Arabische Briefe aus der Papyrussammlung der Hamburger Staats- und Universitäts-Bibliothek*, Veröffentlichungen aus der Hamburger Staats- und Universitätsbibliothek 5(Hamburg
- [15] Dijkstra, J.H.F. 2008. *Philae and the End of Ancient Egyptian Religion: A Regional Study of Religious Transformation (298-642 CE)*, Orientalia Lovaniensia Analecta 173, Leuven
- [16] Engelbach, R. 1938. "A Coptic Ostrakon Mentioning Ieb (Elephantine)", *Annales du Service des antiquités de l'Égypte* 38, 47-52
- [17] Farber, J. J. 1990. "Family Disputes in the Paternouthis Archive" *Bulletin of the American Society of Papyrologists* 27, 111-122
- [18] Hasitzka, M. 1993. *Koptisches Sammelbuch I*, Wien
- [19] Hintze, F. 1977. "Berliner koptische Ostraka", *Zeitschrift für Ägyptische Sprache und Altertumskunde* 104, 97-111
- [20] Hodak, S. – Richter, T.S. – Steinmann, F. 2013. *Coptica. Koptische Ostraka und Papyri, koptische und griechische Grabstelen aus Ägypten und Nubien, spätantike Bauplastik, Textilien und Keramik*. Katalog Ägyptischer Sammlungen in Leipzig, Band 3, Berlin: Manetho Verlag
- [21] Kaiser, W. 1973. "Elephantine. Bericht über die 1. und 2. Grabungskampagne" *Annales du Service des antiquités de l'Égypte* 61, 87-91
- [22] Khoury, R. G. 1993. *Chrestomathie de papyrologie arabe. Documents relatifs à la vie privé sociale et administrative dans les premiers siècles islamiques*, Handbuch der Orientalistik. Ergänzungsband 2, Zweiter Halbband, Leiden
- [23] Lepper, V. M. 2012. Die ägyptische und orientalische „Rubensohn-Bibliothek“ von Elephantine. 4000 Jahre Kulturgeschichte einer altägyptischen Insel. In: Lepper, V.M. (ed.), *Forschung in der Papyrussammlung. Eine Festgabe für das Neue Museum. Ägyptische und Orientalische Papyri und Handschriften des Ägyptischen Museums und Papyrussammlung Berlin 1*, Berlin: Akademie Verlag, 497–508
- [24] MacCoull, L.S.B. 1990. "Christianity at Syene/Elephantine/Philae" *Bulletin of the American Society of Papyrologists* 27, 158-159
- [25] Margoliouth, D. S. – Holmyard, E. J. 1930. "Arabic Documents from the Monneret Collection," *Islamica* 3, 249-271
- [26] Müller, W. 1980. *Die Papyrusgrabung auf Elephantine 1906-1908. Das Grabungstagebuch der 1. und 2. Kampagne*. In: *Forschungen und Berichte*, Bd. 2, 150 Jahre Staatliche Museen zu Berlin, 75-88
- [27] Rubensohn, O. 1907. *Elephantine-Papyri, Ägyptische Urkunden aus den kgl. Museen in Berlin. Griechische Urkunden*, Berlin
- [28] Schaten, S. 2005. Die koptischen Ostraka. In: Dreyer, G. et.al., *Stadt und Tempel von Elephantine: 31./32. Grabungsbericht*, MDAIK 61, 125-128.
- [29] Thomann, J. 2015. An Arabic Ephemeris for the Year 931-932 CE. In: A. Kaplony – D. Potthast – C. Römer (eds.), *From Bāwīt to Marw: Documents from the Medieval Muslim World*, Leiden, 115-153
- [30] Werning, D.A. 2013. „Thesauri:Languages and scripts,, „Thesauri:Materials,, „Thesauri:Colors,, „Thesauri:Support_types,, „Thesauri:Text_types,, „Thesauri:Text_types,, In: *TEI-XML in Egyptology and Coptic Studies* <https://wikis.hu-berlin.de/teiegyptology/>, ed. by Daniel A. Werning
- [31] Werning, D.A. 2015. Conference paper slides TEI and Thesauri in the Rubensohn Project (ÄMP Berlin), International Workshop „Annotated Egyptian Corpora and TopBib Online — Exchange, Convergence, Shared Objectives“ hosted by the Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, 27—29 April 2015
- [32] Werning, D.A. In review. Information Technology and Digital Humanities Workflow in the Rubensohn Project: Research Website and Rich TEI XML Header Encoding, approx. 14 pages
- [33] Worp, K.A. 1990. "Das berliner Ostrakon P. 14735: koptisch oder griechisch?“, *Archiv für Papyrusforschung* 36, 75-77

ToposText: Broadening the Digital Humanities Public

Brady Kiesling

American School of Classical Studies, Athens, Greece, westtothesea@hotmail.com

ABSTRACT

ToposText is a free mobile application (iPhone and Android) and website designed to put in the hands of students and educated travelers a historical gazetteer and library of ancient texts relevant to the topography of ancient Greece, in English translation: now 14 million words, tagged to 6500 mapped places, mostly from the Pleiades Project but enriched with hundreds of minor archaeological sites in Greece. Outside Greece, the gazetteer includes places with an ancient literary footprint. Lessons learned: human obsessiveness is still an acceptable substitute for technological elegance; think flexible interfaces, large rather than small; an imperfect app that exists is superior to a perfect one that doesn't.

A substantial body of clean, machine-readable, searchable texts in English allows cross-disciplinary historical inquiry by non-specialists. ToposText is expanding to include more Latin texts and Roman imperial place names, but help is needed to incorporate the wealth of information available in other projects. Building a user community is labor-intensive but necessary.

Sample research result: Making the ToposText geographic information available in .kml (Google Earth) files has promoted useful insights, for example on the orientation of Greek temples.

KEYWORDS

topography, Greek archaeology, gazetteers, cultural tourism, digital texts

1. OVERVIEW

ToposText was designed to be a multi-purpose ancient history/archaeology mobile application for aspiring scholars but also for a broader public. Its motivational prejudice was a conviction that modern scholarship, written to impress a closed community of academic rivals, often obscures the reality of our all-too-human, quite understandable ancient ancestors. Primary sources, offered in controlled doses *in a context that makes them relevant*, can bring even simple tourists closer to their ancient roots. Whether my vision of reading ancient sources *in situ* is naïve or not, the evolution of ToposText may offer useful lessons for scholars aspiring to make their DH projects tools to be used rather than artworks to be admired.

The core of ToposText is a growing corpus of texts from the Greek and Roman *oikoumene*, composed before 500 CE or drawn from earlier texts, in English translation, collected from the internet or home-translated, then cleaned up and formatted in very basic html. There is a slight collection bias in favor of ancient religion, partly out of personal interest, partly because mythology is the most reliable entry point to antiquity for most potential users. Almost all texts have hyperlinks to the original Greek or Latin text. Each paragraph is tagged with a pair of year dates, of composition and of the events described. The full text library is resident on-device (Android or iOS smart phones and tablets).

These texts are indexed against a growing database of ancient places, drawn originally from Pleiades (and more recently from the Digital Atlas of the Roman Empire), but manually updated with archaeological sites and museums that might offer useful information to visitors. Place mentions (Naxos in the Aegean, Naxos in Sicily) are manually disambiguated, a labor-intensive work that could be automated somewhat more than it has been. Where possible, the coordinates have been corrected on Google Earth to allow navigation to within visual range of the ancient remains. These places are then mapped in Google Maps, with type icons, a navigational link and a link to a brief description and to a key-word-in-context index of all the tagged instances of the various forms of the place name.

Some 10500 other proper names (gods, people, tribes, some animals) are included in a separate key-word-in-context index. A half-hearted attempt has been made to link variant spellings of the same name (Pantheia and Panthea are the same person, Panthia is different) and to disambiguate important names: Plato the philosopher from Plato the comic poet, Alexander the Great from a swarm of lesser figures.

Many sites are linked to early traveler illustrations via the Travelogues.gr website. Many sites in Greece have modern site descriptions from various sources, particularly McGilcrist's Greek Islands and the Greek Culture Ministry's Odysseus website.

The TT database currently includes 702 texts (defined by their distinct CTS identifier) totaling 14 million words. A small number of inscriptions have been included, thanks to the Attic Inscriptions Online project. Likewise, two key ancient dictionaries, the Suda and Harpokration, have been made available via online collaborative projects. The TT library does not include most medical and philosophical texts, which have few place names. A more important gap is ancient scribal commentary and later Roman/Byzantine texts, because translations are for the most part not yet in the public domain.

The tools for creating ToposText were basic indeed: Excel for the places, works, and personal names databases; Notepad++, a free text editor with a good regular expression engine for text formatting and search; Strawberry Perl for the tagging, indexing, transliteration, and formatting scripts; Abbyy Finereader for OCR; Google Earth for locating and assigning WGS84 coordinates in digital degrees.

2. RESULTS AND NEXT STEPS

The initial focus of ToposText was Classical Greece, an arbitrary distinction that caused me to throw out (and later restore with much wasted effort) important data. Though the text corpus and places/proper names databases now (as of 12/2017) includes the standard Roman historical and literary texts and place information for Rome and much of western Europe in reasonable detail, little has been done to publicize ToposText as a resource for Roman history and culture. Success of the project can should be measured in the degree to which useful feedback, additions, and corrections begin to flow. So far, this is a mere trickle. The most difficult part of any such project is building a user community – a job as energy-intensive as creating the database. A small number of elite users – graduate students in archaeology in Greece for the most part – have become enthusiastic users. One good sign is that Greeks are complaining about the lack of a modern Greek version. I hope Italians will complain as well.

ToposText includes standard text and places metadata identifiers (Pleiades, CTS) and is a good fit with the Linked Open Data paradigm. Work is underway to expose ToposText data to display in the Peripleo interface of the Pelagios Project. Other necessary steps include substantial improvements to the user interface, to make navigation faster, smoother, and more intuitive and, crucially, to implement off-line maps in sufficient detail to keep ToposText useful in the absence of an internet connection. Enabling proximity searches (two words or search strings close to one another) on the web interface is a key desideratum.

Ultimately, intellectual property rights obstacles will be overcome so we can gather in one searchable corpus modern-language translations of all significant Greek and Latin texts. This will be a substantial incentive for amateur historians who lack the expertise and access passwords to conduct such searches in Greek or Latin. The ability to compare multiple references to a specific place or person across multiple sources spanning a millennium is very useful.

For example, a simple regular expression search, ‘vow.{1-100}[tT]emple’ generates 50-odd examples of temples being built in response to a vow, usually but not always connected with a military adventure or a bout of the plague. The extensive collection of temple and other site locations assembled in ToposText and viewed in Google Earth offers interesting insights about the way the axial orientation of temples reflects the political as well as religious purposes ancient temple construction served.

3. ACKNOWLEDGEMENTS

ToposText relies on the kind support of the Aikaterina Laskaridis Foundation, Piraeus, Greece, which subsidized the app programming and the topostext.org website, both the work of Pavla A.E., an Athens-based technology company that performed much of the work with little prospect of compensation. Data points have been incorporated from the Pleiades Project, Johan Åhlfeldt’s Digital Atlas of the Roman Empire, the Archaeological Atlas of Thessaly, kastra.eu, and other resources. Texts have been culled from many sources, but scholarship owes a huge debt of gratitude to the Perseus Project, to Bill Thayer’s [Lacus Curtius](http://LacusCurtius), to Roger Pearse’s Tertullian.org, and to archive.org.

TRI: a tool for the diachronic analysis of large corpora and social media

Pierpaolo Basile¹, Annalina Caputo², Giovanni Semeraro¹

¹ Department of Computer Science – University of Bari Aldo Moro - {firstname.lastname}@uniba.it

² ADAPT Centre, Trinity College Dublin, Dublin, Ireland - annalina.caputo@adaptcentre.ie

ABSTRACT

In this paper, we describe TRI a tool for the diachronic analysis of large corpora. TRI is based on Distributional Semantic Models, and it is able to build several WordSpaces for different time periods. In these WordSpaces, words are represented as mathematical points in a geometric space and two words are similar if they are close in the space. Using this approach it is possible to analyze how a word changes its meaning over time, or how the relatedness between two words changes. We report several case studies and an evaluation on Italian corpora. Finally, we describe a preliminary evaluation in the context of social media.

KEYWORDS

Diachronic analysis, Distributional Semantic Models

1. INTRODUCTION

During the last decade, the surge in available data spanning different epochs has inspired a new analysis trend of cultural, social, and linguistic phenomena from a temporal perspective. Language is dynamic and evolves, it varies to reflect the shift in topics we talk about, which in turn follows cultural changes.

So far, the automatic analysis of language has been based on datasets that represented a snapshot of a given domain or a time period (synchronic approach). However, since the rise of big data, which has made large corpora of data spanning several periods of time available, the diachronic analysis of language has emerged as a new approach to study linguistic and cultural trends over time by analysing these new sources of information.

This large amount of data needs efficient computational approaches. In this scenario, Distributional Semantic Models (DSMs) represent a promising solution. DSMs are able to represent words as points in a geometric space, generally called WordSpace [1, 2] simply analysing how words are used in a corpus. However, a WordSpace represents a snapshot of a specific corpus, and it does not take into account temporal information.

In our proposal, we describe a tool [3, 4] that enables the analysis of the time evolution of the meaning of a word, called Temporal Random Indexing (TRI). TRI is able to build WordSpaces taking into account temporal information. We exploit this methodology in order to build geometrical spaces of word meanings that consider several periods of time. The TRI framework provides all the necessary tools for building WordSpaces over different time periods and performing such temporal linguistic analysis.

Previous methods for building WordSpaces are able to create separate spaces for each time period, but generally these spaces are not directly comparable. These approaches [5, 6] require complex mathematical operations to track semantic changes over different vector spaces. Conversely, TRI is able to build several WordSpaces for each different period, and it makes all these spaces comparable to each other by enabling the analysis of word meaning changes over time as simple vector operations.

In particular, TRI allows the analysis of semantic changes of words over time through a Distributional Semantic Model built through an algebraic technique called Random Indexing (RI) [7]. RI is a locality-sensitive hashing approach based on random projection and it is able to preserve the distance between points in a geometric space by projecting them in a smallest space through a random matrix [8]. RI has the advantages of being very simple, since it is based on an incremental approach, and easily adaptable to the requirements of a temporal analysis.

The WordSpace is built taking into account word co-occurrences which, according to the distributional hypothesis [9], states that words sharing the same linguistic contexts are related in meaning. In our case, the linguistic context is defined as the words that co-occur in the same period with the target (temporal) word, i.e. the word under the temporal analysis.

The idea behind RI has its origin in Kanerva work [10] about Sparse Distributed Memory. RI assigns a random vector to each context unit, in our case represented by a word. The random vector is generated as a high-dimensional random vector with a high number of zero elements and a few number of elements equal to 1 or -1 randomly distributed over the vector dimensions. Vectors built using this approach generate a nearly orthogonal space. During the incremental step, a vector is assigned to each temporal word as the sum of the random vectors representing the context in which the temporal element is observed. In our case the target element is a word, while the contexts are the other co-occurring words that we observe in the same time period by analysing a large corpus of documents. Finally, we compute the cosine similarity between the vector representations of word pairs in order to compute their relatedness. The detailed description of the TRI method is reported in [3].

The TRI tool is a simple and effective way for building geometrical spaces of concepts by skimming through huge corpora of text in order to learn the context of usage of words. Moreover, TRI is able to encode temporal information into the WordSpace and provide tools for analysing when words change their meanings.

TRI is available as an open-source software¹ and it has been evaluated on several domains, such as a collection of Italian books from the Gutenberg project [1], English scientific papers [1] and the Italian version of the Google N-gram dataset [2].

An interesting domain, not yet explored by TRI, is represented by social media such as Twitter or Facebook. In this domain, words can rapidly change their meanings according to social/cultural phenomena or particular events, and new words can appear. A particular case in this direction is represented by hashtags, neologisms or slang.

In this proposal, we want to summarize some results obtained by the usage of TRI on Italian corpora and we show some preliminary results about the analysis of a collection of tweets that spans about three years.

2. EVALUATION ON ITALIAN CORPORA

The first collection consists of Italian books, with publication year, made available in text format by the Project Gutenberg². The total number of collected books is 349 ranging from year 1810 to year 1922. All the books are processed using our tool TRI by creating a WordSpace for each available year in the dataset. For our analysis we created two macro temporal periods, before 1900 (T_{pre900}) and after 1900 ($T_{post900}$). The space T_{pre900} contains information about the period 1800-1899, while the space $T_{post900}$ contains information about all the documents in the corpus.

T_{pre900}	$T_{post900}$
libertà	libertà
opera	gloria
pari	giustizia
comune	comune
gloria	legge
nostra	pari
causa	virtù
italia	onore
giustizia	opera
guerra	popolo

Table 1: Neighbourhood of patria (homeland)

As a first example, we analyse how the neighbourhood of the word *patria* (*homeland*) changes in T_{pre900} and $T_{post900}$. Table 1 shows the ten most similar words to *patria* in the two periods; differences between them are reported in bold.

Some words (*legge*, *virtù*, *onore*)³ related to fascism propaganda occur in $T_{post900}$, while in T_{pre900} we can observe some concepts (*nostra*, *causa*, *italia*)⁴ probably more related to independence movements in Italy.

The second use case has the aim of finding words that change their meaning over time. As an example, analysing word meaning evolution over time, we observed that the word *cinematografo* (*cinema*) clearly changes its semantics: the cosine similarity of the word *cinematografo* with itself in the two spaces is very low, about 0.40.

To understand this change we analysed the neighbourhood in the two spaces and we noticed that the word *sonoro* (*sound*) is strongly related to *cinematografo* in $T_{post900}$. This phenomenon can be ascribed to the sound introduction after 1900.

¹ <https://github.com/pippokill/tri>

² <http://www.gutenberg.org/>

³ In English: (law/order, virtue, honor).

⁴ In English: (our, reason, Italy).

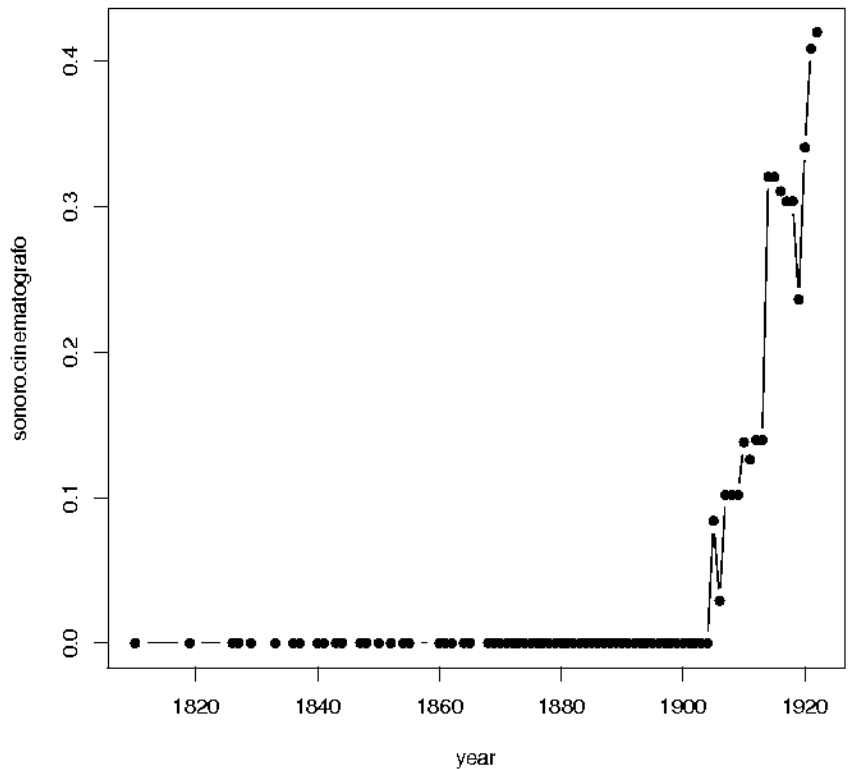


Figure 1: Word-to-word similarity variation over time for Sonoro (sound) and Cinematografo (cinema) in the Gutenberg dataset.

This behaviour is highlighted in Figure 1 in which we plot the cosine similarity between *cinematografo* and *sonoro* over the time. This semantic relatedness starts to increase in 1905, but only in 1914 we observe a substantial level of relatedness between the two terms. We report in Figure 2 a similar case between the words *telefono* (*telephone*) and *chiamare* (*call, as verb*). Their relatedness starts to increase in 1879, while a stronger level of cosine similarity is obtained after 1895.

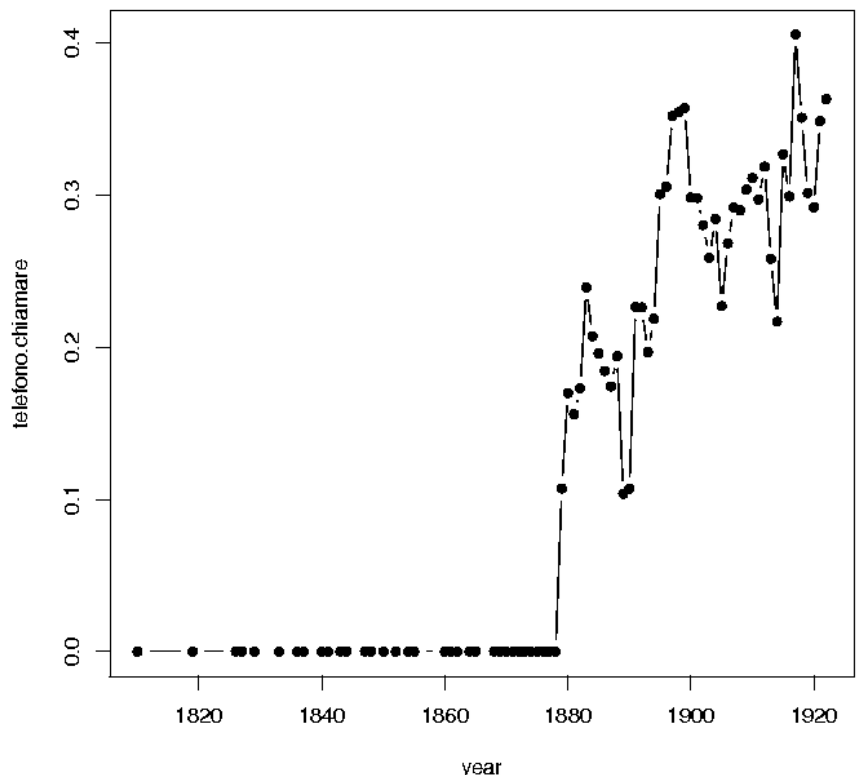


Figure 2: Word-to-word similarity variation over time for Telefono (telephone) and Chiamare (call) in the Gutenberg dataset

Moreover, we provide a further Italian evaluation [4] with the following aims: 1) build a standard benchmarking for meaning shift detection for the Italian language; 2) evaluate the performance of the proposed methods and compare them with the baseline model based on the word frequency.

A list of meaning shifts for the Italian language is not available, then we build a new dataset. We select 50 words using a polling strategy and manually check for each word if a meaning shift occurs by exploiting some dictionaries. We use two dictionaries: the “*Sabatino Coletti*” and the “*Dizionario Etimologico Zanichelli*”. Finally, we obtain a gold standard that consists of 40 words and their corresponding change points. We adopt accuracy as evaluation metric. Given a list of n change points returned by the TRI system, we compute the ratio between the number of change points correctly identified in the gold standard and n .

In order to identify the correct change points, we consider not only the word, but also the year of the change point. In particular, the year predicted by the system must be equal or greater than one of the years reported in the gold standard. We compute the accuracy using different values of n (10, 100, ALL). The analysis of the results shows that TRI provides better results than the simple baseline based on word frequency. A detailed description of the results is reported in [4].

3. PRELIMINARY EVALUATION ON ITALIAN TWEETS

In this preliminary evaluation, we run TRI on a collection of Italian tweets. In particular, the collection spans more than three years, from February 2012 to September 2015, and consists of about 500 million tweets. We build a WordSpace for each month by exploiting the TRI tool. In the context of social media we need to use a shorter time period in order to capture words (and hashtags) that rapidly change their meanings.

The aim of the first analysis is to trace how the semantic relatedness of a particular hashtag changes over time with respect to an emoticon. For example, if we choose as emoticon “:)” or “:(“ we can trace the sentiment (positive or negative) with respect to a particular topic.

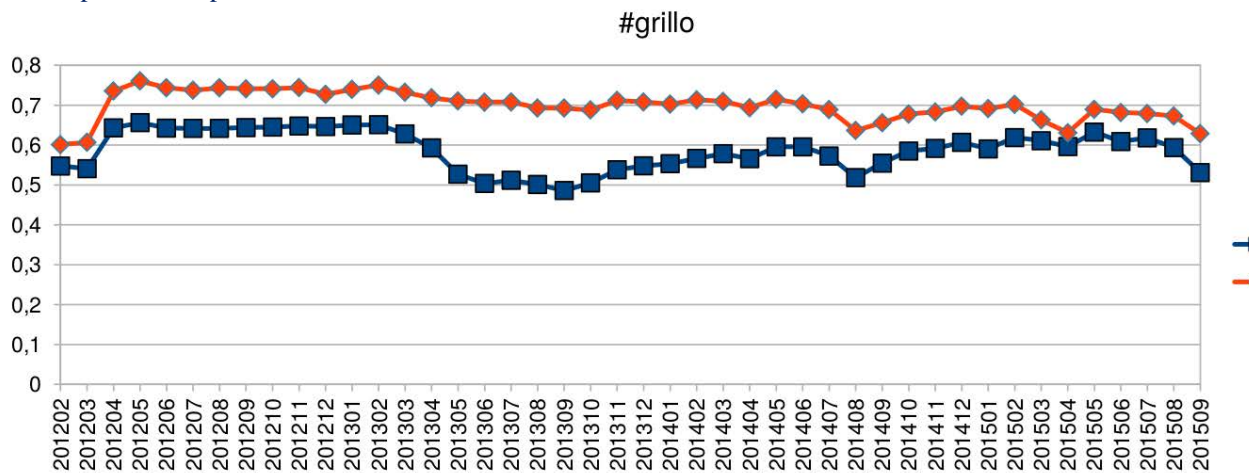


Figure 3: Semantic Relatedness between #grillo and positive/negative emoticons over time.

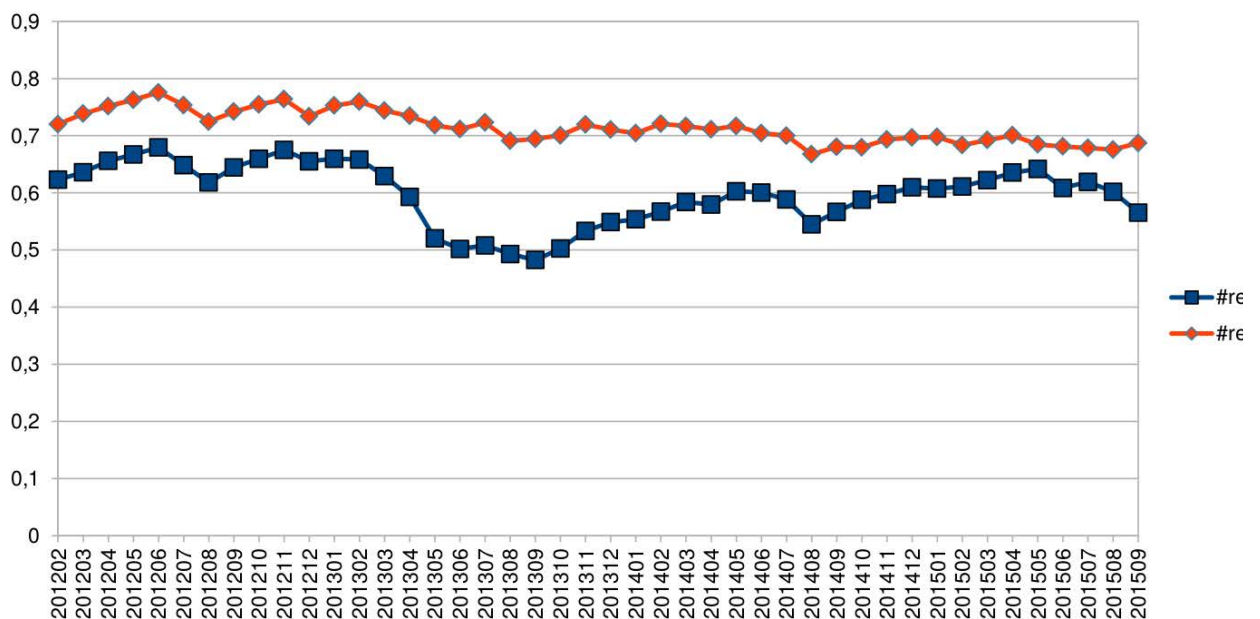


Figure 4: Semantic Relatedness between #renzi and positive/negative emoticons over time.

Figure 3 and 4 shows how the cosine similarity between #grillo and #renzi changes over time with respect to the positive/negative emoticons. In both cases the two hashtags are more related to a negative sentiment.

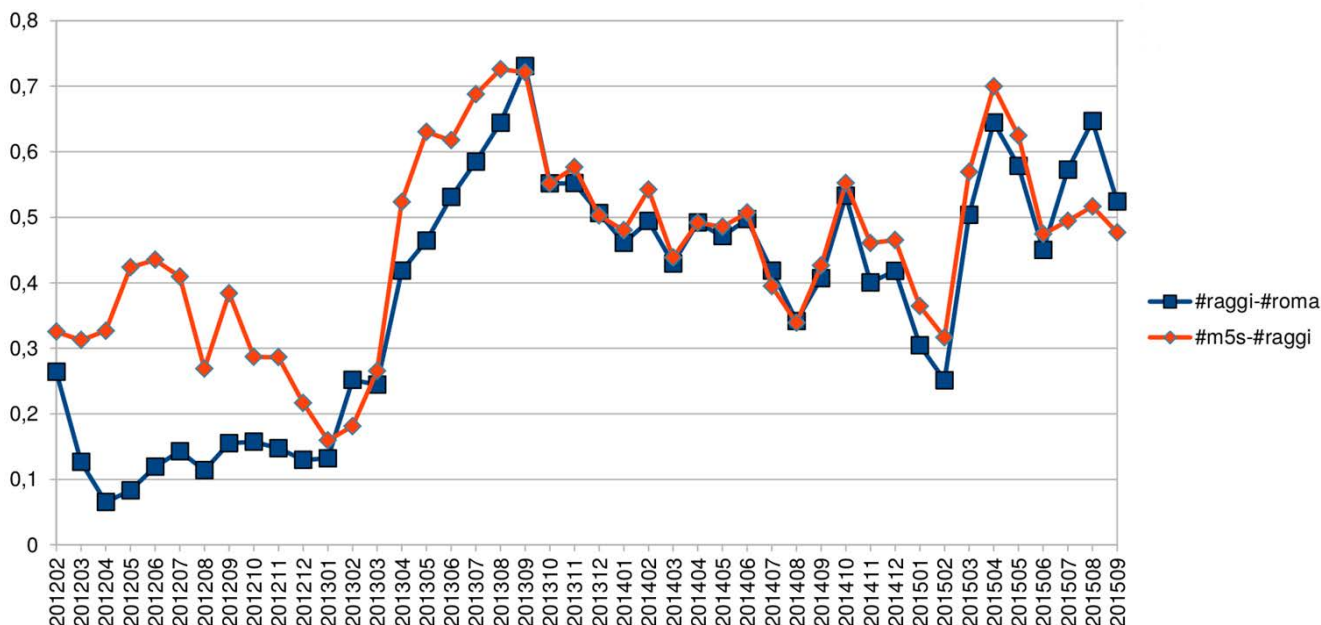


Figure 5

Figure 5 shows some interesting analyses about two pair of hashtags: #raggi-#roma and #m5s-#raggi. This plot depicts some interesting trends:

- In the first period #raggi is more related to #m5s than #roma, most probably because Virginia Raggi was not involved in the politic scenario of Rome municipality at that time;
- There is a rising trend in April-September 2013 during the administrative election in Rome, where Raggi was candidate as councilor;
- After Raggi was appointed as councilor, and the following crisis of the mayor Marino, she was presented as candidate for mayor in the next election and her relatedness with #roma increases again.

This preliminary case study of TRI shows interesting analyses that can be conducted in the context of social media for understating how the shift in meaning is correlated to particular social-cultural events. Similar analyses can be performed on any collection of documents with temporal information.

4. REFERENCES

- [1] Hinrich Schütze. 1993. Word space. *Advances in neural information processing systems*, 5:895–902, 1993.
- [2] Magnus Sahlgren. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, 2006.
- [3] P. Basile, A. Caputo, and G. Semeraro. Temporal random indexing: A system for analysing word meaning over time. *IJCoL*, 1(1), 2015.
- [4] P. Basile, A. Caputo, R. Luisi, G. Semeraro. Diachronic Analysis of the Italian Language exploiting Google Ngram. In *Proceedings of CLiC-it & EVALITA 2016, CEUR-WS vol. 1749*, 2016.
- [5] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. *CoRR*, abs/1605.09096, 2016.
- [6] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. ACM, 2015.
- [7] Sahlgren, Magnus. *An introduction to random indexing*, 2005
- [8] Dasgupta, Sanjoy, and Anupam Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *International Computer Science Institute, Technical Report (1999): 99-006*
- [9] Harris, Z.S.: *Mathematical Structures of Language*. New York: Interscience (1968)
- [10] Kanerva, P.: *Sparse Distributed Memory*. MIT Press (1988)

TriMED: banca dati terminologica multilingue

Federica Vezzani¹, Giorgio Maria Di Nunzio², Geneviève Henrot¹

¹ Dip. di Studi linguistici e letterari, Università di Padova, Italia - federica.vezzani@phd.unipd.it, genevieve.henrot@unipd.it

² Dip. di Ingegneria dell'Informazione, Università di Padova, Italia - giorgiomaria.dinunzio@unipd.it

ABSTRACT

Tre categorie di persone si confrontano con la complessità del linguaggio medico, ciascuna con le proprie esigenze di rimedio: medici, traduttori tecnico scientifici e pazienti. Il presente lavoro propone di elaborare uno strumento che contribuisca a porre rimedio all'opacità che caratterizza la comunicazione in ambito medico tra i suoi vari attori: soddisfare la comunicazione tra pari, fornire una risorsa regolarmente aggiornata ai traduttori tecnico scientifici e facilitare la comprensione delle informazioni da parte del grande pubblico: una risorsa terminologico-fraseologica multilingue. La banca dati si compone di schede terminologiche progettate per creare un ponte fra i vari registri individuati (specialistico, semi-specialistico, non specialistico) nelle lingue considerate. Limitatamente al settore oncologico dei trattamenti per il cancro al seno, i termini da trattare sono estratti da un corpus in lingua inglese, corredati di tutte le informazioni e le proprietà linguisticamente rilevanti, e ricollegati al loro equivalente pragmatico in italiano e in francese.

PAROLE CHIAVE

Linguaggio medico, terminologia, traduzione, metodologie e strumenti per la costruzione di risorse linguistiche.

1. INTRODUZIONE E STATO DELL'ARTE

Il presente intervento si inserisce nel contesto della valutazione della comunicazione in ambito sanitario e nell'analisi del linguaggio medico. Esso si caratterizza da una nomenclatura rigida e chiusa, tale da produrre un lessico opaco e carico di termini di difficile comprensione. Tra le problematiche in esso riscontrabili, le ambiguità semantiche derivano dal fenomeno della polisemia, dall'uso non corretto dei suffissi, dal mantenimento di arcaismi e dalle incongruenze etimologiche.

Numerosi studi hanno dimostrato la difficoltà a comprendere le informazioni medico-sanitarie contenute, ad esempio, nei foglietti illustrativi dei medicinali (PATEL et al., 2002), nei siti web (MCCRAY, 2005; CENTER, 2008) e più in generale nella comunicazione tra medici e pazienti (MCCRAY, 2005; JUCKS e BROMME, 2007; TRAN et al., 2009).

I pazienti sono spesso esposti ad una terminologia medica complessa: numerose ricerche pongono il proprio focus sul concetto di "comprensibilità". Lo studio di GRABAR *et al.* del 2014, ad esempio, muove un primo passo verso la semplificazione dei testi medici, in quanto propone una raccolta terminologica specifica per la valutazione di termini che rischiano di non venire compresi e che quindi richiedono ulteriori delucidazioni.

In questo quadro, ci proponiamo di costruire uno strumento terminologico multilingue che possa risolvere, per le diverse categorie di utenti individuate, i problemi correlati alla comunicazione in ambito medico.

2. TRE CATEGORIE DI UTENTI

In termini di divulgazione e diffusione delle conoscenze scientifiche, abbiamo identificato tre categorie di persone che, secondo diversi aspetti e a diversi livelli di comunicazione, possono risentire dei problemi collegati all'opacità del lessico medico: medici, traduttori tecnico-scientifici e pazienti.

2.1 MEDICI

La divulgazione su scala internazionale delle conoscenze mediche implica che la maggior parte dei testi scientifici vengano redatti in lingua inglese. In termini di diffusione di nuovi protocolli di cura e scoperte scientifiche, la lingua potrebbe pertanto ostacolare le transazioni di servizio tra colleghi medici specialisti parlanti lingue diverse, in quanto la perfetta conoscenza e padronanza della lingua straniera non è un dato scontato.

Ad un livello di comunicazione tra pari, e quindi specialistico-specialistico, ai medici serve poter valicare queste barriere linguistiche e accedere alle informazioni sulla ricerca scientifica nella propria lingua madre, in modo da importare le nuove conoscenze sul territorio nazionale, ma anche da esportare le proprie scoperte scientifiche inserendole nel circuito internazionale.

2.2 TRADUTTORI TECNICO-SCIENTIFICI

Nella sfera della comunicazione e della divulgazione delle informazioni, la traduzione, intesa come disciplina pratica, acquisisce un ruolo fondamentale per la corretta trasmissione di informazioni in lingue diverse. Affinché il testo tradotto risponda al principio deontologico cardine della disciplina, ovvero quello della fedeltà al testo di partenza, il traduttore tecnico-scientifico si trova a dover procedere tramite le fasi di decodificazione e transcodificazione del linguaggio medico al fine di riuscire a tradurre ciò che effettivamente comprende. Si rientra, in questo caso, in un livello di comunicazione specialistico-semi-specialistico.

Tuttavia, le esigenze del mercato della traduzione non concedono il tempo di condurre ricerche terminologiche approfondite, costringendo quindi il professionista a saltare alcuni passaggi chiave per un lavoro ottimale. Al traduttore servono risorse terminologiche regolarmente aggiornate che lo assecondino nella creazione del prodotto finale.

2.3 PAZIENTI

Lo sviluppo scientifico e tecnologico, che tanto influisce sulla medicina e sulla sua capacità diagnostica e terapeutica, ha progressivamente spostato il focus dell'attenzione dei medici non più sul paziente ma sulla stessa malattia, e questo ha condotto ad una crisi nel rapporto medico-paziente: il grande pubblico riscontra una difficoltà notevole nel comprendere informazioni relative alla propria salute, nonostante leggi e politiche enfatizzino la reale necessità di documentare i vari aspetti sanitari in modo più comprensibile¹.

L'interazione medico-paziente implica un livello di comunicazione specialistico-non specialistico, per cui, affinché la comunicazione sia efficace, i pazienti (o più in generale il grande pubblico) necessiterebbero di comprendere il termine medico specializzato tramite l'utilizzo di una sua riformulazione in lingua "popolare" o, più in generale, tramite l'uso di un linguaggio adeguatamente tarato sul livello dell'interlocutore non-specialista.

3. BANCA DATI TRIMED

Le necessità precedentemente delineate per le tre categorie di utenti così identificati trovano una valida risposta in una banca terminologica multilingue. Ci siamo proposti di sviluppare un database terminologico multilingue che raccolga un insieme di schede terminologiche disegnate e compilate per ciascun termine selezionato, a ciascun livello pragmatico individuato. Disponibile gratuitamente in rete, la risorsa prende il nome di TriMED, in quanto il carattere di tripartizione è insito nello strumento: tre sono le lingue di lavoro (inglese, francese e italiano), tre sono le categorie di utenti identificati e, di conseguenza, tre sono i livelli di comunicazione oggetto di analisi dello strumento terminologico.

3.1 STRUTTURA

Le schede di TriMED si articoleranno su tre livelli di comunicazione:

- specializzato, fornendo sia le definizioni scientifiche per soddisfare la comunicazione tra specialisti del settore, sia il corrispondente termine tradotto per rispettare i livelli di registro linguistico;
- semi-specializzato, fornendo informazioni utili al traduttore tecnico-scientifico per la traduzione di testi settoriali;
- non-specializzato, fornendo sia definizioni divulgative per facilitare la corretta comprensione da parte del grande pubblico, sia l'equivalente del termine comunemente utilizzato nel linguaggio popolare.

Le schede terminologiche mirano, quindi, a rendere il termine medico limpido e semanticamente accessibile anche nell'ottica della variazione diastratica.

3.1.1 RACCOLTA DATI

In questa fase, abbiamo compilato il corpus di partenza e i corpora paralleli e abbiamo estratto i termini tecnici.

Prima di tutto, abbiamo selezionato un insieme di articoli in lingua inglese per costituire il nostro corpus di partenza. I documenti sono stati attinti a riviste specialistiche online quali: *Breast Cancer research and treatment*², *Archives of Breast Cancer*³, e *European Journal of cancer care*⁴. Successivamente, abbiamo compilato dei corpora paralleli per le altre due lingue prese in analisi: l'italiano e il francese. Le fonti da cui abbiamo attinto i documenti in lingua italiana sono: Fondazione Umberto Veronesi (in particolare i protocolli dell'iniziativa "Pink is Good"), AIMaC - Associazione Italiana Malati di Cancro, e A.N.D.O.S - Associazione nazionale donne operate al seno. Mentre per la lingua francese: *Fondation du Cancer du Sein du Québec*, *l'Association francophone pour les soins oncologiques de support* (AFSOS) e *Cairn.info*. In seguito, abbiamo estratto manualmente i termini medici per i tre corpora scelti data l'entità contenuta del corpus di partenza. Non ci siamo avvalsi del sussidio di altre risorse terminologiche disponibili sul mercato in quanto lo scopo di TriMED è quello di raccogliere e analizzare i termini presenti nei corpora di partenza selezionati.

3.2 APPLICAZIONE WEB

In questa sezione, applichiamo la metodologia precedentemente delineata tramite la creazione di un'applicazione Web. Per ogni termine tecnico, abbiamo provveduto a fornire le seguenti informazioni:

- equivalente in lingua popolare e divulgativa
- definizione
- analisi semica del termine
- caratteristiche formali.

Su queste informazioni abbiamo disegnato l'interfaccia TriMED che quindi si presenterà come uno strumento "a tre teste": dalla pagina iniziale, si potrà selezionare la categoria nella quale l'utente si identifica (medico, traduttore, paziente) e si potrà accedere alle relative informazioni in base alle proprie esigenze. Abbiamo implementato l'applicazione con Shiny R package (CHANG, 2015): la demo è disponibile online per mostrare il funzionamento dell'interazione fra i tre livelli di comunicazione⁵.

¹ <https://www.kff.org/medicaid/report/ensuring-linguistic-access-in-health-care-settings-2/>

² <http://www.springer.com/medicine/oncology/journal/10549>

³ <http://archbreastcancer.com/>

⁴ [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-2354](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2354)

⁵ <https://gmdn.shinyapps.io/TriMED/>

Accedendo all'interfaccia nella modalità "paziente" (Immagine 1), è possibile ricercare il termine tecnico (con suggerimento di completamento del termine) e fornire il corrispettivo in lingua popolare e la sua definizione. Nell'ottica della semplificazione del linguaggio medico, questo strumento permette quindi al paziente di "tradurre" il termine medico con una terminologia semplificata o, comunque, comunemente usata. Il paziente potrà visualizzare anche la definizione dello stesso termine. Inoltre, considerando le linee programmatiche ternarie (specialisti, semi-specialisti, non-specialisti), abbiamo pensato di fornire la possibilità di consultazione di una variante divulgativa per il termine in questione, la quale è da considerarsi diversa dal termine in lingua popolare.

Questa implementazione è ancora in fase di sviluppo e accertamento: distinguiamo tra

- termine popolare ricorrente prevalentemente in discorsi orali presso i pazienti (per esempio nelle sale d'attesa di ambulatori e ospedali), e
- termine semi-popolare (o divulgativo) usato in articoli divulgativi e che quindi estratto da corpus scritto.

Figura 1 Visualizzazione dati per l'utente di tipo paziente

Per quanto riguarda la consultazione dell'interfaccia nella modalità "traduttore" (Figura 2 e 3), sono fornite all'utente le informazioni prettamente linguistiche e terminologiche che stanno alla base del processo traduttivo.

Dopo aver selezionato il termine, l'utente può accedere ad una schermata che si presenterà sotto forma di una tabella con duplice lettura: verticale e orizzontale.

Procedendo con un'analisi verticale, l'utente potrà consultare la traduzione del termine nella sua versione di linguaggio scientifico e popolare.

Figura 2 Visualizzazione dati per l'utente di tipo traduttore

In seguito, verrà fornita la definizione del termine e la relativa analisi semica: tale operazione ha lo scopo di scomporre il significato dell'unità lessematica o morfematica in unità atomiche non ulteriormente segmentabili.

Infine l'utente può accedere alle caratteristiche formali del termine necessarie per il suo inquadramento lessicale:

- genere;
- pronuncia;
- derivazione e composizione del termine.

<p>Definition</p> <p>An abnormal new growth of disorganized tissue in animals or plants.</p>	<p>Définition</p> <p>Formation pathologique d'un tissu nouveau; résultat de cette formation.</p>
<p>Analysis</p> <p>/abnormal//growth//cell//disorganized//tissue//human/</p>	<p>Analyse</p> <p>/formation//pathologie//tissu//corps//cellule//croissance//anormalité/</p>
<p>Features</p> <p>Formal features: Part of speech: n. Pronunciation: /'ni:əu.plæzəm/ Other: From neo- + -plasm, after German Neoplasma.</p>	<p>Caractéristiques</p> <p>Caractéristiques formelles: Genre: s.f Dérivation/composition: Formé des élém. néo-* et -plasie*.</p>

Figura 3 Visualizzazione dati per l'utente di tipo traduttore

La trasmissione di queste informazioni aiuta il traduttore/la traduttrice a scegliere il candidato traduttore del termine preso in analisi.

Mantenendo la tabella con il termine di partenza, il traduttore potrà impostare la lingua di arrivo e consultare le informazioni del traduttore scelto che appariranno in una tabella adiacente a quella di partenza. In questo modo, l'utente può accedere alle stesse informazioni per il traduttore selezionato e potrà consultare orizzontalmente le informazioni per i due termini.

Infine, la modalità di accesso come utente “specialista” (Figura 4), si presenta come un'interfaccia che offre nell'ottica traduttologica l'opportunità al medico di consultare il termine nella propria lingua madre. Anche questo utente potrà selezionare la lingua di partenza e la lingua di arrivo in cui desidera consultare il termine stesso e la definizione ad esso correlata.

Per questa categoria di utente, è stato previsto anche un collegamento diretto con il MeSH term ad esso correlato. Cliccando sul termine in questione, si potrà quindi accedere alle varie informazioni fornite direttamente dalla National Library of Medicine⁶.

<p>TRIMed Patient Translator Physician</p>	
<p>Technical Term</p> <p>Erythrocyte</p>	<p>Terme technique</p> <p>Erythrocyte</p>
<p>Informative Term</p> <p>Red blood cell</p>	<p>Terme vulgarisateur</p> <p>Globule rouge</p>
<p>Definition</p> <p>Erythrocytes are the most common type of blood cell and the vertebrate's principal means of delivering oxygen (O2) to the body tissues—via blood flow through the circulatory system.</p>	
<p>MeSH Term: Erythrocytes</p>	

Figura 4 Visualizzazione dati per l'utente di tipo medico

⁶ <https://www.nlm.nih.gov/mesh/>

4. CONCLUSIONI

La nostra analisi delle problematiche comunicative in ambito medico ci ha portati a concepire una risorsa linguistica che fosse di aiuto per le categorie di utenti. Attualmente TriMED comprende un totale di 200 termini per le lingue francese e italiano. Stiamo lavorando sulla versione inglese in modo da adempiere agli obiettivi multilingue precedentemente delineati.

Il nostro lavoro, benché agli inizi, copre la terminologia medica nel settore oncologico e cerca di proporre una risorsa che non comprenda esclusivamente la lingua inglese ma anche altre lingue di lavoro che necessitano di terminologia documentata nel settore. Il nostro obiettivo è quindi quello di fornire uno strumento terminologico che comprenda 2500-3000 termini per questo settore specifico. Attualmente l'estrazione terminologica è stata compiuta manualmente ma prevediamo di semi-automatizzare questo processo così come l'allineamento di termini in traduzione: questo ci condurrà anche a considerazioni critiche sull'affidabilità dell'estrazione automatica in termini di granularità e pertinenza.

I dati, che attualmente sono salvati in formato tabellare, verranno modellati secondo i principi di progettazione di una base di dati relazionale. Questo tipo di organizzazione permetterà l'esportazione dei dati in vari formati. Uno degli obiettivi di questa applicazione web è quello di permettere ai traduttori tecnico-scientifici di esportare i dati in formato TermBase Exchange (TBX) che è lo standard internazionale per la rappresentazione in XML di dati terminologici⁷, in modo da seguire i principi del FAIR (Findable Accessible Interoperable Re-usable) Open Data della European Open Science Cloud⁸.

La struttura e l'idea che stanno alla base di TriMED lasciano spazio ad implementazioni future. Ad esempio, per quanto riguarda l'arricchimento della modalità "medico", ci proponiamo di permettere all'utente la visualizzazione di articoli provenienti dalle riviste scientifiche correlati a quel preciso termine tecnico. Per i termini in lingua inglese potrebbe essere utile un collegamento diretto tra i Mesh terms che rimandi a sua volta agli articoli di PubMed correlati al termine stesso. Mentre per i termini in lingua italiana, il collegamento potrebbe essere fatto con la banca dati bilingue dei Mesh a cura dell'Istituto superiore di sanità di Roma e successivo collegamento agli articoli di PubMed.

Allo stesso modo, per la terminologia francese si potrebbe procedere con il collegamento diretto al sito dell'Inserm⁹ che in cooperazione con l'Inist-CNRS contribuisce all'aggiornamento della versione francese dei Mesh terms dal 2004.

Inoltre, considerata la reale utilità di una risorsa terminologica di questo tipo pensata nell'ottica della valutazione multilingue del linguaggio medico, ci proponiamo di ampliare le lingue di lavoro includendo anche lingue extra-europee e di estendere il settore d'interesse all'intera sfera dell'oncologia e quindi non limitandoci allo studio delle cure per il cancro al seno.

5. BIBLIOGRAFIA

- [1] BOUAMOR, D., LLANOS, L. C., LIGOZAT, A.-L., ROSSET, S., and ZWEIGENBAUM, P. (2016). Transferbased learning-to-rank assessment of medical term technicality. In Proceedings of the LREC 2016, Portorož, Slovenia, May. ELRA.
- [2] CENTER, O. P. (2008). Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. technical report, agency for healthcare research and quality. Oregon Evidence-based Practice Center.
- [3] CHANG, W., (2015). Shiny: Web Application Framework for R. R package version 0.11.
- [4] ELHADAD, N. and SUTARIA, K. (2007). Mining a lexicon of technical terms and lay equivalents. In Proc BioNLP Workshop, pages 49-56. ACL.
- [5] FRANÇOIS, T. and FAIRON, C. (2013). Les apports du TAL à la lisibilité du français langue étrangère. TAL, 54(1) : 171-202.
- [6] GRABAR, N., VAN ZYL, I., DE LA HARPE, R., and HAMON, T. (2014). The comprehension of medical words. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5, BIOSTEC 2014, pages 334-342, Portugal. SCITEPRESS - Science and Technology Publications, Lda.
- [7] JUCCS, R. and BROMME, R. (2007). Choice of words in doctor-patient communication: an analysis of health-related internet sites. Health Commun, 21(3):267-77.
- [8] KESELMAN, A., TSE, T., CROWELL, J., BROWNE, A., NGO, L., and ZENG, Q. (2007). Assessing consumer health vocabulary familiarity: an exploratory study. J Med Internet Res, 9(1): e5.
- [9] LEROY, G., HELMREICH, S., COWIE, J., MILLER, T., and ZHENG, W. (2008). Evaluating online health information: Beyond readability formulas. In AMIA Annual Symposium Proceedings.
- [10] MCCRAY, A. (2005). Promoting health literacy. J of Am Med Infor Ass, 12:152-163. NEVEOL, A., GROSJEAN, J., DARMONI, S., and ZWEIGENBAUM, P. (2014). Language resources for French in the biomedical domain. In Proceedings of the LREC 2014, Reykjavik, Iceland, May. ELRA.
- [11] PATEL, V., BRANCH, T., and AROCHA, J. (2002). Errors in interpreting quantities as procedures: The case of pharmaceutical labels. International journal of medical informatics, 65(3):193-211.
- [12] SERIANNI, L. (2005). Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente. Milano, Garzanti Libri.
- [13] TRAN, T., CHEKROUD, H., THIERY, P., and JULIENNE, A. (2009). Internet et soins : un tiers invisible dans la relation médecine/patient. Health Commun, 21(3): 26777.
- [14] VITALI E.D., (1967), Considerazioni sul problema del linguaggio in medicina, Protagora.
- [15] ZENG, Q., KIM, E., and CROWELL, J. and TSE, T. (2005). A text corpora-based estimation of the familiarity of health terminology. In ISBMDA, pages 184-192, Aveiro, Portugal.

⁷ http://www.ttt.org/oscarstandards/tbx/tbx_oscar.pdf

⁸ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

⁹ <http://mesh.inserm.fr/mesh/>

“Two days we have passed with the ancients...”: a Digital Resource of Historical Travel Writings on Italy

Rachele Sprugnoli

Digital Humanities Group, Fondazione Bruno Kessler - sprugnoli@fbk.eu

1. INTRODUCTION

Non-fictional travel writings are powerful sources of information for many research areas, such as art history, ethnography, geography and cultural history (Burke, 1997). By collecting several books on the same place, it is possible to study how material and cultural aspects change over time (Schulz, 2012). Moreover, travel writings can give insight not only into the places and people visited by the traveller but also on his/her culture and worldviews (Quintana, 2011). Despite the importance of this type of texts, digital collections of travel writings are not easy to retrieve. Texts are often not available in a format that can be straightforwardly processed by a computer, or there is no direct download of the collection (see for example, “A Vision of Britain through Time”¹), or the documents are scattered in vast digital archives.

In order to overcome these limitations, we release the first version of a corpus of more than 2 millions words of historical English travel writings about Italy which we have retrieved from freely available sources (Project Gutenberg² and Project Gutenberg Australia) and we make them available in a cleaned text format and in TEI-XML through the following website: <https://sites.google.com/view/travelwritingsonitaly>.

2. CORPUS DESCRIPTION

The current version of the corpus is made by 30 texts including both travel narratives (reports, diaries, collections of letters) and guidebooks, for a total of 2,413,838 tokens. The authors, of both genders, belong to different nationalities (UK, US, Australia). Some books are focused on specific cities or regions, others cover different parts of Italy. All the books were published between the end of the XIX Century and the beginning of the XX Century. We choose this period because at that time the tradition of the Grand Tour decayed and leisure-oriented tourists travels emerged (Ellerbee, 2010). This radical transformation was due to innovations occurred in transportation, economics and ideology: e.g., the diffusion of the railway network and of steam-powered ships, the increase in wealth for many Americans, a greater emancipation of women that led to having more female travelers (Schriber, 1995). In addition, Italian unification allowed the opening of new routes to the South of the country that were not taken into account during the Romantic era (Ouditt and Polezzi, 2012).

Chosen texts were processed with GutenTag³ so to automatically obtain their XML version based on the TEI standard⁴ and then manually checked and integrated using the Oxygen editor⁵. In this latter phase, particular effort has been devoted to enhance the description of each file adding missing biographical information on the author and adopting an unique standard for text classification. To this end, we took advantage of the Online Computer Library Center web service⁶ to obtain the Library of Congress Classification class of each book and a set of FAST subject headings to be used as keywords (Dean, 2004). As for text structure, the main additions involved the annotation of poems by using the *<lg>* element and of the formulaic opening of letters with the *<opener>* element.

3. TOWARDS A GEOGRAPHICAL ANALYSIS OF TRAVEL WRITINGS

The plain text corpus has been processed with the Stanford CoreNLP (Manning et al., 2014) named entity recognizer to automatically detect place names and provide a preliminary analysis of the geographical dimension of travel writings. More than 4,600 locations were extracted and then georeferenced using Google's services. Table 1 shows the top 10 places with their frequency while Figure 1 displays the 200 most recurring locations.

¹ <http://www.visionofbritain.org.uk/travellers/>

² <http://www.gutenberg.org> - <http://gutenberg.net.au/>

³ <http://www.projectgutentag.org>

⁴ <http://www.tei-c.org/>

⁵ <https://www.oxygenxml.com/>

⁶ <http://classify.oclc.org/classify2/>

PLACE	FREQ
Rome	2287
Venice	1291
Florence	1136
Naples	769
Pisa	394
Genoa	343
Sicily	312
Siena	206
Vatican	174
Tuscany	172

Table 1. Top mentioned places.



Figure 1. Map of the 200 most mentioned locations.

In the map all regions are represented but travellers have focused their attention in particular on central Italy. As for the South, Figure 2 presents a zoom on the most frequent places in the Neapolitan area. Not much has changed with respect to the most popular contemporary tourist spots but it is worth noticing the use of ancient spelling of cities ("Puteoli", "Misenum", "Baiae") related to the strong interest for the historical/archeological aspects of travels at that time. A fine-grained analysis is also possible: as shown in Figure 3, CoreNLP can detect the names of palaces, churches and squares providing a detailed view of historical tourist attractions.

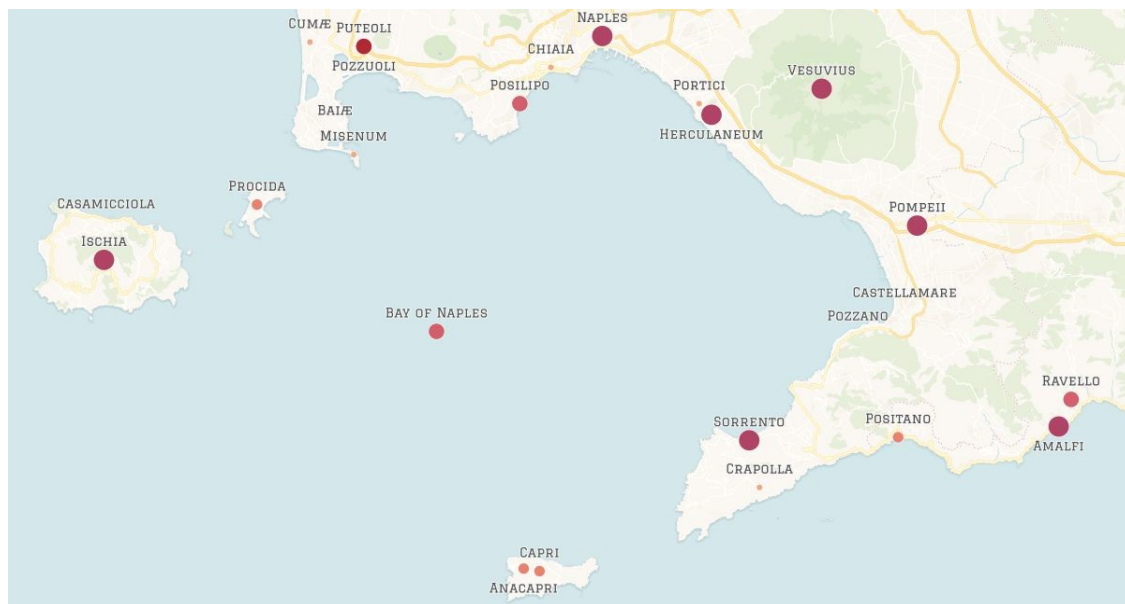


Figure 2. Most recurring place names in the Neapolitan area.



Figure 3. Locations extracted from "A Wanderer in Venice" (Lucas, 1914).

4. FUTURE WORK

In this proposal we describe a new digital resource for which many future works can be envisaged. First of all, the corpus could be expanded adding new texts or covering different periods and languages thus presenting further viewpoints on Italy. TEI annotation can be widened as well: the encoding of citations to determine the attribution of quotes would be particularly interesting so to study literary references and their correlation with the mentioned locations. The corpus could play a central role also in the field of Natural Language Processing applied to the Digital Humanities, opening up language technologies to a type of text not much explored so far. Recently the corpus has been used to detect Content Types, i.e. clauses with specific functional and semantic roles, and to analyse the phenomenon of languaging in travel discourse (Sprugnoli et al., 2017a; Sprugnoli et al., 2017b). As for the geographical analysis presented in the previous Section, it can be improved by retraining CoreNLP so to obtain a more accurate output: an evaluation performed on a subset of 10,000 tokens registered a precision (percentage of correct positive predictions by CoreNLP) of 89.15%, a recall (percentage of positive cases recognized by the system) of 78.98% and an F-measure (weighted harmonic mean of precision and recall) of 83.76% showing that there is room for improvements. Another aspect worth studying with language technologies would be the detection of stereotypes that are pervasive in travel writings and often persist nowadays in the perception of Italian people (Calaresu, 2007).

5. REFERENCES

- [1] Brooke, Julian, Adam Hammond, and Graeme Hirst. 2015. GutenTag: an NLP-driven Tool for Digital Humanities Research in the Project Gutenberg Corpus. *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*. Association for Computational Linguistics.
- [2] Burke, Peter. *Varieties of cultural history*. Cornell University Press, 1997.
- [3] Calaresu, Melissa. 2007. From the Street to Stereotype: Urban Space, Travel and the Picturesque in Late Eighteenth-Century Naples. *Italian Studies* 62.2: 189-203.
- [4] Colavizza, Giovanni, and Frédéric Kaplan. 2015. On Mining Citations to Primary and Secondary Sources in Historiography. *Proceedings of CLiC-it 2015*.
- [5] Dean, Rebecca J. 2004. FAST: Development of simplified headings for metadata. *Cataloging & Classification Quarterly* 39.1-2 (2004): 331-352.
- [6] Ellerbee, Genevieve. 2010. Voyage to Italia: Americans in Italy in the Nineteenth Century. *Sheldon Museum of Art Catalogues and Publications*. 83.
- [7] Lucas, Edward Verrall. 1914. *A wanderer in Venice*. The Macmillan Company.
- [8] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.
- [9] Ouditt, Sharon, and Loredana Polezzi. 2012. Introduction: Italy as place and space. *Studies in Travel Writing* 16.2 (2012): 97-105.

- [10] Quintana, Ángel Gurría. 2001. Travelling Through Discourse, Discoursing on Travel: Recent Writing on Travel Literature and British Travellers in Mexico. *Studies in Travel Writing* 5.1 (2001): 172-188.
- [11] Schriber, Mary Suzanne. 1995. Women's Place in Travel Texts. *Prospects*, (1995)20: 161-179.
- [12] Schulz, Janice. 2012. Travel Writings in the Archives and Rare Books Library. Online: <http://libapps.libraries.uc.edu/liblog/2012/04/travel-writings-in-the-archives-and-rare-books-library/>
- [13] Sprugnoli, Rachele, Tommaso Caselli, Sara Tonelli, and Giovanni Moretti. 2017a. The Content Types Dataset: a New Resource to Explore Semantic and Functional Characteristics of Texts. In *Proceedings of EACL 2017*.
- [14] Sprugnoli, Rachele, Sara Tonelli, Giovanni Moretti and Stefano Menini. 2017b. A little bit of bella pianura: Detecting Code-Mixing in Historical English Travel Writing. To appear in *Proceedings of CLiC-it 2017*.

Using Formal Ontologies for the Annotation and Study of Literary Texts

Fahad Khan¹, Gloria Mugelli², Federico Boschetti³, Francesca Frontini³, Andrea Bellandi¹

¹ ILC-CNR, Italy - firstname.secondname@ilc.cnr.it

² Università di Pisa, Italy - gloria.mugelli@gmail.com

³ PRAXILING UMR 5267 Univ Paul Valéry Montpellier 3, France - francescafrontini@gmail.com

ABSTRACT

In this poster we give a brief motivation for the use of ontologies in literary analysis and of the use of the semantic web based OWL ontology language. As part of this motivation we present a case study based on ongoing work on an ancient Greek corpus.

KEYWORDS

Linked open data, ontologies, ancient greek

1. INTRODUCTION

Perhaps the single most popular definition of the term ontology, as it is used nowadays in informatics, is that given by (Studer et al, 1998) in which an ontology is described as a “formal, explicit specification of a shared conceptualization.” Note in particular the presence of the modifier “shared” here: ontologies have the potential to play an important role in promoting interoperability across individual resources precisely because they (are supposed to) describe shared agreement over the ‘meaning’ of the terms of discourse used by a given community (or communities)¹ -- and this applies to scholarly communities, and within and across individual academic disciplines, just as much as in any other case. And although they don’t tend to be commonly used at present in the humanities, the success that ontologies have achieved in other domains seems to offer prima facie evidence for their suitability and potential utility in humanities disciplines too. Indeed, they could potentially facilitate the formalisation of a great deal of scholarly knowledge in the humanities, and render explicit, as well as -- and this is what makes ontologies especially powerful -- machine operationable, much that is currently implicit, assumed, or at least partially taken for granted.

2. ONTOLOGIES APPLIED TO THE STUDY OF LITERARY TEXTS

If we narrow our focus to the study of literary texts then a number of specialised applications of formal ontologies immediately suggest themselves. For instance they are especially well suited to giving ‘machine actionable’ descriptions of the formatting of texts, of story plots and narrative arcs, as well as of the various kinds of relationships (citational and allusional) between different literary works. Another important way in which they can be applied to this domain is through the use of annotation tagsets structured as ontologies. Structuring tagsets in this way this will help to codify much of a researcher’s underlying hypotheses and assumptions, especially those that have been inherited from her research discipline. It also has important practical benefits. Indeed by using taxonomies to structure tagsets, especially large tagsets, we are already able to write simpler and more concise queries than would be the case otherwise, i.e., without having to enumerate subclasses or instances of concepts. But of course ontologies allow us to do much more than just structure a domain using hyponymic relations as is the case with a simple taxonomy. We can also include ‘part_of’ relations, spatial and geographical properties and relations, or in fact almost any sort of relation, meaningfully defined within the bounds of the expressive limitations inherent in the particular formalism being used. Thus we can search for example for all texts tagged with parts of the body (without having to enumerate these or even to know which of these have been explicitly added to the tagset) or names of locations within the Sasanian empire. In this regard it is useful if the ontology in question is described in a formal language for which there exist reasoners for deriving new axioms and checking consistency. The immense popularity of the Web Ontology Language (OWL) as a language for Knowledge Representation and Reasoning can be attributed in large part to the ready availability of (open source) tools, both for reasoning and inference as well as for visualisation and editing, for that language. Furthermore the use of OWL gives us access to a whole ecosystem of Semantic Web standards, resources and technologies, and allows the possibility of linking both ontological tagset and texts to a rich semantically structured network of interlinked knowledge resources. In particular there exist numerous linked data bibliographic databases, gazetteers, authority databases, historical knowledge bases to which an ontology in OWL can be linked.

In the next section we present a case study based on ongoing work carried out by the authors of this article in which an OWL ontology has been integrated into a system for querying a corpus of Ancient Greek text.

¹ Clearly, the larger and more diffuse the community in question the more difficult it will be to find agreement on, or a shared understanding of, specific terms. In such cases, however, ontologies can help to facilitate interoperability between different, and perhaps ostensibly incompatible, sets of metadata by mapping their categories to a common, upper level, ontology. This is done by breaking down the meanings of such categories in terms of more basic semantic constituents, and using these latter constituents to create mappings to the upper level ontology. One such approach that has been successful in the past uses the CIDOC-CRM upper level ontology to model metadata records as propositions using ontological classes and relations (Doerr 2003).

*Euporia, Ritual in Ancient Greek Tragedy*² is an ongoing project for the annotation and retrieval of ritual and religious facts in ancient Greek tragic texts. The project was designed to offer digital support to historico-anthropological research on the form and function of rituals in ancient Greek tragedy. The first stage of the project involved the annotation of the corpus of 33 surviving ancient Greek tragedies by a classical scholar using an annotation interface based on the annotation practices of classicists (Mugelli et al. 2016). The tagset used in the annotation was defined by the expert herself using a bottom-up approach, and was kept open, with the possibility of modifications up until the very last phases of the annotation process. Moreover, it was completely flat: complex concepts were expressed by combining different tags. Afterwards, and in order to overcome the limitations of this flat representation the tagset was converted into an OWL ontology (Mugelli et al., 2017).

The result was a tagset ontology that essentially describes the ‘meaning’ of the tags in the tagset and the multiple kinds of semantic interrelations between them, and allows this information to be used in querying the textual data itself. Moreover the tagset takes on further value as a resource in its own right: as a scholarly ontology describing ritual practice in Ancient Greek society, one that, thanks to its status as a linked data dataset, can be easily re-used in part or in whole in other projects, and which will therefore help to link together other datasets.

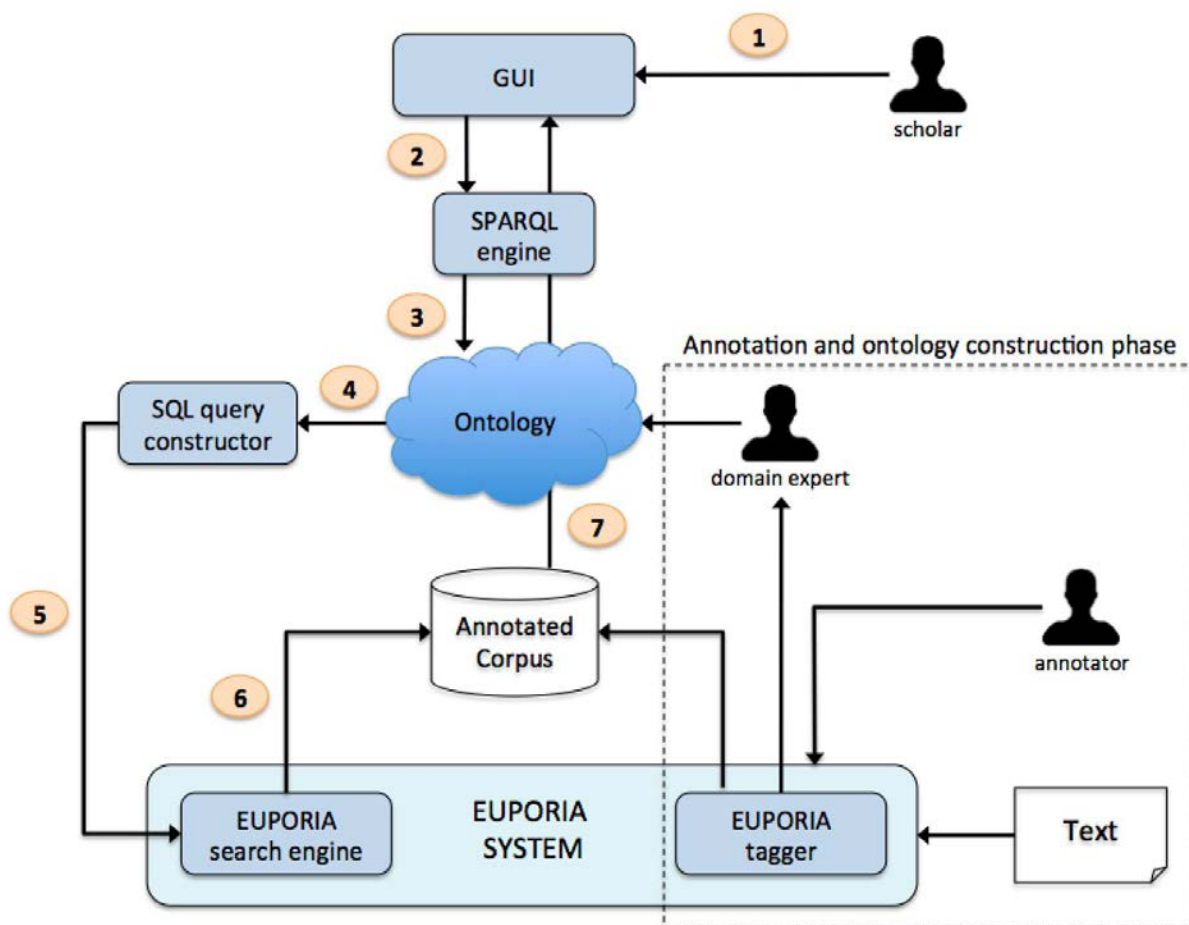


Figure 2. An overview of the Euporia Annotated Text Query System

In Figure 2 we present a schema for the entire corpus query system. The user interacts with the system via a GUI containing a series of forms and drop down menus; these assist her in the process of formulating her query to the system. Since the texts in the corpus are specifically annotated for information relating to religious and ritual events -- sacrifices in particular -- the user interface focuses on enabling the user to search texts for information pertaining to these events, their sub-events, their participants and the different roles they may play. The query entered by the user is subsequently translated into SPARQL and used to query the tagset ontology. The results of this query are used to create SQL queries to query the Euporia system³ and this allows for the expansion and enrichment of the user’s initial input queries via the ontology. One such expansion strategy is through hyponymic expansion/instance enumeration; others take advantage of the different

² See <http://www.himeros.eu/euporiaRAGT/>.

³ Note that in order to carry out queries on the annotated text the Euporia system needs to calculate the overlap between textual indices. For this reason we couldn’t directly use SPARQL to query the annotated corpus with Euporia.

ontological relations between and properties of concepts. In addition to allowing users to write more succinct queries, the ontology can also guide users by dynamically suggesting or disallowing categories while the user is in the process of entering queries into the interface. And since the ‘semantic’ information that feeds into the query interface is stored in the ontology it is easy to alter or update this information simply by updating this single resource. In future work we intend to link the tagset and the text to other Semantic Web datasets as was alluded to above, for instance Pelagios⁴.

3. REFERENCES

- [1] Doerr, M. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3), 75.
- [2] Doerr, M., Ore, C.-E., & Stead, S. 2007. The CIDOC conceptual reference model: a new standard for knowledge sharing. In *Proceedings Tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling* (pp. 51–56). Auckland, New Zealand: Australian Computer Society.
- [3] Khan, F., Arrigoni, S., Boschetti, F. and Frontini, F. 2016a. Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying. *MATLIT: Materialities of Literature*, 4(2): 11–27.
- [4] Khan, F., Bellandi, A., Benotto, G., Frontini, F., Giovannetti, E. and Reboul, M. 2016b. Leveraging A Narrative Ontology to Query a Literary Text. In Miller, B., Lieto, A., Ronfard, R., Ware, S. G. and Finlayson, M. A. (eds), *Proceedings of the 7th International Workshop on Computational Models of Narrative (CMN'16) - Krakow (Poland)*, vol. 53. (OpenAccess Series in Informatics (OASIS)). Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, pp. 124–133 doi:<http://dx.doi.org/10.4230/OASIS.CMN.2016.10>. <http://drops.dagstuhl.de/opus/volltexte/2016/6711>.
- [5] Mugelli, G., Boschetti, F., Del Gratta, R., Del Grosso, A. M., Khan, A. F., Taddei, A. 2016. A user-centred design to annotate ritual facts in ancient greek tragedies. *Bulletin of the Institute of Classical Studies* 59, 2.
- [6] Mugelli, G., Khan, A. F., Bellandi, A. and Boschetti, F. 2017. Designing an Ontology for the Study of Ritual in Ancient Greek Tragedy. In Frontini, F., Grčić Simeunović, L., Vintar, Š., Khan, A. F. and Parvizi, A. (eds), *Proceedings of the Workshop on Language, Ontology, Terminology and Knowledge Structures - LOTKS 2017*. Montpellier (France).
- [7] Studer, R., Benjamins, V. R. and Fensel, D. 1998. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2): 161–97 doi:10.1016/S0169-023X(97)00056-6.
- [8] Tambassi, T. (2017). *The Philosophy of Geo-Ontologies*. Springer
- [9] Tomasi, F., Ciotti, F., Daquino, M. and Lana, M. 2014. Using ontologies as a faceted browsing for heterogeneous cultural heritage collections. vol. *Proceedings of Digital Humanities 2014*. Lausanne, pp. 494–96.

⁴ <http://commons.pelagios.org/>

AIUCD 2018

Workshops

Data Mining and Image Analytics

Ajda Pretnar¹, Lan Žagar²

Faculty of Computer and Information Science, University of Ljubljana

¹ajda.pretnar@fri.uni-lj.si - ²lan.zagar@fri.uni-lj.si

KEYWORDS

image analytics, deep embeddings, neural networks, image maps, data mining

1. IMAGE ANALYTICS

Recent surge in digitization of material culture enabled access to a large amount of data to broader public (see Newell 2012). These data include digitized images of paintings, postcard, and photographs, which offer a rich source of information. But extracting this information can be painstaking and messy. Are researchers really required to sift manually through hundreds of images?

Not necessarily. This workshop will introduce the participants to data mining methods through the analysis of images. We will exploit pre-trained deep neural networks to represent images with profile vectors, cluster them and project them in 2D to find groups of images with similar motifs and styles. This is a hands-on workshop, where the participants will actively construct analytical workflows and go through case studies with the help of the instructors. Participants will also be given a task to apply the learnt techniques on their own set of images. At the end of the workshop, the participants will know how to use visual programming to seamlessly construct powerful data analysis workflows and apply them to the analysis of images.

2. STRUCTURE OF THE WORKSHOP

2.1. PART 1: VISUAL PROGRAMMING WORKFLOWS AND DATA INPUT

We will show the basics of visual programming in Orange, an open-source data mining and machine learning software (Demšar et al. 2013). We will prepare small image repositories, load them into Orange, inspect them in the data table and visualize them in an album-based visualisation. A particular emphasis will be on the structure of image data input, how it differs from a traditional spreadsheet data and how to properly analyse it.

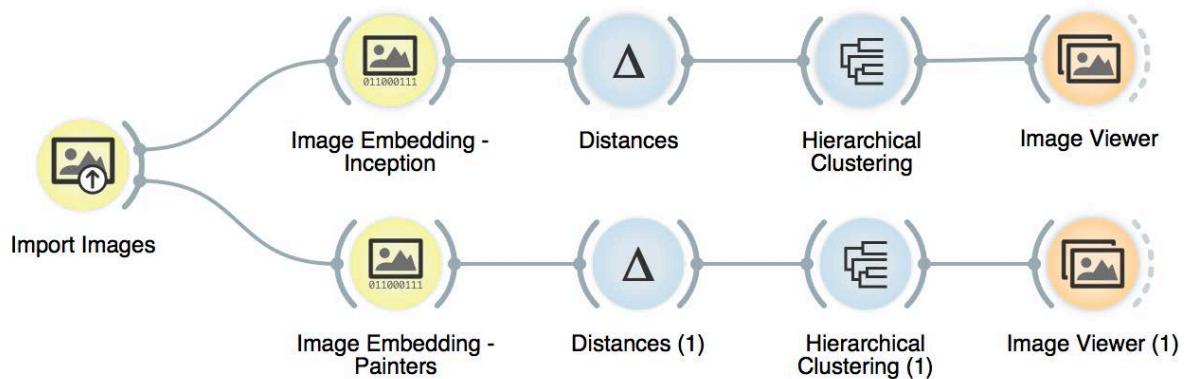


Figure 1: Comparing two embedders through the results of hierarchical clustering.

2.2. PART 2: DEEP-LEARNING-BASED EMBEDDINGS FOR IMAGE ANALYTICS

We will learn how to transform images into feature vectors and uncover similar images. We will discuss different measures of similarity and use hierarchical clustering to find interesting groups (Figure 1). Finally, we will compare the performance of two deep models on paintings from Europeana collection (Purday 2009) and discuss the differences. One model was trained on images from everyday life and the other specifically on paintings. Which one works better for which situation?

2.3. PART 3: IMAGE MAPS AND NEIGHBOURS

In the third part of the workshop, we will learn how to use a reference image to find the most similar image from a collection of paintings. A reference painting will serve as a point of departure for the exploration of the immediate similarity-based neighbourhood of the image (Figure 2). We will also learn about image maps – a projection technique that maps all images onto a two-dimensional plane. We will use these projections to see whether Monet is really so different from Manet (Figure 3).

Finally, participants will be encouraged to create their own repository of paintings from Europeana collection (or other resources) and apply the learnt workflows (clustering, visualization and neighbour exploration) to their own data.

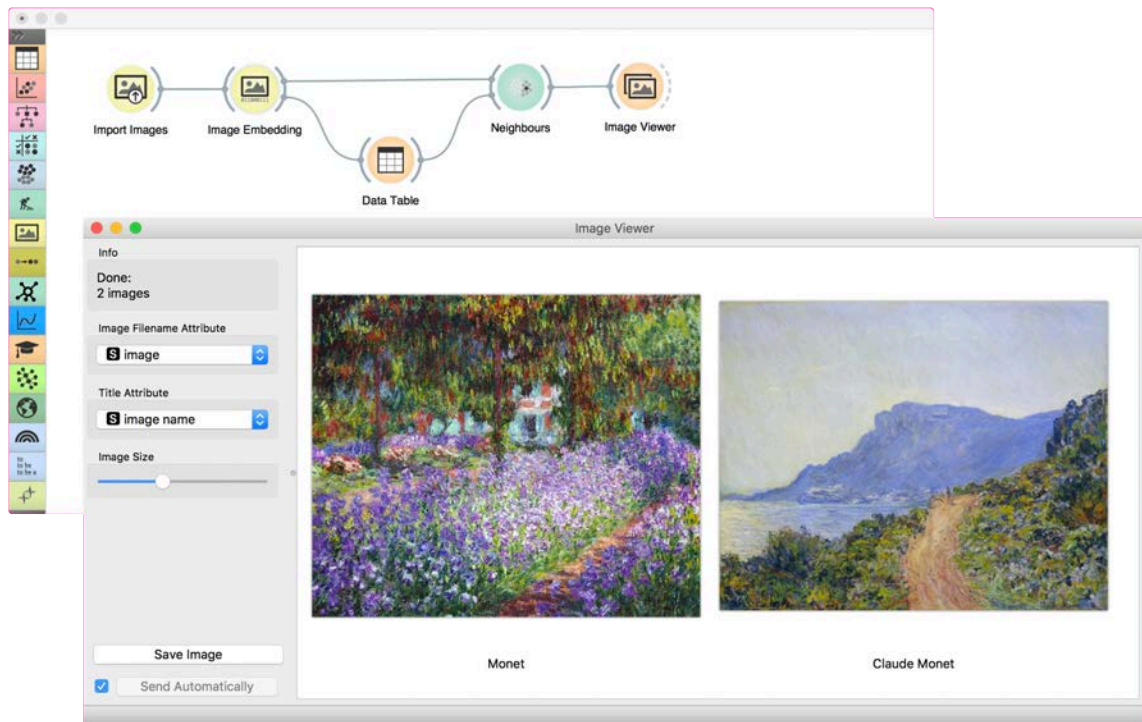


Figure 2: The closes neighbour to Monet's painting is another one of his paintings.

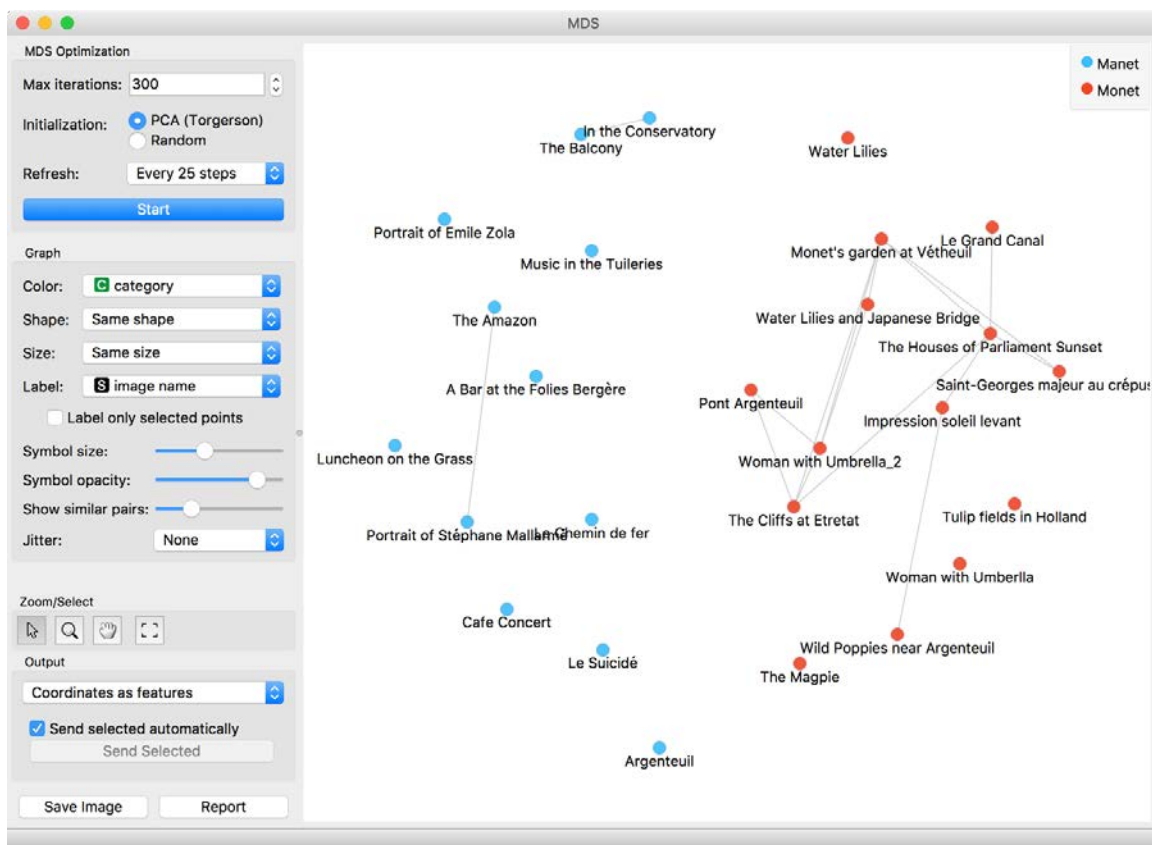


Figure 3: Projection of paintings of Monet and Manet by multi-dimensional scaling. Seems quite obvious these two painters are different.

3. REFERENCES

- [1] Demšar, J. et al., 2013. Orange: data mining toolbox in Python. *The Journal of Machine Learning Research*, 14(1), pp. 2349-2353.
- [2] Newell, J., 2012. Old objects, new media: Historical collections, digitization and affect. *Journal of Material Culture*, 17(3), pp. 287-306.
- [3] Purday, J., 2009. Think culture: Europeana.eu from concept to construction. *The Electronic Library*, 27(6), pp. 919-937.

How to Create a Memory Box with Raspberry PI and RFID Without Being a Programmer

Liviu Pop

The Institute 'Folklore Archive of the Romanian Academy', Cluj-Napoca, Romania - liviu.o.pop@gmail.com



This workshop attempts to teach non-technical participants how to create a memory box that can play sounds using a Raspberry PI, an RFID sensor and RFID tags. The box as a prototype was designed as a tool for returning songs and poems from folklore archives back to the villages where were recorded, a two or three generations later. The concept of open folklore is more deeply related to the structure of the research field than one might think initially. Folklore is open source, as Dorothy Noyes explains in “From Homeric Epic to Open-Source Software: Towards a Network Model of Invention” and we should treat it like that. Inspired by the links between open source and folklore I explore the idea of an alternative method of preserving the digitized content from the folklore archives.

It is of great importance if the records from a folklore archive can be made publicly accessible not only from a cultural heritage point of view, but also from a personal and emotional point of view. We could see different facets of a folklore archive by using different metaphors for it. First, we can view the archive from the point of view of the zoo. During a certain time, some researchers went into the “wild” and gathered certain “folklore species” from their “natural environment” - the traditional villages, for example. Many of the informants left behind very few traces of their existence, and even fewer kept in a public institution. It might even happen that they were the last ones to know a particular song or story. Once that folklore piece was “captured” / recorded on tape or paper, its survival is assured for an indefinite time. It seems only fair to “repopulate” the former natural habitat with long gone folklore species in case of need. Sending the records from the archive back to the places they were taken from in a memory box might be a gesture of courtesy, but the emotional impact on the informants or their offsprings might be surprising.

In order to have a stronger connection with those memory boxes, we can envision different custom enclosures for them made from traditional materials, and in this particular case, from wood. We added small wooden coins with RFID tags that act as interfaces to the small memory boxes, thus lowering and hiding the technical aspects and making them more friendly and more familiar. We aimed to obtain a hidden tech interactive installation that will be easy to use. By using small wooden coins that have RFID tags embed (similar to the access cards used in automatically opening the doors), we hope to achieve a nonintrusive, friendly and familiar object instead of an alien hightech totem that nobody would approach and use.

This workshop is done mostly by hacking preexisting software, so no need for programming is required. I myself I’m not a programmer and I managed to create a functional prototype by tinkering with a very friendly hardware and following tutorials online. For the workshop there will be several Raspberry Pi’s 3, Zero and Zero W available and some RFID sensors, but if you have your own Raspberry Pi, please bring it with you. The workshop is dedicated to those interested in practical solutions related to the *Internet of Things for Cultural Heritage and the Humanities* or *Methodologies and technologies of the digitization process for the production and preservation of digital cultural heritage*.

AIUCD 2018

Panels

Competenze culturali per un Umanesimo Digitale. La via italiana verso Europa 2018

Panel organizzato da AIUCD e DiCultHer non sottoposto a valutazione

Nicola Barbuti¹, Carmine Marinucci², Simonetta Buttò³, Fabio Ciotti⁴, Emiliano Degli Innocenti⁵, Luigi Spadari⁶, Giovanna Barni⁷, Fabio Viola⁸, Serge Noiret⁹

¹ Polo Apulian DiCultHer per la Formazione e Ricerca nel Digital Cultural Heritage, Arts and Humanities

² Scuola a Rete per la Formazione nel Digital Cultural Heritage, Arts and Humanities - DiCultHer

³ Istituto Centrale per il Catalogo Unico ICCU - MiBACT

⁴ Associazione per l'Informatica Umanistica e la Cultura Digitale

⁵ Digital Research Infrastructure for the Arts and Humanities - DARIAH-IT

⁶ Comando Carabinieri Tutela Patrimonio Culturale

⁷ CoopCulture - Società Cooperativa Culture

⁸ Associazione Culturale TuoMuseo - Impresa sociale

⁹ European University Institute

ABSTRACT

Il panel intende proporre un confronto tra i rappresentanti dei principali soggetti pubblici e privati nazionali impegnati nella ricerca, formazione, gestione, tutela, valorizzazione del digital cultural heritage e del patrimonio culturale tangibile e intangibile, sul tema dei saperi da integrare nella formazione del sistema di competenze umanistiche digitali già oggi indispensabile per gestire, tutelare e valorizzare il cultural heritage nell'evo digitale contemporaneo, garantendo nel contempo la conservazione e la sopravvivenza nel futuro del nuovo patrimonio culturale digitale.

KEYWORDS

Competenze digitali, Competenze culturali, Digital Humanities

Common Strategies and Aims of Italian Research Centers for Digital Humanities

Panel organizzato da AIUCD non sottoposto a valutazione

Fabio Ciraci¹, Enrica Salvatori², Cristina Marras³, Franco Nicolucci⁴, Gianfranco Crupi⁵,
Francesca Tomasi⁶, Luigi Catalani⁷, Nicola Barbuti⁸, Fabio Ciotti⁹

¹ Centro interdipartimentale di ricerca in Digital Humanities, Università del Salento, Italy - fabio.ciraci@unisalento.it

² Laboratorio di Cultura Digitale, Università degli Studi di Pisa - enrica.salvatori@unipi.it

³ CNR – ILIESI - cristina.marras@cnr.it

⁴ VASTLAB, Università degli Studi di Firenze - franco.nicolucci@gmail.com

⁵ DigiLab, Università degli Studi di Roma - gianfranco.crupi@uniroma1.it

⁶ Centro di Risorse per la Ricerca Multimediale, Università degli Studi di Bologna - francesca.tomasi@unibo.it

⁷ Sezione Digital Humanities del Centro FiTMU dell'Università degli Studi di Salerno - lcatalani@unisa.it

⁸ DICULTHER – The Digital Cultural Heritage, Arts and Humanities School - nicola.barbuti@uniba.it

⁹ AIUCD - Associazione per l'Informatica Umanistica e la Cultura Digitale - fabio.ciotti@uniroma2.it

ABSTRACT

The purpose of the panel is to present Research Center activities, discuss their role within the university system, establish a common strategy and identify shared goals for the dissemination and disciplinary recognition of Digital Humanities.

KEYWORDS

strategies, community, recognition

Archivi digitali d'autore: ipotesi di lavoro

Pierluigi Feliciati¹, Stefano Allegrezza², Ernesto Belisario³, Salvatore Vassallo⁴, Emmanuela Carbé⁵, Primo Baldini⁶, Paul Gabriele Weston⁷ (coordinatore del panel)

¹ Università di Macerata, Italia - pierluigi.feliciati@unimc.it

² Università di Udine, Italia - stefano.allegrezza@uniud.it

³ e-lex, Italia - ebelisario@e-lex.it

⁴ Archivum Romanum Societas Iesu, Italia - vassallosalvatore@gmail.com

⁵ Università di Pavia, Italia - emmanuela.carbe@gmail.com

⁶ Università di Pavia, Italia - primo.baldini@unipv.it

⁷ Università di Pavia, Italia - paul.weston@unipv.it

ABSTRACT

Il panel mira ad attraversare su più livelli il tema della conservazione del digitale nativo nell'ambito degli archivi di persona, con particolare riguardo alla conservazione e gestione dei materiali di autori contemporanei. L'intento è quello di avviare una riflessione condivisa con la comunità scientifica su metodologie e strumenti da adottare in questo campo di indagine non del tutto esplorato e ancora carico di criticità teoriche e operative. Se è chiaro ormai che la conservazione degli archivi nativi digitali è una sfida sempre più importante nei più diversi settori della nostra società (l'ambito della pubblica amministrazione e quello forense sono tra gli esempi più noti), è altrettanto vero che per gli archivi di persona, nel contesto specifico della produzione letteraria, i casi di studio a nostra disposizione sono ancora sporadici e non sufficienti per individuare del tutto delle buone pratiche. Pare dunque necessario un confronto all'interno della comunità scientifica delle DH, non tanto per condividere risultati quanto per illustrare le problematiche maggiori e individuare delle ipotesi di lavoro. Il panel vuole aprirsi al dibattito con un approccio multidisciplinare, cercando di affrontare le questioni archivistiche più importanti relative all'acquisizione e conservazione dei fondi, alla loro archiviazione, gestione e fruizione. Verrà dato particolare rilievo alla scelta di formati e standard, alla qualità dei dati e dei servizi, ai numerosi aspetti critici sotto il profilo giuridico, alla gestione del versioning e agli strumenti per la ricerca e la didattica. Nel corso della sessione saranno esaminate due esperienze italiane: il lavoro sull'archivio ibrido di Massimo Vannucci e il progetto PAD – Pavia Archivi Digitali, che ha avviato nel 2010 un percorso di sperimentazione su archivi di scrittori contemporanei, perfezionando negli anni alcune strategie per le fasi di preparazione dei conferimenti e acquisizione dei materiali.

PAROLE CHIAVE

Born-digital archiving, Digital Curation, Digital Preservation, Private Papers Archiving, Born-digital papers, Personal Digital Archiving, RiC Model

1. INTRODUZIONE

La sfida rappresentata dalla creazione di depositi per la documentazione nativamente digitale non è data tanto dalla individuazione e messa in atto delle procedure di riversamento dei file, quanto dall'impegno richiesto per consentirne nel tempo la fruizione. Non c'è dubbio che il numero dei file che sono oggetto di archiviazione pone già notevoli difficoltà in quanto solitamente eccede di gran lunga quello di un archivio di natura cartacea, con l'ulteriore difficoltà di gestire una grande varietà di formati e di tenere in vita supporti connessi a generazioni tecnologiche ormai in completo disuso. Sfida particolarmente impegnativa è quella della descrizione degli archivi, in primo luogo per il numero davvero notevole dei file e per la diversità dei formati. È, tuttavia, la fluidità del mezzo digitale che rappresenta per l'archivista-conservatore la vera novità rispetto all'archivio tradizionale. Questa fluidità determina un maggior numero di relazioni e di intrecci tra i contenuti, ma soprattutto la stratificazione dei file. Ogni versamento di documenti, che in un archivio tradizionale è incrementale e non produce sostituzioni, in un archivio digitale può produrre il fenomeno opposto: la sostituzione di contenuti, lo spostamento di un file da una cartella all'altra, i cambiamenti di metadati tecnici derivanti dalla sostituzione di un computer – o di una componente software – non sono che alcune delle manifestazioni più comuni. Inoltre, la gestione dell'archivio digitale presuppone l'aggiornamento costante dei modelli di dati, degli standard e delle procedure per far fronte alla crescente varietà delle fonti documentarie.

In aggiunta ai noti rischi ai quali sono esposti gli oggetti digitali, nonché alla stratificazione delle versioni prima accennata, intervengono come altri fattori di particolare criticità la tutela dei dati sensibili, la gestione dei diritti, le licenze d'uso degli applicativi software (nell'evenienza di trasferimenti di pacchetti software dal soggetto utilizzatore a quello che si occupa della conservazione), l'organizzazione e le responsabilità della curatela digitale e, infine, il rapporto con l'utenza.

Proprio quest'ultimo fattore rappresenta uno degli elementi di maggiore incertezza rispetto alle soluzioni da adottare per la restituzione dei dati. Idealmente si fronteggiano due approcci: quello archivistico, in cui la descrizione mira a riprodurre il contesto di provenienza, ossia l'organizzazione che l'autore ha dato alle proprie memorie elettroniche, e quello bibliocentrico, in cui sono le singole opere – anche in senso astratto – a fungere da elemento di aggregazione dei contenuti associati: il materiale pretestuale, le diverse stesure del testo, il materiale scambiato con l'editore, le diverse edizioni parziali o integrali, le presentazioni delle opere, le recensioni, la voce dei lettori, i rapporti con le altre opere e così via. Evidentemente questo secondo approccio, che per certe fasce di utenza può essere più funzionale e intuitivo, ci dice assai

poco sulle abitudini di scrittura dell'autore, sul modo in cui lui/lei hanno organizzato i propri spazi mentali, digitali, relazionali.

Il successo di un progetto finalizzato alla conservazione a lungo termine di archivi privati digitali nativi dipende dalla capacità di individuare fin dall'inizio le soluzioni più adeguate (riguardo a linee guida, standard, buone pratiche, ecc.), con l'intesa che abbiano anche la caratteristica della sostenibilità.

Il caso degli archivi nativi digitali letterari presenta, rispetto ad altri casi come quello offerto da PARADIGM (Personal Archives Accessible in Digital Media), una serie di specificità che richiedono soluzioni *ad hoc*. Importanti progetti sono già stati avviati da tempo, come quello dell'Harry Ransom Center e dell'Emory University's Manuscript, Archives and Rare Book Library. Nel corso di questa sessione, che offrirà un attraversamento su più livelli delle tematiche qui menzionate, saranno illustrati anche due casi italiani: il prezioso lavoro sull'archivio ibrido di Massimo Vannucci, che rappresenta nel panorama nazionale forse il primo tentativo di inventariazione e riordinamento di un archivio di questo tipo; e l'esperienza del progetto PAD – Pavia Archivi Digitali, che dal 2010 opera a fianco del Centro Manoscritti di Pavia per acquisire e conservare archivi digitali nativi di autori contemporanei, con lo scopo futuro di metterli a disposizione dell'utenza per attività di studio e ricerca.

Considerando l'insieme delle problematiche che la documentazione, gestione e conservazione di questi archivi sottendono, si rende dunque evidente la necessità di integrare fra loro competenze appartenenti ad ambiti molto diversi: discipline letterarie e filologiche, ma anche giuridiche, tecniche archivistiche, tecnologia dell'informazione, aspetti amministrativi. Si pensi ad esempio alla conservazione a lungo termine e ai risvolti metodologici, tecnologici e organizzativi connessi, conosciuti in gran parte, ma su cui ancora non c'è la garanzia di stabilità e affidabilità che sarebbe auspicabile. O ancora, alla descrizione archivistica: pur basandosi su standard ben conosciuti e in genere periodicamente aggiornati, le caratteristiche di questi archivi hanno un impatto particolare che necessita di specifiche evoluzioni, per la numerosità dei file conservati, il *versioning*, la presenza degli stessi file in più cartelle, l'intreccio di relazioni tra file, eventi, responsabilità e agenti, l'assenza di politiche condivise, di massimari di scarto e di piani di conservazione. Una particolare attenzione al modello FRBRoo può essere opportuna, oltre all'osservazione della diffusione di standard come RDA, modalità di pubblicazione come LOD e formati di scambio come JSON. Altre competenze necessarie riguardano lo studio di ambito filologico, per la corretta valorizzazione delle "carte", attraverso una lettura e interpretazione in cui si tenga conto anche delle caratteristiche tecniche dei file. Poiché è probabilmente irrealistico pensare che in ogni contesto siano presenti da subito queste e altre competenze, è forse opportuno pensare da un lato a progetti di messa a fattor comune della ricerca e dello sviluppo, dall'altro a programmi di formazione mirata.

2. P. FELICIATI – LA QUALITÀ DEGLI ARCHIVI LETTERARI DIGITALI NATIVI: DALLA SEDIMENTAZIONE AL SERVIZIO

La complessità richiesta per la corretta progettazione, creazione, gestione e conservazione a lungo termine degli archivi digitali, nella dialettica continua tra sviluppo delle soluzioni tecnologiche, standard tecnici e procedurali e normativa nazionale e internazionale, è nota e molto studiata da molti. Uno degli snodi critici riguarda la garanzia di contestualizzazione del documento digitale, inseguendo da un lato il modello di vincolo proprio della tradizione archivistica e guardando dall'altro alla sfida che pone la multidimensionalità della sedimentazione, della strutturazione degli elementi del documento e le conseguenze che tutto ciò ha sull'accesso finale. Se si considerano gli archivi letterari nativi digitali o ibridi, le questioni si fanno più complesse e intriganti: quale vincoli tra entità informative (che siano *files* o no) sono da mantenere o creare ex-novo per non svilire il senso complessivo della sedimentazione documentaria dell'attività di uno scrittore? Come gestire il corpus documentario, fatto di entità informative e delle loro relazioni qualificate? Quali scelte selettive ci possiamo permettere di compiere? E come rappresentarlo al fine dell'accesso da parte di utenti finali, ammesso che le caratteristiche della *collezione* siano state studiate e costituite stabilmente? Questo contributo si propone di discutere di queste e altre questioni, inserendole in un modello di qualità che garantisca la migliore gestione e il miglior uso possibile degli archivi letterari digitali nativi.

3. S. ALLEGREZZA – PERSONAL DIGITAL ARCHIVING: CRITICITÀ E SOLUZIONI NELLA CONSERVAZIONE A LUNGO TERMINE DEGLI ARCHIVI LETTERARI DIGITALI

L'intervento si propone di affrontare il tema del *Personal Digital Archiving* con particolare riferimento agli archivi di letterati, fornendo da una parte un quadro aggiornato sugli sviluppi più recenti del dibattito in corso e dall'altro descrivendo l'intervento d'inventariazione e riordinamento dell'archivio "ibrido" – contenente cioè documenti sia cartacei che digitali – di Massimo Vannucci (1957-2012), politico e imprenditore. Si tratta probabilmente del primo caso in Italia di intervento su un archivio digitale di persona: la documentazione risale prevalentemente al periodo compreso fra il 2005 e il 2012, quindi all'epoca dell'attività politica di Vannucci, ed è proprio la natura ibrida dei materiali a determinare l'originalità dell'intervento: e-mail archiviate sia nel client di posta che sulla webmail; documenti rinvenuti nel computer e nello smartphone; post sui social media, si aggiungono alla documentazione cartacea, costituendo a tutti gli effetti un archivio "ibrido" di persona. Oltre a illustrare il lavoro svolto, l'intervento si pone l'obiettivo di aprire una riflessione sulle criticità connesse con la formazione degli archivi di persona in ambiente digitale: tematiche di importanza fondamentale se si pensa che gli archivi – non soltanto di persona – diventeranno totalmente digitali nel giro di poco tempo, dato che i trend evolutivi degli ultimi anni hanno registrato una fortissima accelerazione in tale direzione. È urgente, quindi, aprirsi al confronto su

queste tematiche e individuare soluzioni che permettano di conservare per il futuro gli archivi digitali che si stanno oggi formando.

4. S. VASSALLO – GLI STANDARD ARCHIVISTICI ALLA SFIDA DEGLI ARCHIVI LETTERARI

Di fronte alla descrizione di Archivi Letterari e di altri archivi che non ricadono nella stabile casistica degli archivi di enti, gli standard archivistici (come ISAD, International Standard for Archival Description) sembrano fallire nel tentativo di imbrigliare documentazione per sua natura fluida, con versioni differenti dello stesso documento che si sovrappongono come onde, in una sorta di griglia autocompilativa.

Ciò in gran parte è dovuto a un'erronea interpretazione degli standard, degenerati da filosofie descrittive a precetti, tabelle descrittive fino a diventare anche rigidi e unici modelli di visualizzazione e accesso per l'utente.

RiC (Records in Contexts) da questo punto di vista guarda all'utente, lo asseconda rappresentando il mondo non solo e non tanto come una rigida impalcatura gerarchica in cui incasellarlo quanto come una rete di significati e livelli differenti di lettura intrecciati fra di loro.

La sfida raccolta dai nuovi standard archivistici è proprio quella di fornire le chiavi per poter dar conto e poter descrivere un mondo complesso che si snoda a cavallo fra beni culturali differenti e metodologie per descriverli diverse.

Progetti come SNAC (Social Networks and Archival Context), a titolo di esempio, mostrano infatti come i monoliti descrittivi non possano funzionare in un mondo, quello delle digital humanities, fatto di condivisione e comunicazione.

5. E. BELISARIO – CREA, GESTISCI, CONSERVA: LE NORME SUL CICLO DI VITA DEL DOCUMENTO INFORMATICO

Nel corso degli ultimi vent'anni, il legislatore si è occupato ripetutamente delle norme sulla formazione, gestione, archiviazione e conservazione dei documenti informatici.

Obiettivo primario è stato consentire di avviare il difficile percorso di digitalizzazione della pubblica amministrazione e di dematerializzazione dei suoi archivi. Tuttavia, le stesse norme si applicano anche ai privati che vogliono formare documenti e archivi aventi piena validità giuridica.

Con la recente riforma del Codice dell'Amministrazione Digitale (Decreto Legislativo n. 82/2005), operata a seguito della riforma della Pubblica Amministrazione, il quadro regolatorio è ormai completo e costringe tutti i soggetti giuridici ad un complesso adeguamento: tecnologico, organizzativo, giuridico e contrattuale.

Le decisioni da assumere in questa fase sono fondamentali se si vuole conservare non solo l'integrità e la disponibilità, ma anche il valore legale degli archivi.

Per questo motivo, è opportuno definire le caratteristiche delle soluzioni e delle procedure che consentano di gestire e conservare in modo adeguato il patrimonio documentale.

6. E. CARBÉ, P. BALDINI – I PROCESSI DI ACQUISIZIONE DI ARCHIVI LETTERARI DIGITALI NATIVI: IL CASO PAD

I fondi conservati in PAD – Pavia Archivi Digitali, difforni tra loro per mole, tipologia e supporti di provenienza, rappresentano una casistica davvero ampia che dimostra le molteplici criticità nella gestione di un archivio digitale nativo di tipo letterario, a partire dalla fase di ingest. L'intervento mira a illustrare le metodologie e gli strumenti raffinati in questi anni nell'ambito dei processi di acquisizione dei fondi, su cui ci si aspettano ulteriori miglioramenti grazie a un proficuo confronto con la comunità scientifica. La preparazione del conferimento prevede una fase conoscitiva preliminare con l'autore, sia per la definizione delle modalità di acquisizione, regolate per gli aspetti legali da un contratto, sia per una prima raccolta di informazioni utili alla preservazione del contesto di origine del materiale, effettuata anche grazie a un questionario, che viene inviato all'autore insieme con una Guida al conferimento. L'acquisizione degli archivi avviene tramite un hard disk dotato di crittografia hardware, qualora il trasferimento dei file avvenisse da un computer; in altri casi (ad esempio il salvataggio di siti internet, o ancora il recupero di file contenuti in vecchi floppy disk) sono messe in atto soluzioni alternative, sempre documentate per preservare la storia archivistica del fondo. All'arrivo del materiale a Pavia, sono attivate procedure per la messa in sicurezza e per il controllo dell'integrità. Il materiale viene poi verificato insieme con l'autore e infine convalidato. Solo a questo punto il conferimento risulta concluso: si passa dunque al trattamento archivistico dei materiali, ora in fase di sperimentazione, e alle più consolidate procedure di conservazione a lungo termine.

7. BIBLIOGRAFIA

- [1] Alfieri A., Feliciati P. 2017. *Gli archivi online per gli utenti: premesse per un modello di gestione della qualità*. In: J LIS.it, 8, 1: 22-28. DOI: <http://dx.doi.org/10.4403/jlis.it-12269>.
- [2] Allegrezza S., Gorgolini L. 2016. *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*. Bologna, Il Mulino. ISBN: 978-88-152648-00.
- [3] Archivio Massimo Vannucci. URL: <http://www.archiviomassimovannucci.it>.
- [4] Barrera-Gomez J., Erway R. 2013. *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house*. Dublin, OH, OCLC Research. URL: <http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>.

- [5] Carbé E. 2017. *Walking in My Shoes: A Case Study from a Born-digital Archive*. In: CEUR, 2021: 137-146. URL: <http://ceur-ws.org/Vol-2021/>
- [6] Carroll L., Farr E., Hornsby P., Ranker B. 2011. *A Comprehensive Approach to Born-Digital Archives*. In: *Archivaria*, 72: 61-92. DOI: <https://open.library.emory.edu/publications/emory:cksgv/pdf/>.
- [7] Consultative Committee for Space Data Systems 2012. *Reference Model for an Open Archival Information System (OAIS)*. Washington DC, CCSDS Secretariat. URL: <https://public.ccsds.org/pubs/650x0m2.pdf>.
- [8] Delneri, F. 2017. *Il documento amministrativo informatico: un cammino per approssimazione. Criticità e risposte possibili, tra normativa e prassi, dalla formazione alla conservazione*. In: *JLIS.it*, 8, 3: 26-38. URL: <https://www.jlis.it/article/view/12193/11275>.
- [9] Dooley J. 2015. *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials*, Dublin, Ohio, OCLC Research. URL: <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015.pdf>.
- [10] Erway, R. 2010. *Defining "Born Digital". An Essay*. Dublin, Ohio: OCLC Research. URL: <http://www.oclc.org/research/activities/hiddencollections/borndigital.pdf>.
- [11] Erway, R. 2012. *You've got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*. Dublin, Ohio, OCLC Research. URL: <http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>.
- [12] Erway R., Goldman B., McKinley M. 2014. *Agreement Elements for Outsourcing Transfer of Born Digital Content*. Dublin, Ohio, OCLC Research. URL: <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-born-digital-content-transfer-2014.pdf>.
- [13] Experts Group on Archival Description. *Records in Contexts – Conceptual Model*. Versione per commenti pubblici. URL: <https://www.ica.org/en/egad-ric-conceptual-model>.
- [14] Feliciati P. 2016. *L'usabilità degli ambienti bibliotecari e archivistici digitali come requisite di qualità: contest, modelli e strumenti di valutazione*. In: *JLIS.it*, 7, 1: 113-130. DOI: <http://dx.doi.org/10.4403/jlis.it-11512>.
- [15] Gueguen, G., da Fonseca V.M.M., Pitti D.V., Sibille-de Grimouard, C. 2013. *Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description*. In: *The American Archivist*, 76, 2: 566-583. URL: https://www.ica.org/sites/default/files/EGAD_English.pdf.
- [16] Guercio M. 2013. *Conservare il digitale. Principi, metodi e procedure per la conservazione a lungo termine di documenti digitali*. Roma-Bari, Laterza. ISBN: 978-88-593000-9-0.
- [17] Guercio M. 2016. *Archivi personali: la sfida del digitale. Una riflessione*. In: *Gli archivi di persona nell'era digitale. Il caso dell'archivio di Massimo Vannucci*, a cura di S. Allegrezza e L. Gorgolini. Bologna, Il Mulino: 133-148.
- [18] Kirschenbaum M.G., Farr E.L., Kraus K.M., Nelson N., Peters C.S., Redwine G., Reside D. 2009. *Digital Materiality: preserving access to computer as complete environments*. In: *iPres 2009. The Sixth International Conference on Preservation of Digital Objects*. California Digital Library, UC Office of the President: 105-112. URL: <https://escholarship.org/uc/item/7d3465vg>.
- [19] Kirschenbaum M.G., Ovenden R., Redwine G. 2010. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington DC, Council on Library and Information Resources. ISBN: 978-1-932326-37-6. URL: <https://www.clir.org/pubs/reports/pub149/pub149.pdf>.
- [20] Kuny T. 1998. *A Digital Dark Ages? Challenges in the Preservation of Electronic Information*. In: *International Preservation News*, 17. URL: <https://archive.ifla.org/IV/ifla63/63kuny1.pdf>.
- [21] International Council of Archive. *General International Standard for Archival Description*. URL: <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>.
- [22] Lavoie, B. 2014. *The Open Archival Information System (OAIS) Research Model: Introductory Guide (2nd Edition)*. Dublin, Ohio, OCLC Research, Digital Preservation Coalition. DOI=<http://dx.doi.org/10.7207/TWR14-02>.
- [23] Lenzini L. 2016. *Don't Save! Fortini, un Mac e le ultime cose*. In: *L'Ulisse. Rivista di poesia, arti e scritture*, 19, 96-101. URL: <http://www.lietocolle.com/cms/wp-content/uploads/2016/04/ULISSE-19.pdf>.
- [24] Lynch T. J. 2014. *Social Networks and Archival Context Project: A Case Study of Emerging Cyberinfrastructure*. In: *Digital Humanities Quarterly*, 8, 3. URL: <http://www.digitalhumanities.org/dhq/vol/8/3/000184/000184.html>.
- [25] PAD – Pavia Archivi Digitali: pad.unipv.it.
- [26] Stollar Peters, C. 2006. *When Not All Papers are Paper: A Case Study in Digital Archiving*. In: *Journal of the Society of Georgia Archivists*, 24: 22-34. URL: <http://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1068&context=provenance>.
- [27] Vassallo S. 2013. *L'integrazione tra archivi e biblioteche alla prova del web semantico*. In: *Biblioteche in cerca di alleati. Oltre la cooperazione, verso nuove strategie di condivisione*. Milano: Editrice Bibliografica. ISBN: 978-88-707576-5-1.
- [28] Weston P.G. 2013. *Digital curation: nuove opportunità per bibliotecari e archivisti*. In: *Il libro, gli archivi e la memoria digitale*, a cura di Paul Gabriele Weston, Milano, Unicopli: 9-66. ISBN: 978-88-400168-1-8.
- [29] Weston P.G., Carbé E., Baldini P. 2016. *Hold it All Together: a Case Study in Quality Control for Born-Digital Archiving*. In: *Qualitative and Quantitative Methods in Libraries (QQML)*, 5: 695-710. URL: http://www.qqml.net/papers/September_2016_Issue/5313QQML_Journal_2016_Westonet_695-710.pdf.
- [30] Weston P.G., Carbé E., Baldini P. 2017. *Se i bit non bastano: pratiche di conservazione del contesto di origine per gli archivi letterari nativi digitali*. In: *Bibliothecae*, 6, 1: 154-177. DOI: <https://doi.org/10.6092/issn.2283-9364/7027>.
- [31] *Workbook on Digital Private Papers 2008*. URL: <http://www.paradigm.ac.uk/workbook/index.html>.

Digital practices and teaching Humanities

Antonio Brusa¹, Annalisa Caputo², Giuliano De Felice³,
Isidoro Davide Mortellaro⁴, Claudia Villani⁵ (coordinatrice del panel)

¹ Università di Bari - brusantonio@gmail.com

² Università degli studi di Bari, Italia - annalisa.caputo@uniba.it

³ Università degli studi di Foggia, Italia - giuliano.defelice@unifg.it

⁴ Università degli studi di Bari, Italia - isidorodavide.mortellaro@uniba.it

⁵ Università degli studi di Bari, Italia - claudia.villani@uniba.it

ABSTRACT

DH can become a chance for rethinking and reinventing humanities. Indeed, several scholars have emphasized the importance of DH-driven methodological and epistemological reflections in research and in the dissemination and teaching of humanities. Moreover, DH encourage scholars to go beyond the reductionist specialism of separate academic disciplines. DH-based cultural projects are indeed characterized by interdisciplinarity and Openness (Digital Manifesto 2.0) and are elaborated and realized within a digital craft (digicraft) laboratory, which is an outcome of the combination of diverse skills.

The aim of this panel is to reflect critically on the teaching of humanities in the digital era and on the relationship between books and digital technology on the basis of the concrete experience of teachers, professors and researchers from different academic fields. What these diverse experiences have in common is, in particular, the awareness that in the teaching of humanities DH are not only a tool to reach new digital skills: most importantly, they offer a chance to reflect on methodological approaches and elaborate new ways of thinking about humanities themselves, their epistemological statute, their dissemination and didactic mediation. In short, DH represent on the one hand a form of theoretical and methodological training that allows us to rediscover (but also to re-think) the tools and foundations of the very disciplines, on the other a way to answer the growth of one-sided and functional public uses of the past by the most varied public and private actors.

KEY WORDS

Digital, Teaching, History, Philosophy, Archaeology, Publishing.

1. PANEL ABSTRACT

International surveys and reports show that the use of digital resources, methods, tools and applications is spreading among the scholars of humanities and highlight the positive or problematic aspects entailed by such spreading (Holm, Jarrik, Scott 2015; DIMPO 2017). Nobody denies that digital resources and Web 2.0 change radically the environment, the methodologies and the concrete practices of scientific research in several disciplinary fields. Some, for instance, see DH as a building made up of several rooms that can be entered from different doors, looking into digital technologies through the lenses of the problems, the themes and the needs of the separate disciplines (Robertson 2016).

What we are most interested in here is the reflection on the relationship between the didactic mediation of humanities and digital technologies, for we are convinced that a “reasonable and well-reasoned” use of digital resources can contribute to a rethinking of the forms of the very didactic mediation, helping both in moving beyond the rigidity of traditional didactics (in its prevailing transmissive form) and in experimenting innovative, constructivist and participatory practices, in accordance with the most recent European requests. Furthermore, as it becomes clearer in the cases of history and archaeology, the digital medium provides the opportunity to learn/experiment specific methodologies and various approaches to academic disciplines, strengthening fundamental disciplinary skills (for example in the critique of sources) and widening them so that they can include those digital skills that are crucial for a scholar (and citizen) of the digital and global era. Moreover, in other cases, such as that of philosophy, digital technology gives the opportunity to look differently at the academic discipline and at its spreading and didactic mediation, especially by making it possible to connect different languages (the philosophical, the artistic, the musical, the literary ones, etc.) and to rethink, promote and develop in an original way basic philosophical skills, thus fruitfully contaminating other disciplinary fields.

In addition, in all cases digital technologies condition also the traditional tools of dissemination and didactic mediation: the book, the journal, the textbook.

This panel aims thus at reflecting on the teaching of humanities in the digital era, on the relationship between the book (especially the school textbook and the university textbook) and the digital medium, on the basis of the concrete experience of teachers, professors and researchers from different academic fields. What all these various experiences have in common is the awareness that in the teaching of humanities DH are not only a tool to acquire new digital skills; most importantly, they offer a chance to reflect on methodological approaches and elaborate new ways of thinking about humanities themselves, their epistemological statute, their dissemination and didactic mediation.

Indeed, the importance of DH-driven methodological and epistemological reflection in the research activities of these disciplines has been highlighted (Thaller 2012). DH, in sum, are a form of theoretical and methodological training that allows us to rediscover (but also to rethink) the tools and foundations of the very disciplines (Mahony-Pierazzo 2012).

On the other hand, as it has been observed, any cultural project within DH is a complex cultural project, characterized by some basic elements: interdisciplinarity, Openness (Digital Manifesto 2.0) and the elaboration and realization within a digital craft (digicraft) laboratory that is an outcome of the combination of diverse skills.

DH force scholars to go beyond the reductionist specialism of separate academic disciplines and test both research methodologies, dissemination practices and didactic methodologies (Salvatori 2017). Indeed, openness and interdisciplinarity refer to the other aspect characterizing the relationship between humanities and the digital medium: what historians call “public use”. The digital medium has multiplied in a staggering way the practice of putting online one’s materials in order to reach colleagues, scholars, but also amateurs, and disseminate the results among a wider audience and among teachers and students. That is another reason why a “reasonable and well-reasoned” use of the digital medium, starting from the knowledge, the approaches and methodologies peculiar to humanities, appear as not only crucial, but necessary. Otherwise, one runs the risk of leaving the public use of humanities (whether an historical source, a work of art, a literary or philosophical text, etc.) to others, to the various ways people today look at the past in order to trace back some elements that can be useful for their own identity strategy. Widening Pierre Nora’s reflections, we can argue here that today not only history, but all humanities are involved in the competition, confrontation and conflict with public uses and policies of memory carried out by various actors (Nora 2011).

DH can thus be conceived as a field for research and for the rethinking and reinvention of humanities, of didactic methodologies and of dissemination practices, but also as a qualified answer to challenges brought about by the one-sided and functional public uses of the past by the most varied public and private actors (one might even argue for the existence of Public Digital Humanities).

2. ANTONIO BRUSA, HISTORY AND DIGITAL AND MEDIA EDUCATION.

A grammar of images. An instrument to overcome the iconic functional illiteracy.

The advent of technology has powerfully strengthened a fundamental function of teaching: the communicative function. Materials, images, texts, documents of all kinds: the network is in fact an inexhaustible deposit, which can be put at the service of the communicative function of the teacher. At the same time, it should be noted that the network does not lend itself, with the same facility, to another fundamental function of teaching: that of the laboratory. Apparently the network is presented as a laboratory accessible to students, and not only to professors or researchers. In reality, as it is well known, it is a deceptive laboratory. The case of images is exemplary. Students tend to capture images and use them to illustrate their texts. How to teach them to use them as “documents”? The “grammar of images”, modeled on the “grammar of documents”, experimented in schools for a long time, can be an answer to this question. In this report the fundamental features and problems of teaching are presented.

3. ANNALISA CAPUTO, PHILOSOPHY AND DIGITAL PRACTICES

The triadic relationship between digital practices, philosophical theory and teaching philosophy can be approached from different perspectives.

- Philosophical reflection on digital world: how this new way of being and thinking has changed anthropology and social relations, and therefore the human sciences themselves
- Philosophy of Information and Computer Ethics; issues related to 'big data' (dictatorship of transparency or amplification of possibilities?),
- Philosophy and 'media': how the digital fruition of images, works of art, music, film, video, etc. can change the dialogue between philosophical language and other languages;
- Philosophy and digital communication: the new possibility of philosophical diffusion on the web (sites, blogs, open access online magazines, etc.)
- Philosophy and teaching through digital practices: how to change disciplinary teaching (at University and High School), using digital instruments; and, on the contrary, how to exploit websites and web channels to spread educational experiments and didactic research.

Starting from this theoretical scenario, this paper aims to present a project called “Playing with Thinking” / “Logoi”, a university project that put 'into play':

- different languages (philosophy, art, music, cinema, literature), linking them into an international peer reviewed journal, open access (called “logoi.ph”)
- university and social context: in particular the school world, through a particular didactic methodology (called *Philosophia ludens*, PhL) that is used to teach philosophy from Primary School (PhL for children) to High School. All these educational experiments converge on the on-line Journal “Logoi”, becoming shared practices and networking suggestions for new practices.

4. GIULIANO DE FELICE, DIGITAL ARCHAEOLOGY

The recent discussion on the ratification of the Faro Convention (Convention on the Value of Cultural Heritage for Society (Faro Convention, 2005), Montella 2016; Volpe 2017) has finally triggered also in Italy a deep review of the concept of Cultural Heritage, no longer understood as a set of objects and goods (both material and immaterial) but as a complex network of relations between community and heritage (Manacorda 2017). This important innovation lays the foundations

for an innovative conception of didactics, with the task of transmitting to the future generations the value of cultural heritage, *enhancing the value through its identification, study, interpretation, protection, conservation and presentation* (Faro Convention, Article 5b). In this scenario, doing archaeology becomes a very complex challenge, which requires a deep innovation of the themes and tools normally used. Consequently, teaching archaeology becomes a maybe more challenge too, directly connected with the need to create a new awareness of cultural heritage in the future generations. An important role in overcoming this challenge can be played by digital technologies (already very present in the archaeological field especially on the side of research and development) and by their applications in university courses, in a more organic and effective way. The most interesting perspective is that of designing innovative forms of teaching that can exploit the ability of digital technologies to contaminate the boundary between production and fruition to overcome a purely transmissive structure in favour of a participatory logic (G. Volpe, G. De Felice 2014).

5. ISIDORO DAVIDE MORTELLARO, PUBLISHING AND DIGITAL WORLD

Everything was so clear: we had to get dissolved into information. But just as one with ourselves: eyes and ears, accustomed spectators for Popper's and Sartori's TV. And jet before, Orwell, in 1984 had warned us about totalitarian control. But just in 1984, history had seen the newcomer Mac forcing his brother of those times, IBM, to change skin, to become friendly, personal, PC. Nobody, however, would have bet in those days, on thumbs and its virtues. Thanks to its power, we jumped from the tactile seduction of our remote control to the frantic habit with tablets and smartphones, as they are the true revolutionary characters of our return to the writing habit. Also if syncopated, reactive, stupid, in the tons of bites we waste every day. We are all authors, millions of us. And sometimes we are also ourselves editors. In this arrogant return to writing we feel the exact proportion of blogs and social networks phenomena. In this messy horizon, we find the matter of books moving. Never too healthy, in Italy. To be honest, quite weak. We're almost the last for reading percentages in Europe. Yet, elitist: 7% of strong readers (5% buyers) who move the 44% of the whole Italian books bundle. A market in which digital is stronger everyday: in 2016 74,020 books: a concrete jump forward compared to the 9 thousand in 2010, with the 5% of concrete sells. Even more radical is the mutation in the sell and buy method, if we consider that the global amount of the on-line sales and the weight of Amazon is nowadays enormous. Made-over habits, also in the commercial field. Well: we never made ourselves Italians, as Massimo D'Azeleglio suggested. Mother RAI made it, just a bit. Where will actual communication bring us?

6. BIBLIOGRAPHY

- [1] Bradley P., <http://chronicle.com/blogs/profhacker/where-are-the-philosopherthoughts-from-thatcamp-pedagogy/37408> [8-10-2017]
- [2] Brusa A. (2014), *Io non condivido. Immagini di guerra, social e didattica della storia*, <http://www.historialudens.it/didattica-della-storia/153-io-non-condivido.html> Caputo A. (2011), *Philosophia ludens. 240 attività per giocare in classe con la storia della filosofia*, La Meridiana, Molfetta (BA) (con CD allegato)
- [3] Cataluccio F. (2010), *Che fine faremo i libri?*, Nottetempo
- [4] Ciotti F. (2014), Digital Humanities in Italy and their role in DARIAH research infrastructure, *Leggere, scrivere e far di conto. Informatica*
- [5] Convenzione di Faro (testo ufficiale italiano): http://www.ufficiostudi.beniculturali.it/mibac/export/UfficioStudi/sito-UfficioStudi/Contenuti/Pubblicazioni/Volumi/Volumi-pubblicati/visualizza_asset.html_917365394.html [8-10-2017]
- [6] *European survey on scholarly practices and digital needs in the arts and humanities*, Digital Methods and Practices Observatory Working Group (DiMPO) Survey highlights EN. [Research Report] DARIAH; DIMPO. 2017. <https://hal.archives-ouvertes.fr/DARIAH/hal-01449002v1> [8-10-2017]
- [7] De Natale F. – Caputo A. – Mercante A. – Baldassarra R. (2011), *Un pensiero in gioco*, Stilo, Bari
- [8] Floridi L. (2012), *La rivoluzione dell'informazione* (Codice, Torino)
- [9] Id. (2014), *The Fourth Revolution - How the infosphere is reshaping human reality* (Oxford University Press, 2014)
- [10] Guaraldi M. (2016), *Radici di carta frutti digitali*, Guaraldi
- [11] Haydin T. (2011), *History teaching and ICT*, in I. Davies (dir), *Debates in History Teaching*, Routledge, London
- [12] Holm P., A. Jarrickt, D. Scott, 2015. *Humanities World Report 2015*. New York: Palgrave MacMillan, pp.64-83
- [13] Mahony S., Pierazzo E. (2012), *Teaching Skills or Teaching Methodology?*, in Hirsch ed. *Digital Humanities Pedagogy: Practices, Principles and Politics*, Open Book Publishers <<http://www.openbookpublishers.com/product/161>>. [8-10-2017]
- [14] Manacorda D. (2016), *Patrimonio culturale: un patrimonio collettivo*, in R. Auriemma (ed), *La democrazia della conoscenza: patrimoni culturali, sistemi informativi e open data: accesso libero ai beni comuni?* Atti del convegno (Trieste 28-29 gennaio 2016), Udine, pp. 117-124.
- [15] Marazzini C. (2014) (ed.), *L'editoria italiana nell'era digitale*, Accademia della Crusca
- [16] Montella M. (2016), *La "Convenzione di Faro" e la tradizione culturale italiana*, in P. Feliciati (ed), *La valorizzazione dell'eredità culturale in Italia*. Atti del convegno Supplemento 5/2016 a Il Capitale Culturale, Studies on the Value of Cultural Heritage (<http://riviste.unimc.it/index.php/cap-cult/issue/view/81>), pp. 13-17. [8-10-2017]
- [17] Noiret S. (2015), *Digital Public History: bringing the public back*, Public History Weekly, 3:13, <http://public-history-weekly.oldenbourg-verlag.de/3-2015-13/digital-public-history-bringingthe-public-back-in/> [8-10-2017]
- [18] Nora P. (2011), *Recent history and the new dangers of politicization*, <http://www.eurozine.com/recent-history-and-the-new-dangers-of-politicization/> [8-10-2017]
- [19] Robertson S. (2014), *The Differences between Digital History and Digital Humanities*, CHNM Blog Post (23/1/2014), <http://drstephenrobertson.com/blog-post/the-differences-between-digitalhistory-and-digital-humanities/> [8-10-2017]
- [20] Roncaglia G. (2012), *La quarta rivoluzione. Sei lezioni sul futuro del libro*, Laterza

- [21] Salvatori E. (2017), *Digicraft and pensiero sistemico: le prospettive delle DH*, AIUCD 2017, Book of abstract, http://aiucd2017.aiucd.it/?page_id=14 [8-10-2017]
- [22] Thaller M. (2012), *Controversies around the Digital Humanities: An Agenda*, Historical Social Research / Historische Sozialforschung, 37/ 3: 7-23 <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html> [8-10-2017]
- [23] *The Cambridge Handbook of Information and Computer Ethics*, ed. by L. Floridi, (Cambridge University Press, 2010).
- [24] *The Digital Humanities Manifesto 2.0* http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf [8-10-2017]24
- [25] Volpe G. (2016), *Un'eredità da riconquistare. Libero accesso e riforma dei beni culturali*, in R. Auriemma (ed), *La democrazia della conoscenza: patrimoni culturali, sistemi informativi e open data: accesso libero ai beni comuni?* Atti del convegno (Trieste 28-29 gennaio 2016), Udine, pp. 13-22.
- [26] Volpe G., De Felice G. (2014), *Comunicazione e progetto culturale, archeologia e società*, in *European Journal of Post-Classical Archaeologies* 4/2014, pp. 405-424. ISSN 2039-7895s. www.logoi.ph

Digitalizzare la memoria.

Studi, ricerche e sperimentazioni in Terra d'Otranto

Giuliana Iurlano¹, Anna Maria Colaci², Demetrio Ria³,
Deborah De Blasi⁴, Giovanna Bino⁵, Francesca Salvatore⁶

¹ Università del Salento-Centro Studi Relazioni Atlantico-Mediterranee, Italia - giuliana.iurlano@unisalento.it

² Università del Salento, Italia – annamaria.colaci@unisalento.it

³ Università del Salento, Italia – demetrio.ria@unisalento.it

⁴ USP Lecce, Italia – deborahdeblasi@alice.it

⁵ MiBACT, Italia – giovanna.bino@beniculturali.it

⁶ Università del Salento-Centro Studi Relazioni Atlantico-Mediterranee, Italia – kikkasalvatore@hotmail.it

ABSTRACT

L'idea di digitalizzare la memoria storica locale, costruendo un archivio in cui riversare fonti inedite reperite in vario modo sul territorio, è nata nell'ambito del progetto "Cento anni fa... la Grande Guerra". L'esperienza ha permesso la creazione di un archivio in cui riversare fonti inedite reperite in vario modo sul territorio. L'intero ciclo di creazione, diffusione ed uso delle risorse digitali, secondo quella che viene ormai definita come "*precognitive utility net*", costituisce una vera e propria infrastruttura di supporto all'apprendimento. Si sta cercando di realizzare è un vero e proprio modello integrato per l'apprendimento e la ricerca (secondo l'idea di *e-science* o anche di *cyberinfrastructure*).

In tale prospettiva, diventa importante la didattica multimediale delle discipline storico-sociali, sostenuta dalle TIC. Lo sviluppo di grandi archivi materiali per sostenere la ricerca scientifica e promuovere la manutenzione-conservazione di artefatti del patrimonio culturale potrebbero consentire un cambio di rotta nella didattica delle discipline umanistiche, in particolare nell'ambito storico-sociale. Ciò può essere realizzato attraverso modelli didattici integrati con le tecnologie digitali. In questo contesto si colloca la riscoperta e la ricostruzione digitale del percorso storico che ricostruisca dal punto di vista educativo, il periodo del fascismo in Terra d'Otranto.

La digitalizzazione della memoria può riguardare anche due aspetti ancora poco diffusi sul piano didattico. Il primo è quello relativo alla lettura espressiva di un diario di un reduce della Grande Guerra, accompagnata dalla ricostruzione sonora dei suoni legati agli eventi bellici. L'altro aspetto è legato al mondo del *game* nella sua accezione più ampia. La storia, ormai, non può essere esente dal processo di *gamification*, basato sull'idea di inoculare l'apprendimento profondo all'interno di una esperienza piacevole di gioco digitale, per fortuna "riabilitato" dalla didattica della storia.

L'esperienza di una estesa rete territoriale, realizzata nel corso degli ultimi tre anni nell'ambito del progetto "Cento anni fa... la Grande Guerra", ha visto – tra le innumerevoli iniziative – anche alcuni interventi di reperimento di fonti cartacee e materiali inedite e di recupero e tutela del patrimonio culturale, in particolare archivistico-bibliotecario, proprio di alcuni istituti scolastici storici di Terra d'Otranto. È nata, così, l'idea di "digitalizzare" la memoria storica locale, costruendo un archivio in cui riversare fonti inedite reperite in vario modo sul territorio, in un processo la cui funzionalità è estremamente flessibile e molteplice: innanzi tutto, una piattaforma digitale consentirebbe di rinforzare la rete territoriale esistente, che già di per sé è uno strumento di integrazione socio-culturale molto forte; poi, costituirebbe lo strumento più efficace per individuare/reperire/recuperare fonti inedite spesso di origine familiare o scolastica, mettendole in rete mediante un accesso veloce, esteso a tutti e, soprattutto, coprendo l'intero ciclo di creazione, diffusione ed uso delle risorse digitali, secondo quella che viene ormai definita come "*precognitive utility net*", una vera e propria infrastruttura di supporto all'apprendimento. Ciò che si sta cercando di realizzare è un vero e proprio modello collaborativo per l'apprendimento e la ricerca (secondo un modello di *e-science* o anche di *cyberinfrastructure*), nel quale la ricerca sul campo-effettuata dagli studenti delle scuole secondarie, con la guida e la supervisione dell'Accademia- si coniuga con il processo di apprendimento della metodologia della ricerca storica, della conoscenza degli argomenti oggetti di studio, e dell'utilizzo dell'innovazione digitale per trasferire la fruizione delle fonti individuate, reperite in ambito più ampio. L'archivio digitale sarà infatti realizzato dagli stessi studenti anche in modalità di alternanza scuola-lavoro e costituirà la fase successiva della ricerca. Potrà servire anche per la progettazione e realizzazione di percorsi culturali, turistici e virtuali nel territorio di Terra d'Otranto.

In tale prospettiva, allora, si comprende l'importanza della didattica multimediale delle discipline storico-sociali, sostenuta dalle tante opportunità che le TIC – sempre più impegnate nel campo dei beni culturali – stanno offrendo. Lo sviluppo di grandi archivi materiali per sostenere la ricerca scientifica e promuovere la manutenzione/conservazione di artefatti del patrimonio culturale potrebbero consentire un cambio di rotta nella didattica delle discipline umanistiche, in particolare nell'ambito storico-sociale. Ciò potrebbe avvenire individuando modelli didattici integrati con le tecnologie digitali, che contribuiscano a far crescere il senso di tutela del patrimonio, come pure la capacità di riconoscerlo nella fase di formazione e di valorizzarlo a livello locale.

In questo contesto si colloca la riscoperta e la ricostruzione digitale del percorso ideologico costruito dal fascismo nell'ambito dell'educazione. All'interno della didattica della scuola elementare di epoca fascista, infatti, è interessante portare alla luce – e alla memoria storica attraverso la digitalizzazione delle fonti – i metodi e le tecniche formative attuate dai maestri al fine di costruire una coscienza patriottica e di popolo, ma anche tutte le fonti iconografiche e fotografiche che sostenevano il processo di rafforzamento della propaganda in tutti gli strati della società.

La fruizione delle fonti pubbliche, dunque, diventerebbe un momento essenziale della trasformazione digitale dei documenti presenti negli archivi di stato, comunali, scolastici o privati: il “documento”, così, non sarebbe più solo oggetto di studio di utenti accademici, ma un elemento della cui complessità la digitalizzazione può dar conto, focalizzandone tutti gli aspetti e le possibilità di lettura, di interpretazione e di fruizione.

Ma la digitalizzazione della memoria può riguardare anche due aspetti ancora poco diffusi sul piano didattico. Il primo è quello, già sperimentato in alcune classi del primo ciclo della scuola primaria, relativo alla lettura espressiva di un diario di un reduce della Grande Guerra, accompagnata dalla ricostruzione sonora dei suoni legati agli eventi bellici, ma anche a quelli della quiete dopo lo scontro, suoni tutti riprodotti ed analizzati nella loro caratteristica ritmica, dinamica e timbrica. L'altro aspetto è legato al mondo del *game* nella sua accezione più ampia. La storia, ormai, non può essere esente dal processo di *gamification*, basato sull'idea di inoculare l'apprendimento profondo all'interno di una esperienza piacevole di gioco digitale, per fortuna “riabilitato” dalla didattica della storia. La capacità ludica della *game history*, infatti, può offrire esperienze di immedesimazioni in azioni e processi decisionali degli attori del passato.

PAROLE CHIAVE

Archivio digitale; Public history; Fonti educative; Didattica multimediale; Digital humanities; Digitalizzazione; Innovazione didattica; Gamehistory.

1. “MEMORIE DAL SOTTOSUOLO”. UN ARCHIVIO DIGITALE ONLINE DI PUBLIC HISTORY

“Memorie dal sottosuolo” è la prosecuzione ad un livello qualitativamente più alto del progetto “Cento anni fa...la Grande Guerra”. Esso intende utilizzare la rete già esistente delle scuole che hanno aderito e realizzato i laboratori scolastici per reperire, censire ed elaborare fonti e reperti inediti sulla Grande Guerra. Esso consiste nella realizzazione di un polo archivistico di ambito territoriale (l'ex Terra d'Otranto), in cui possano essere riversate tutte le fonti inedite reperite attraverso i laboratori scolastici, sotto la supervisione scientifica del CeSRAM (Centro Studi Relazioni Atlantico-Mediterranee), dell'Istituto per la Storia del Risorgimento Italiano-Comitato di Lecce e dell'Università del Salento.

2. LE FONTI EDUCATIVE DEL FASCISMO. RISCOPERTA E RICOSTRUZIONE DIGITALE DI UN PERCORSO IDEOLOGICO

Il presente intervento si propone di ricercare all'interno della didattica della scuola elementare di epoca fascista i metodi e le tecniche formative attuate dai maestri al fine di costruire una coscienza patriottica e di popolo. Lo studio scaturisce da una ricerca d'archivio condotta in alcune scuole elementari di Terra d'Otranto, approfondendo la condizione della realtà rurale, andando alla scoperta dei costumi, dell'igiene, del rapporto delle famiglie con la realtà scolastica e della programmazione mensile ed annuale, indirizzata verso il culto della nazione e del regime fascista. Importanti, in tal senso, sono le testimonianze dirette dei maestri che, attraverso le annotazioni riportate sui giornali di classe, hanno consentito una peculiare ricostruzione dell'ambiente scolastico e della gioventù che si appropiava alla crescita sotto l'egida del fascismo.

3. DIDATTICA MULTIMEDIALE DELLE DISCIPLINE STORICO-SOCIALI: TRA RICERCA EDUCATIVA E VALORIZZAZIONE DEL PATRIMONIO CULTURALE LOCALE

La formazione critica verso la tutela del patrimonio culturale nella maggior parte dei paesi europei (in particolare in Italia) è stata per lungo tempo basata sui metodi didattici tradizionali. Oggi le opportunità che le TIC, sempre più impegnate nel campo dei beni culturali, stanno offrendo sono tantissime. Lo sviluppo di grandi archivi di materiali per sostenere la ricerca scientifica e promuovere la manutenzione/conservazione di artefatti del patrimonio culturale potrebbero consentire un cambio di rotta nella didattica delle discipline umanistiche, in particolare dell'ambito storico-sociale. L'obiettivo principale è quello di individuare modelli didattici integrati con le tecnologie digitali che possano contribuire a far crescere il senso di tutela del patrimonio, come pure la capacità di riconoscerlo nella fase di formazione. Un valore aggiunto nell'adozione delle TIC nell'educazione, poi, è che ciò possa incidere nella valorizzazione del patrimonio culturale, in particolare di quello locale. La domanda che ci si pone è: le TIC possono realmente fornire un valore aggiunto alla pedagogia, all'educazione e all'apprendimento del patrimonio culturale? L'impatto pedagogico che può derivare da un accesso più facile e una visione multi-prospettica dei manufatti del patrimonio culturale può anche arricchire e migliorare l'educazione ai beni culturali, sociali e comunitari grazie all'adozione di metodi innovativi di apprendimento/insegnamento.

4. SUONI E COLORI DALLA GUERRA. LA DIGITALIZZAZIONE COME ESPERIENZA DI RICERCA DIDATTICA

In alcune classi della scuola primaria è stata affrontata la tematica della Grande Guerra attraverso le pagine del diario di un reduce, che scrisse della sua esperienza di trincea, pagine che sono state trasformate in lettura espressiva e video-animata finalizzata all'ascolto e all'approfondimento. Dopo la visione del breve filmato in cui erano state sintetizzate due pagine della narrazione (la prima legata ad una fase di scontro bellico, la seconda legata ad un momento di tregua, ad una mattina di apparente quiete), tali momenti – diametralmente opposti in termini di eventi, condizioni emotive, rumori e suoni – sono stati visualizzati sulla LIM di classe da tutti gli alunni in classi aperte, riprodotti e analizzati nelle loro caratteristiche

ritmica, dinamica e timbrica. Tali “condizioni” sonore dell’animo e dell’habitat sono state espresse dai bambini prima attraverso la propria voce, il proprio corpo, successivamente con l’uso del tablet. Ciascun alunno è stato fornito di questo strumento, sul quale era stato caricato un programma specifico in grado di fornire una vasta gamma di strumenti sonori e ritmici con i quali creare le sonorizzazioni. Ciascun bambino, avendo già approfondito la conoscenza del testo narrativo ed avendo già ben discusso e condiviso le sensazioni sonore ad esso legate, ha cercato, nel proprio tablet, il colore, o timbro, adeguato ad esprimere il messaggio. È nato, così, il tappeto sonoro che ha sotteso l’intero brano testuale e che ne ha rafforzato il senso e gli obiettivi.

5. L’ARCHIVIO PRET À PORTER. LA FRUIZIONE DIGITALE DELLE FONTI PUBBLICHE

La cultura rappresenta un *driver* importante nel quadro della strategia dedicata allo sviluppo della macroregione adriatico-ionica, area nella quale il “patrimonio culturale”, in particolare il “patrimonio di carte” depositato e fruito negli Archivi di Stato, viene inteso quale specifica dotazione territoriale, in quanto infrastruttura materiale o immateriale. Tale riserva “aurea” per la cultura ha necessità di un linguaggio differente di comunicazione e di immediata circolazione, non essendo più, infatti, oggetto di studio di soli utenti accademici. Il primo obiettivo strategico è, dunque, quello di puntare sulla sua trasformazione digitale, cioè sulla “crescita digitale” del “documento”, in un’ottica centrata sull’utente. Ciò consentirà di entrare in archivio sfogliando le sue pagine web, di curiosare nella sua storia, di passare dalle antiche pergamene della città e della sua provincia per approdare alle opportunità offerte dalla tecnologia più avanzata, il tutto raccontato per immagini e suoni. Non solo Archivi di Stato, ma anche archivi e biblioteche scolastiche, spesso intrappolate negli scaffali polverosi di luoghi sino ad oggi poco frequentati, costituiscono il patrimonio “rivelato” da rendere fruibile in rete alla città, in quanto storia di istituzioni e storia dei suoi cittadini.

6. LA GAMEHISTORY. L’INNOVAZIONE DIDATTICA FRA GIOCO E REALTÀ

Una volta considerato responsabile dei comportamenti violenti e asociali dei giovanissimi, il mondo del *game*, considerato nella sua accezione più larga, oggi viene celebrato per la sua capacità potenziale di trasformare tecniche didattiche e pratiche di apprendimento. La storia non è esente da questo processo di “*gamification*” che si concentra sull’idea di inoculare l’apprendimento profondo all’interno di una esperienza piacevole: il gioco digitale, infatti, offre la possibilità di sperimentare l’esecuzione di obiettivi di sviluppo della conoscenza storica senza essere consapevoli dell’obiettivo stesso (conoscenza *by stealth*, ovvero “in modalità nascosta”). Ma il mondo della *game history*, come articolazione della *public history*, non si rivolge ormai più solo ai piccoli, bensì a un pubblico più adulto, che richiede, come mercato, un elevato livello di autenticità storica. I prodotti di *gaming* impiegano, infatti, sempre più accademici qualificati per ricercare contenuti, consigliare gli sviluppatori e fornire al giocatore materiale di supporto per migliorare le sue esperienze in-game. Se si accettano i giochi come facilitatori di conoscenza storica, è semplice pensare come questi possano offrire al giocatore esperienze di immedesimazione in azioni e processi decisionali degli attori del passato. La capacità ludica della *game history* può consentire, infatti, una comprensione approfondita di fatti, date, persone e eventi ma anche la complessa rete di condizioni, contingenze, errori e circostanze che contribuiscono ad una vera comprensione della storia.

7. BIBLIOGRAFIA

- [1] Atkins, D.E. 2003. *NSF Blue Ribbon Advisory Panel on Cyberinfrastructure*. Alexandria: National Science Foundation.
- [2] Borgman, C. 2000. *Digital libraries and the continuum of scholarly communication*. In: Journal of Documentation.
- [3] Bortone, R. 2015. *Fare scuola in un mondo che cambia*. Aradeo: Nikeditrice.
- [4] Bostrom, R.P. - Gupta, S. - Hill, J.R. 2008. *Peer-to-peer technology in collaborative learning networks: Applications and research issues*. In: International Journal of Knowledge and Learning.
- [5] Branchesi, L., Branchesi, L. 2006. *Il patrimonio culturale e la sua pedagogia per l’Europa: bilancio e prospettive*. Roma: Armando
- [6] Brown, J.S. - Collins, A. - Duguid, P. 1989. *Situated cognition and the culture of learning*, In: Educational Researcher.-Iurlano, G., “Notes from Underground”: a Digital Archive of Public History on Network, in «Journal of Education & Social Policy», IV, 3, Sept. 2017, pp. 28-32
- [7] Jedlowski, P. 2016. *Intenzioni di memoria. Sfera pubblica e memoria autocritica*. Milano: Mimesis.
- [8] Lytras, M.D. 2007. *Teaching in the knowledge society: An art of passion*. In: International Journal of Teaching and Case Studies.
- [9] March, T. 2003. *The learning power of WebQuests*. In: Educational Leadership.
- [10] Meyer, É., Grussenmeyer, P., Perrin, J. P., Durand, A., & Drap, P. A. 2007. *Web information system for the management and the dissemination of Cultural Heritage data*. In: Journal of Cultural Heritage.
- [11] Noiret, S. 2011. *La Public History: una disciplina fantasma?* In: Memoria e ricerca.
- [12] Orlandi, T. 2009. *Quantità e qualità. I testi, le biblioteche e l’accesso alle informazioni*. In: Informatica umanistica.
- [13] Politou, E.A. - Pavlidis, G.P. - Chamzas, C. 2004. *JPEG2000 and the dissemination of Cultural Heritage databases over the Internet*. In: IEEE Transactions on Image Processing.
- [14] Propp, V. 1988. *Morfologia della fiaba*. Torino: Einaudi.
- [15] Riva, E. 2013. *Digital Humanities e Digital History; una nuova cittadinanza dei saperi*. In: Annali di Storia Moderna e Contemporanea.
- [16] Rodari, G. 1973. *Grammatica della fantasia*. Torino: Einaudi.
- [17] Roussou, M. 2002. *Virtual heritage: From the research lab to the broad public*, In: Proceedings of the VAST Euroconference.
- [18] Sacks, O. 2008. *Musicofilia*. Milano: Adelphi.
- [19] Sefton-Green, J. 2004. *Informal learning with technology outside school*. Bristol: Futurelab

[20] Testa, A. 2010. *La trama lucente*. Milano: Rizzoli.

[21] Vitali, S. 2004. *Passato digitale. Le fonti dello storico nell'era del computer*. Milano: Bruno Mondadori.

[22] Veltman, K. H. 2005. *Challenges for ICT/UCT applications in Cultural Heritage*. In: C. Carreras (Ed.). *ICT and heritage*.

[23] Vygotsky, L. 1962. *Thought and language*. Cambridge: M.I.T. Press (Originally published in 1934).

La storia alla prova dei like.

Il caso del neoborbonismo mediatico e digitale

The history to the test of the like.

The case of media and digital neoborbonism

Annastella Carrino¹ (coordinatrice del panel), Gian Luca Fruci² (coordinatore del panel),
Christopher Calefati³, Antonella Fiorio⁴, Maria Teresa Milicia⁵,
Federico Palmieri⁶, Carmine Pinto⁷, Silvia Sonetti⁸, Claudia Villani⁹

¹Università di Bari, annastella.carrino@uniba.it

²Università di Bari, gianluca.fruci@uniba.it

³Università di Bari, christophercalefati@libero.it

⁴Università di Bari, antonella.fiorio@libero.it

⁵Università di Padova, mariateresa.milicia@unipd.it

⁶Università di Bari, fopalmieri@hotmail.it

⁷Università di Salerno, cpinto@unisa.it

⁸Università di Salerno, ssonetti@unisa.it

⁹Università di Bari, claudia.villani@uniba.it

ABSTRACT

Nei contesti democratici occidentali la dialettica delle conoscenze si è a lungo mantenuta dentro un ambito “costituzionale” di garanzia. Oggi la messa in discussione di tale democrazia “garantista” investe anche la definizione – non solo accademica – dei saperi esperti. Chi conquista più consensi o più *like* si sente abilitato a delegittimare luoghi e attori di produzione neutri, professionali, ed esterni alle arene politiche e ai processi mitografici ad esse funzionali.

La ricerca storica viene accusata di essere élitaria e oscura, e, di fatto, di non prestarsi alle operazioni neoidentitarie e micronazionalistiche in atto in ogni parte del mondo globalizzato. Nell’era post-industriale e digitale un ruolo fondamentale rivestono i circuiti comunicativi tramite cui si irradiano i saperi umanistici. Attraverso alcuni di questi canali, i saperi che dovrebbero essere “esperti” circolano in maniera spesso incontrollata, decontestualizzata; procedono per *slogans* e per accumulazione; si basano su reazioni immediate ed emotive, finendo per confondere storia e memorie, per raccontare “verità” storiche dimentiche di fonti e metodologie.

Il panel intende indagare la tensione che si determina fra flussi di informazioni e narrazioni storiche (spesso sotto forma di *scoops* sensazionalistici) disponibili online su siti, blog, pagine facebook, canali youtube, e la loro selezione critica da parte sia del pubblico ‘generalista’ della rete sia delle agenzie educative e delle istituzioni rappresentative chiamate rispettivamente a lavorare o a deliberare sul passato. Al corollario immaginifico di un’accessibilità autonoma e diretta alle verità scientifiche fa riscontro, infatti, nella realtà l’esigenza di un *surplus* – e non di una diminuzione – di intermediazione critica al fine di radicare il discorso pubblico su una base di conoscenze correttamente accertate. Ci si focalizzerà sul caso del neoborbonismo mediatico e digitale allo scopo di analizzare la costruzione, la circolazione e la ricezione delle mitografie sudiste e neomeridionali e la loro performatività sia politica che istituzionale.

In the democratic Western contexts the cultural dialectics of knowledge has long been maintained within a framework of "constitutional" warranty. Today the crisis of this guaranteed democracy also affects the definition of the expert knowledge, not only academic. Those who conquer more consensus or more like feel empowered to undermine the places and the actors of the neutral and professional production and in more general terms all what that is external to the political arenas and to mythographic processes. The historical research is accused of being elitist and obscure, but, in fact, the historians are accused of lack of support the identitary and micronationalistic movements that grow everywhere in the globalized world.

In the post-industrial and digital era, the communication circuits play a fundamental role in attacking the humanistic knowledge. Through some of these channels, the knowledge that should be expert circulates in a manner often uncontrolled; the circulation of historical narratives is based on slogans, i.e. on immediate and emotional of explication and presentation. History and memories are confused; the historical "truth" is told without sources and appropriate methodologies.

The panel intends to investigate the tension that is determined between the flows of information and historical narratives (often in the form of sensational scoops). This tension is particularly available online (sites, blogs, Facebook pages, youtube channels): the corollary is that of an autonomous and direct accessibility to the scientific truths. In truth, this tension demands a surplus and not a diminution of critical intermediation: the purpose should be to found the public discourse on knowledges correctly verified. Finally the panel will focus on the case of the media and digital neoborbonism with a view to analysing the building, the circulation and the reception of the new southern mytographies and their performativity, both in the political and in the institutional sphere.

PAROLE CHIAVE

(Neo)borbonismo, circuiti comunicativi multimediali, public history, antistoria, memoria

Neoborbonism, communication circuits, public history, antihistory, memory

1. PANEL ABSTRACT

Nei contesti democratici occidentali la dialettica delle conoscenze si è a lungo mantenuta all'interno di un ambito "costituzionale" di inquadramento e di garanzia, entro il quale hanno circolato i saperi esperti. Oggi la messa in discussione di questa forma di democrazia "garantista" investe anche la definizione – larga e non esclusivamente accademica – delle conoscenze. Chi conquista più consensi o più *like* si ritiene legittimato a farsene portatore, finendo per delegittimare – attaccandoli – luoghi e attori di produzione neutri perché professionali, esterni alle arene politiche e ai processi mitografici e identitari ad esse funzionali. Questa dinamica fa parte di una più ampia contestazione degli istituti di garanzia, visti come un unico universo oligarchico che opprime il flusso sciolto di pensieri e opinioni "fai da te".

In particolare, la ricerca storica viene attaccata con l'accusa di essere élitaria, oscura, di mirare solo ad autoconservarsi e, in fondo, di non prestarsi alle molteplici operazioni neoidentitarie e micronazionalistiche in atto in ogni parte del mondo globalizzato. Ricerca, appunto, ovvero risultato di un lavoro lungo e puntuale, fondato su tecniche e metodologie precise, e su un'attitudine alla revisione permanente, a prescindere dalla collocazione universitaria e non dei suoi attori. Le produzioni di sapere sono, infatti, sofferte e contraddittorie, e perciò stesso affidabili, proprio perché non dogmatiche; pronte a rimettersi in discussione, ma su basi scientifiche e seriamente documentate. Questo dato è assai evidente negli ambiti medico-biologico, ma non è meno vero per le scienze sociali e umanistiche.

Le società si danno proprie memorie: dunque è normale che gruppi, comunità, formazioni politiche elaborino, ieri come oggi, visioni del passato funzionali alla costruzione di identità vecchie e nuove, sulle quali non c'è – e non può esserci – alcun monopolio accademico, né giornalistico o politico. Le memorie – al plurale – sono inevitabilmente divise e divisive, stratificate e conflittuali. Il punto cruciale è quindi, da un lato, il contesto della loro formazione e i toni del dibattito in cui esse si collocano; dall'altro, l'apparente "cordialità" e accessibilità del linguaggio e del mestiere dello storico che legittima chiunque ad appropriarsene, in modi talvolta avventurosi o, peggio, strumentali. Infine, fondamentali diventano nell'era post-industriale e digitale i circuiti comunicativi su cui viaggiano e con cui si irradiano i saperi umanistici, e in particolare la conoscenza storica. Attraverso alcuni di questi canali, quei saperi che dovrebbero essere "esperti" circolano in maniera spesso incontrollata, decontestualizzata; procedono per slogan e per accumulazione; si nutrono di consensi determinati da reazioni impulsive ed emotive, sganciati dal sapere stesso, finendo per confondere storia e memorie, per raccontare una "verità" storica che ha dimenticato le sue fonti ed i suoi presupposti teorici e metodologici.

Scopo di questo panel è indagare la tensione che si determina continuamente fra i flussi di informazioni, dati, interpretazioni, narrazioni storiche (non di rado sotto forma di presunti *scoops* sensazionalistici) disponibili online tramite siti, blog, pagine facebook, canali youtube, e la loro gestione e selezione critica da parte sia del pubblico "generalist" della rete sia delle agenzie educative e delle istituzioni rappresentative chiamate rispettivamente a lavorare o a deliberare sul passato. Al corollario immaginifico di un'accessibilità diretta, autonoma e immediata alle verità scientifiche fa riscontro, infatti, nella realtà l'esigenza di un *surplus* – e non già di una diminuzione – di intermediazione critica ed esperta al fine di radicare il discorso pubblico – in qualsiasi direzione orientato – su una base condivisa di conoscenze storiche correttamente accertate. In particolare, i fuochi analitici saranno orientati sul caso della del neoborbonismo mediatico e digitale al fine di indagare la costruzione, la circolazione e la ricezione delle mitografie sudiste e neomeridionali ad esso connesse e la loro performatività sia politica che istituzionale, come dimostra, nei mesi di marzo e di luglio scorsi, l'approvazione da parte dei consigli regionali di Basilicata e Puglia di mozioni miranti alla celebrazione di giornate della memoria delle vittime meridionali del Risorgimento in coincidenza con la vicenda e la data iconiche del legittimismo borbonico e del nazional-patriottismo duosiciliano: la resa di Gaeta del 13 febbraio 1861.

Il panel si articolerà in sei interventi che intendono, da un lato, interrogarsi più in generale sulla nuova sfera pubblica digitale costellata di *media memories* indagandone la segmentazione dei caratteri originali (Claudia Villani) e le principali piattaforme italiane online in ambito storico (Christopher Calefati – Antonella Fiorio), dall'altro analizzare in particolare il revival neoborbonico attraverso una ricognizione della sua articolata e variegata galassia web (Federico Palmieri), un'archeologia della rielaborazione prima postrisorgimentale e poi postnovocentesca del discorso legittimista borbonico dilatato e rilanciato recentemente grazie all'apporto di un articolato universo comunicativo multimediale (Gian Luca Fruci – Carmine Pinto), un *case study* dedicato a Maria Sofia di Borbone quale icona della controrivoluzione transnazionale fra passato e presente (Silvia Sonetti) e una ricerca partecipante di taglio antropologico condotta all'interno della galassia digitale postmeridionalista e neoborb (Maria Teresa Milicia).

2. Claudia Villani, MEMORIE COLLETTIVE E NUOVI MEDIA DIGITALI NELLA GLOBAL PUBLIC SPHERE

Negli anni Settanta si avvia il processo di decolonizzazione (Nora, 2016) e pluralizzazione delle memorie collettive in cui siamo immersi oggi, che scardina le narrazioni storiche maggioritarie, siano esse nazionali, europee, occidentali. Anche dall'esito dello scontro tra diverse memorie collettive dipende, in fondo, «ciò che diventa storia e ciò che la storia diventa», come emerge negli studi sulla costruzione delle narrazioni storiche nazionali (Berger, 2007), sulle cosiddette memorie politiche, istituzionali, funzionali, e sulla loro interazione con altre memorie collettive (culturali, sociali, territoriali, generazionali, ecc.). Gli studi più recenti mostrano come non solo "locale" e "globale" si influenzano a vicenda nella complessiva rinegoziazione delle memorie (attraverso cui passa la ridefinizione di identità, valori e forme di integrazione sociale), ma il ruolo dei media e delle tecnologie della comunicazione diventa sempre più rilevante. Nella nuova *global*

public sphere proliferano *media memories* che si prestano a svariate forme di consumo e di uso, orientate a produrre una identificazione immediata (Assmann-Conrad, 2010). Sono memorie attraversate da processi di iconizzazione ed estetizzazione, «detextualised and isolated, symbolically dramatised as images, and made absolute in its validity» (Kaschuba, 2010), che si sottraggono alla riflessione e all'argomentazione. L'interazione tra memoria e media, tra politica e media, contribuisce quindi a definire le attuali *globalizing memory cultures*, e un nuovo regime di storicità, caratterizzato da alcuni elementi comuni: riparazione, giuridicizzazione, vittimizzazione e denazionalizzazione (Rouso, 2007). In questo intervento confronteremo gli esiti dei più recenti studi su questi temi con il caso di studio proposto nel panel.

Collective Memories and New Digital Media in the Global Public Sphere

It was in the 1970s that began the process of decolonization (Nora, 2016) and pluralization of collective memories we are still immersed in, a process that unhinged most historical narratives, whether national, European or Western ones. After all, «what becomes history and what history becomes» depends also on the outcome of the clash of different collective memories, as shown by the studies on the construction of national historical narratives (Berger, 2007), and by those on the so-called political, institutional, functional memories and their interaction with other collective memories (whether cultural, social, territorial, generational ones, etc.). The most recent studies point out that not only the “local” and the “global” influence each other in this complex renegotiation of memories (through which identities, values and forms of social integration are redefined), but that the role of the media and communication technologies is becoming more and more relevant. In the new global public sphere there is a proliferation of media memories available for various forms of consumption and use and aimed at producing an immediate identification (Assmann-Conrad, 2010). These memories are crossed by processes of iconization and aestheticization, «detextualised and isolated, symbolically dramatised as images, and made absolute in its validity» (Kaschuba, 2010), they escape reflection and reasoning. Thus the interaction between memory and media, between politics and media, helps defining the current globalizing memory cultures and a new regime of historicity, characterized by some shared elements: reparation, juridicization, victimization and denationalization (Rouso, 2007). In this paper of mine I will compare the case study proposed in the panel with the outcomes of the most recent studies on these themes.

3. Christopher Calefati – Antonella Fiorio, STORIA 4.0. IL RISORGIMENTO NELLA RIVOLUZIONE DIGITALE

Nell'era digitale la ricerca storica può svelare insidie dovute all'organizzazione e gestione dei numerosi canali online di divulgazione. Il Web è penetrato nelle discipline umanistiche come potente mezzo di comunicazione e condivisione, lasciando spazio a iniziative poco professionali che parimenti fruiscono delle preziose opportunità messe a disposizione dalla Rete (Noiret, 1999; Ortoleva 1999). A tal proposito, si reputa indispensabile porre il lettore generico nelle condizioni di valutare autonomamente l'attendibilità di siti e blog di carattere storico-culturale durante la navigazione (Bordini, 2008; Minuti, 2015). Partendo dagli studi di avanguardia effettuati nei primi anni 2000 che offrono una rassegna ragionata di una parte consistente dei siti di storia contemporanea (Criscione, Noiret, Spagnolo, Vitali, 2004), il proposito è quello di effettuare una nuova mappatura incentrata sul tema "Risorgimento e Mezzogiorno" seguendo i medesimi criteri di analisi ivi elaborati, valutandone la compatibilità con le nuove frontiere della comunicazione e della tecnologia, come i canali youtube divulgativi. L'intento successivo è quello di promuovere la riflessione sulla varietà degli input provenienti dalla Rete e quindi proporre, se possibile, nuovi parametri da integrare ai precedenti. Si offriranno modelli campione rispondenti ai canoni di una corretta *Public History*, comparati a esempi prodotti in area anglo-americana. Il quesito che emerge da tali presupposti è se sia possibile per le università e gli istituti di ricerca inserirsi in questo sconfinato campo d'azione in maniera concorrenziale e trovare, dunque, una sintesi tra le esigenze della complessità dell'indagine storica e la rapidità di ricezione delle informazioni richiesta dal pubblico dei lettori al tempo di Internet.

History 4.0: The “Risorgimento” into digital revolution

In the Digital Age, the historical research could chow dangers for the organization and management of the great numbers of Web's way of communication. The Web has come inside Human Disciplines as a powerful tool for sharing and reporting, leaving room to unprofessional ventures that could use this “world” in the same way. About that we need to put the Web's reader in condition to evaluate independently, during the Web's surfing, the soundness of historical websites and blogs. Starting from the work published on 2000 by the research group made up of Antonino Criscione, Serge Noiret, Carlo Spagnolo and Stefano Vitali – that have worked on Contemporary History's websites – the intention is to make a new mapping on the theme of “Mezzogiorno e Risorgimento” following the same analysis' standards submitted before, and also trying to evaluate their compatibility with the new tools of Web's technologic development like YouTube Communication Channels. The next step is to promote a thought on the multiplicity of inputs that come from Web and then we will suggest, if it's possible, new evaluation's standards that could integrate the others that already exist. We will, also, offer Anglo-Saxon's sample models of a right use of Public History. The question is whether it is possible for Universities and Research Institutes to get involved into this giant Web world in a competitive way and to find a compromise between the historical research's needs and information's fast reception needed by Web's readers.

4. Gian Luca Fruci – Carmine Pinto, MITOGRAFIE (NEO)BORBONICHE: MORFOLOGIA E CIRCUITI COMUNICATIVI FRA PASSATO E PRESENTE

Il discorso neoborbonico è presentato dai suoi *speakers* odierni come una narrazione inedita e sensazionale, addirittura capace di apportare nuova e veritiera luce sulla conoscenza del processo risorgimentale nel Mezzogiorno (Casalena, 2013). In realtà, si tratta di un *revival* sempre uguale di temi che abitano il discorso pubblico e la società meridionale praticamente dal momento della fine del Regno delle Due Sicilie, ovvero dalla resa di Gaeta del 13 febbraio 1861, diventata in presa diretta data simbolica della memoria legittimista borbonica (Pinto, 2016, 2017). Oggi si assiste alla rielaborazione e – se possibile – alla radicalizzazione postnovocentesca degli argomenti prodotti dalla propaganda duosiciliana prima nella guerra di resistenza allo stato unitario (1861-1866), poi nella nostalgica sopravvivenza della generazione degli sconfitti (1867-1914).

L'intervento intende ripercorrere le dinamiche della costellazione discorsiva borbonica dalla fine dell'Ottocento all'inizio del XXI secolo focalizzandosi sulla sua morfologia e sui suoi circuiti comunicativi multimediali (vecchi e nuovi) allo scopo di mostrare come il mito della mancata nazione duosiciliana si sia plasmato storicamente in contrapposizione mimetica a quello della nascente nazione italiana attingendo al canone nazional-patriottico ottocentesco con un'analoga ossessione per le comunità violate, il sangue dei martiri, l'onore degli eroi e la virtù della eroine, l'empietà dei traditori, la crudeltà dei nemici (Fruci - Pinto, 2017). E che ieri come oggi si assiste a un ribaltamento inversamente simmetrico (in chiave borbonica) della retorica risorgimentale, in cui al nazionalismo panitaliano è contrapposto il neonazionalismo panmeridionale, mentre il brigante Crocco o – nel caso pugliese – il sergente Romano si configurano come icone alternative a Garibaldi e a Vittorio Emanuele II.

Neo-bourbon mythographies: morphology and communication circuits between past and present

The neo-Bourbon argument is presented by its modern speakers as an unprecedented and sensational narration, even capable of bringing a new and truthful light on the knowledge of the Risorgimento process in the Italian Mezzogiorno (Casalena, 2013). In reality, it is a revival of themes always equal to itself that affects the public debate and the southern society from the end of the Two Sicilies Kingdom, more specifically from the capitulation of Gaeta on February 13, 1861, which can be considered a symbol of the legitimist Bourbon memory (Pinto, 2016, 2017).

Today we are witnessing to the re-elaboration and - if possible – to the post-twentieth-century radicalization of the arguments produced by the two Sicilian propaganda in the war of resistance to the unitary state (1861-1866), first, and in the nostalgic survival of the generation of the defeated (1867-1914), afterward. The intervention aims to retrace the dynamics of the Bourbon discursive constellation from the late eighteenth century to the beginning of the twenty-first century focusing on its morphology and its multimedia communication circuits (old and new) to show how the myth of the missing two Sicilian nation was shaped historically in a mimetic contrast to that of the nascent Italian nation drawing on the nineteenth-century national patriotic canon with a similar obsession for the violated communities, the blood of the martyrs, the heroes's honor and the virtue of the heroines, the wickedness of the traitors, the cruelty of enemies (Fruci-Pinto, 2017).

Yesterday as today we are witnessing to an inversely symmetrical (in Bourbon key) reversal of the Risorgimento rhetoric, in which pan-Italian nationalism is opposed to the new southern nationalism, while the brigand Crocco or - in the case of Puglia - the Roman sergeant are configured as alternative icons to Garibaldi and to Victor Emmanuel II.

5. Federico Palmieri, LA GALASSIA NEOMERIDIONALE: IL WEB COME STRUMENTO DI ANTIFORMAZIONE

Negli ultimi anni sul web si assiste a un proliferare di siti sulla vicenda risorgimentale nel Mezzogiorno, narrata in termini di conquista ed impoverimento di un Sud originariamente ricco, prospero e indipendente. Un numero crescente di movimenti, associazioni e testate neomeridionali infatti fanno ampio uso dei social media per diffondere versioni presentate come inedite e semplificazioni di una pagina di storia complessa come fu quella del processo di unificazione nazionale (Patruno, 2011).

Questo contributo si propone di indagare la galassia di movimenti neomeridionali che utilizza il web come un amplificatore e un mezzo con cui raggiungere facilmente e immediatamente migliaia di utenti con modalità proprie del tempo della "democrazia in diretta". Questi siti, pagine facebook, blog utilizzano un linguaggio diretto ed estrapolano immagini da trasformare in slogan facili ed efficaci oppure recuperano vecchie interviste, canzoni, film o serie televisive, da piegare alle ragioni della "causa" neoborb e più ampiamente neomeridionale.

La finalità dello studio è quella di decostruire queste contro-narrazioni, cercando di capire come e quando nascono, chi e cosa le alimenta, evidenziandone limiti e contraddizioni. La ricerca ci porrà di fronte a un gigantesco effetto Dunning-Kruger, in cui chiunque può sentirsi esperto di storia, politica, economia o medicina, pur senza averne le adeguate competenze (Dunning - Kruger, 1999).

The neoborb affair: internet and the success of a fake storytelling

In the last few years, the number of websites relating to the Italian Risorgimento has grown: in fact, the south of Italy before the national unification has been described as a rich, powerful and independent country. What is written in these misleading websites is the result of articles edited by movements, associations, and online newspapers whose aim is to

provide brand-new version of this complex historical process, oversimplifying it and presenting it as an unreleased, original version of the “official” history.

This paper will examine the cluster of these neo-borbonici and neomeridionalisti movements, who use the websites for sharing their ideas and quickly reaching a large number of people. They use everything they can, including images, videos and audios, as proof to support their theories. The methodology and the scientific rigor have no importance at all: all that matters is being simple, effective and believable.

The aim of this article is therefore to deconstruct these counter-narrations, trying to figure out how and when they have been generated, who and what supports them, whilst at the same time underlining their limits and contradictions. The research will highlight a significant Dunning-Kruger’s effect, where everyone is able to express his opinion about everything, including subjects about which he has no competence.

6. Silvia Sonetti, CLIO SALVI LA REGINA. MARIA SOFIA DI BAVIERA TRA LEGGENDA E MISTIFICAZIONI NEL RACCONTO DEL WEB

133mila risultati google, più di mille articoli in vendita su ebay, quasi 1500 video su youtube. Questi alcuni dei numeri in rete di Maria Sofia di Baviera, ultima regina delle Due Sicilie.

Sorella della più famosa Sissi, Maria Sofia giunse a Napoli nel maggio del 1859, poco prima che la morte di Ferdinando II e la guerra nella pianura padana di Cavour cambiassero l’equilibrio geopolitico della penisola. Il suo arrivo sconvolse la città come la corte per i suoi modi anticonformisti e ribelli in evidente contrasto con l’impostazione retriva e bigotta della famiglia reale. Divenne regina a 18 anni e cercò subito di emergere come interlocutore nella vita del Regno sostenendo Filangieri e la via costituzionale (Petacco, 1995).

Dopo la fuga da Napoli il protagonismo politico e mediatico di Maria Sofia si fece sempre più evidente, affiancando e in qualche modo superando anche quello del re. Nei 150 giorni della durata dell’assedio di Gaeta non si risparmiò nel visitare le caserme, fare sopralluoghi o predisporre le cure ai feriti. Utilizzò la sua immagine come strumento di propaganda distribuendo le sue foto ai soldati e mostrandosi in prima linea sui bastioni della cittadella. Gaeta la consacrò indiscussa eroina dell’Antirisorgimento e a partire da lì, il suo mito attraversò i cambiamenti epocali tra Otto e Novecento (De Lorenzo, 2008).

Negli ultimi trent’anni, la combinazione tra il successo del neoborbonismo e l’esplosione degli strumenti di informazione consentita dalla rete, le ha restituito una vita 2.0 in cui la leggenda si confonde con la storia e diventa “verità”. Lo studio di questo caso può contribuire a spiegare le ragioni di un fenomeno, quella della rielaborazione delle culture dei vinti, che, per le sue proporzioni sempre più vaste, non solo anima il discorso pubblico ma interroga e sollecita il dibattito scientifico.

Clio save the queen. Maria Sofia of Bavaria between legend and mystification in the story of the web

133000 google results, more than a thousand items on sale on eBay, almost 1500 videos on YouTube. These are figure as found on line about Maria Sofia of Bavaria, the last queen of the Two Sicilies. Sister of the most famous Sissi, Maria Sofia arrived in Naples in May 1859, just before the death of Ferdinand II and the war in the Po valley of Cavour, two events that changed the geopolitical equilibrium of the Italian peninsula. Her arrival shocked the city as well the court for the unconventional and rebellious manners in clear contrast to the archaic and bigoted approach of the royal family. She became queen when she was 18 and she immediately tried to emerge as aninterlocutor in the life of the Kingdom by supporting Filangieri and the constitutional way. After the escape from Naples, the politic and media attention-seeking behaviour of Maria Sofi became more and more evident, flanking and in some way even surpassing the one of the king.

During the 150 days of the siege of Gaeta, she even visited the barracks, made inspections and arranged for the wounded. She used her image as a propaganda instrument, distributing her photos to the soldiers and showing herself in the front lines of the citadel's ramparts. Gaeta consecrated her the undisputed heroin of Antirisorgimento, and starting from there, her myth crossed the epochal changes between the nineteenth and twentieth centuries. In the last thirty years, the combination of the success of neo-Bourbonism and the explosion of the availability of resources found on line, have made her to return in a 2.0 life in which legend is tangled with history and becomes truth. The study of this case can contribute to explain the reasons for the emerging of such phenomenon, that of the re-elaboration of the cultures of the vanquished. This, due to its ever-expanding proportions, is not only animating the daily public discourse, but it is also questioning and soliciting the current scientific debate.

7. Maria Teresa Milicia, ANTROPOLOGIA DELLE POLITICHE DIGITALI POSTMERIDIONALISTA. MOBILITAZIONE DELLA STORIA, COSTRUZIONE DELLA MEMORIA E RIFONDAZIONE DEL REGNO DELLE DUE SICILIE

La controversa istituzione del “giorno della memoria delle vittime meridionali dell’unità d’Italia” da parte della Regione Puglia nel luglio 2017, seguita a quella meno clamorosa dei comuni della Valle del Torbido in provincia di Reggio Calabria, è l’esito di un lungo processo di espansione politica del movimento neoborbonico, fondato a Napoli nel 1993. La mobilitazione della “vera storia” e la ri-costruzione celebrativa dei “luoghi di memoria” del Regno delle Due Sicilie hanno costituito il nucleo simbolico forte delle rivendicazioni politiche identitarie del movimento. Grazie a un’élite di attivisti appassionati e simpatizzanti di alto rango, si è consolidata una rete di alleanze territoriali capace di resistere, per oltre un ventennio, alla rappresentazione pubblica che la confinava nella marginalità del fenomeno folkloristico. L’avvento della comunicazione digitale segna una svolta per l’attivismo neoborbonico. Le risorse mitopoietiche della rivelazione della vera

storia del Regno delle Due Sicilie si caricano del potenziale performativo della reiterazione nell'ambiente interconnesso dei nuovi media (Couldry – Hepp, 2016). Le proprietà tecnologiche della comunicazione interattiva digitale imprimono una velocità di diffusione inedita ai contenuti che possono essere riappropriati e riprodotti dalla creatività di ogni nodo individuale della rete. Questo intervento si basa sulla rielaborazione dei dati dell'esperienza etnografica all'interno del gruppo facebook "No Lombroso", aggregatosi intorno al movimento di protesta contro l'apertura del Museo Cesare Lombroso di Torino nel 2009 (Milicia 2014). La mobilitazione contro il Museo è partita dai leaders del movimento neoborbonico e ha rappresentato un passaggio decisivo per l'aggregazione delle "moltitudini" digitali – rappresentanze di movimenti territoriali di ispirazione meridionalista e/o singoli individui – alla ricerca di spazi di socializzazione e di riconoscimento (Montaldo, 2012). L'osservazione partecipante delle pratiche comunicative quotidiane all'interno del gruppo facebook consente di proporre un modello rituale di istituzione di una nuova soggettività politica "postmeridionalista", basata sulla costruzione dell'identità vittimaria, della memoria e del senso di appartenenza al Regno delle Due Sicilie (Milicia, 2014a, 2015, 2016, 2106a).

Anthropology of "postmeridionalist" digital politics. Mobilization of history, construction of memory and refoundation of the Kingdom of the Two Sicilies

In July 2017, the Puglia regional council approved the institution of the "Memorial day for the Southern victims of the unification of Italy". The exponents of M5S filed the request in many Southern regions, as Basilicata and Campania. The Valle del Torbido municipality (Reggio Calabria), where is located the Cultural Association of Two Sicilies (Associazione Culturale Due Sicilie), approved the institution of the Memorial day independently from the Five Stars party. The mobilization of the "true history" and the re-construction of the "places of memory" of the pre-unitarian Kingdom of the Two Sicilies constituted the strong symbolic nucleus of the identity politics of the Neo-bourbon movement, since its foundation in Naples in 1993. An élite of enthusiastic activists and sympathizers contributed to create the network of alliances expanding into Southern communities, from South to Northern Italy. For over twenty years, the media represented the Neo-Bourbon as a folk phenomenon, devoid of any political influence. As well as for all social movements, the Internet communication era marks a turning point for Neo-Bourbon activism. The use of the past of the Kingdom of Two Sicilies as the mytho-poietic source of the counter-history of Italian Nation building has been empowered with the performative reiteration in the interconnected environment of new media. The affordances of digital communication give an unprecedented speed to the dissemination of multimodal contents which can be replicated and creatively reappropriated by each social media user.

This contribution draws upon data gathered during the online ethnographic research within Facebook group "No Lombroso" over two years (2012-2014), which is part of a multi-sited fieldwork on the protest movement against the "Cesare Lombroso Museum" in Turin. The mobilization started from the leaders of the Neo-Bourbon movements, in 2009. The Museum craniological collection as evidence of the genocide of Southern peoples is one of the core topic of the Neo-bourbon revisionism that supports the claim for the Memorial day. The online visibility of the Southerner protest movement represented a decisive step for the aggregation of digital "multitudes" including groups and/or individuals – in search of spaces for socialization and recognition.

The participant observation of the daily communication practices within the Facebook group allows me to propose a ritual model of the institution of a new "postmeridionalist" political subjectivity, based on the sharing of the sense of victimhood, on the production of the cultural memory as well as on the mobilizations of history to create the collective sense of belonging to the resurgent past of the Kingdom of Two Sicilies.

8. BIBLIOGRAFIA

- [1] Assmann, A., Conrad, S. 2010, edited by. *Memory in a Global Age. Discourses, Practices and Trajectories*. Basingstoke: Palgrave MacMillan.
- [2] Berger, S. 2007. *Writing National Histories in Europe: Reflections on the Pasts, Presents and the Futures of a Tradition. In Conflicted Memories. Europeanizing Contemporary Histories*, edited by K.H. Jarausch, T. Lindenberger, A. Ramsbrock, 55-68. New York - Oxford: Berghahn Books.
- [3] Bordini, S. 2008. *La storia mediata. Il medioevo visto dal Web: percorsi di ricerca e didattica*. Bologna: CLUEB.
- [4] Casalena, M.P. (a cura di). 2013. *Antirisorgimento. Appropriazioni, critiche, delegittimazioni*, Bologna: Pendragon.
- [5] Couldry, N., Hepp, A. 2016. *The Mediated Construction of Reality*. Cambridge: Polity Press.
- [6] Criscione, A., Noiret, S., Spagnolo, C., Vitali, S. (a cura di). 2004. *La storia a(l) tempo di Internet. Indagine sui siti italiani di storia contemporanea 2001-2003*, Bologna: Pàtron Editore.
- [7] De Lorenzo, R. 2008. *Tempi brevi e lunghi di un regno al femminile: Maria Sofia di Baviera regina del Regno delle Due Sicilie. In Tra res e imago, in memoria di Augusto Placanica*, a cura di M. Mafri e M.R. Pellizzari, 1051-1067. Soveria Mannelli: Rubbettino.
- [8] Dunning, D., Kruger, J. 1999. *Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments*. "Journal of Personality and Social Psychology" (77) 6: 1121-1134.
- [9] Fruci, G.L., Pinto, C. 2017. *Borbonismo e sudismo*. "il Mulino" edizione online, 30 agosto (https://www.rivistailmulino.it/news/newsitem/index/Item/News:NEWS_ITEM:4089)
- [10] Kaschuba, W. 2010. *Iconic Remembering and Religious Icons. Fundamentalist Strategies in European memory Politics? In A European Memory? Contested Histories and Politics of Remembrance*, edited by M. Pakier, B. Stråth, 64-75. New York - Oxford: Berghahn.
- [11] Milicia, M. T. 2014. *Lombroso e il brigante. Storia di un cranio conteso*. Roma: Salerno editrice.

- [12] Milicia, M. T. (2014a). *La protesta “No Lombroso” sul web. Materiali di ricerca sulle narrative identitarie neo-meridionaliste*. “Etnografia e Ricerca Qualitativa” 2: 265–286.
- [13] Milicia, M. T. 2015. *Noi contro tutti. La solidarietà aggressiva nella web communitas No Lombroso*. “EtnoAntropologia” 2: 165–177.
- [14] Milicia, M. T. 2016. *Nel laboratorio sociale dell’odio: un anno di ordinario razzismo su Facebook*. “Voci. Annuale di Scienze Umane” XIII: 124–147.
- [15] Milicia, M. T. 2016a. *Postcolonialismo o postmeridionalismo? Riflessioni sulla teoria postcoloniale a partire dalla ricerca sul campo “Into the heart of Italy”*. “From the European South” 1: 295–306.
- [16] Minuti, R. (a cura di). 2015. *Il web e gli studi storici, guida critica all’uso della rete*. Roma: Carocci.
- [17] Montaldo, S. 2012. *La “fossa comune” del Museo Lombroso e i “lager” di Fenestrelle: il centocinquantesimo dei neoborbonici*. “Passato e presente” 87:105–118.
- [18] Noiret, S. 1999. *Storia e Internet: la ricerca storica all’alba del terzo millennio*. “Memoria e Ricerca” 3: 7-20.
- [19] Nora, P. 2016. *Come si manipola la memoria. Lo storico, il potere, il passato*. Brescia: Editrice La Scuola.
- [20] Ortoleva, P.1999. *La rete e la catena. Mestiere di storico al tempo di Internet*. “Memoria e Ricerca” 3: 31-40.
- [21] Patruno, L. 2011. *Fuoco del Sud. La ribollente galassia dei Movimenti meridionali*. Soveria Mannelli: Rubettino.
- [22] Petacco, A. 1995. *La regina del Sud. Amori e guerre segrete di Maria Sofia di Borbone*, Milano: Mondadori.
- [23] Pinto, C. 2016. *Il patriottismo di guerra napoletano, 1861-1866*. “Nuova rivista storica” (100) 3: 841-869.
- [24] Pinto, C. 2017. *Gli ultimi borbonici. Narrazioni e miti della nazione perduta duo-siciliana (1867-1911)*. “Meridiana” 88: 61-82.
- [25] Rousso H. (2007), *History of Memory, Policies of the Past: What For?*. In *Conflicted Memories. Europeanizing Contemporary Histories*, edited by K.H. Jarausch, T. Lindenberger, A. Ramsbrock, 23-36. New York - Oxford: Berghahn.

Nuove Competenze e Nuovi Servizi per la Ricerca nella European Open Science Cloud: Il Futuro degli Open Research Data nelle Digital Humanities

Giorgio Maria Di Nunzio¹ (coordinatore del panel),
Paolo Manghi², Marisol Occioni³, Francesca Tomasi⁴, Marialaura Vignocchi⁴

¹ Università di Padova - giorgiomaria.dinunzio@unipd.it

² ISTI-CNR Pisa - paolo.manghi@isti.cnr.it

³ Università Ca' Foscari Venezia - occioni@unive.it

⁴ Università di Bologna - [francesca.tomasi, marialaura.vignocchi]@unibo.it

ABSTRACT

A settembre 2017 le Università di Padova, Venezia e Vienna hanno organizzato il seminario “*The European Open Science Cloud (EOSC) versus the single Research Institution. Drawing the scenario at local level*” con l’obiettivo di evidenziare come la realizzazione nel 2018 della *European Open Science Cloud (EOSC)* comporterà dei cambiamenti all’interno delle organizzazioni di ricerca.

In questo panel, focalizzeremo l’attenzione sulla questione dell’Open Science e dei nuovi ruoli e competenze che i ricercatori dovranno avere nel contesto della EOSC e dei nuovi servizi di supporto alla ricerca che gli istituti di ricerca dovranno fornire per poter restare al passo di un cambiamento imminente. Inoltre, discuteremo alcuni punti critici contestualizzandoli nell’ambito delle Digital Humanities e mostrando alcuni risultati di progetti internazionali come EOSCpilot¹ e Digital Research Infrastructure for the Arts and Humanities² (DARIAH) per quel che riguarda lo sviluppo di infrastrutture per la preservazione e la ricerca di dati di collezioni provenienti dall’area di ricerca Umanistica.

PAROLE CHIAVE

European Open Science Cloud, Open Research Data, Data Management Plan, Data Workflow for Digital Humanities

1. INTRODUZIONE

Il tema dell’Open Science (OS) è al centro delle strategie della Commissione Europea volte alla diffusione, alla condivisione della conoscenza e al supporto all’innovazione. A maggio del 2017, il gruppo della *Research, Innovation and Science Policy Experts (RISE)* ha pubblicato un documento riguardante la “Visione Europea” del futuro della ricerca “aperta” che descrive gli elementi fondamentali della ricerca in ambito europeo basata su tre principi: *Open Innovation, Open Science, Open to the World*³. In questo contesto che include anche i progetti Horizon 2020, i finanziatori richiedono che i risultati della ricerca siano “*open access by default with some opt out*” e che gli Open Research Data siano “*as open as possible, as closed as necessary*”⁴.

In questo panel, focalizzeremo l’attenzione sulla questione dell’OS e dei nuovi ruoli e competenze che i ricercatori dovranno avere nel contesto della *European Open Science Cloud (EOSC)* e dei nuovi servizi di supporto alla ricerca che gli istituti di ricerca dovranno fornire per poter restare al passo di un cambiamento imminente. In particolare, nel primo rapporto della commissione *High Level Expert Group (HLEG)* della EOSC⁵ vengono definiti dei punti chiave che definiscono una serie di iniziative per poter permettere la collaborazione tra i vari partner europei in maniera da:

“[...] *enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders. [...]*

based on existing and emerging elements in the Member States, with lightweight international guidance and governance and a large degree of freedom regarding practical implementation.”

La realizzazione della EOSC, pertanto, riguarda due aspetti:

- Fornire dei nuovi servizi di supporto alla ricerca per l’implementazione dell’OS;
- Fornire ai ricercatori una formazione avanzata sulle nuove competenze richieste per gli Open Research Data.

2. DETTAGLIO DELLA PROPOSTA

A settembre 2017 le Università di Padova, Venezia e Vienna hanno organizzato il seminario “*The European Open Science Cloud (EOSC) versus the single Research Institution. Drawing the scenario at local level*” con l’obiettivo di evidenziare come la realizzazione nel 2018 della *European Open Science Cloud* comporterà dei cambiamenti all’interno delle organizzazioni di ricerca.

¹ <https://eoscpilot.eu/>

² <http://www.dariah.eu/>

³ <https://ec.europa.eu/research/openvision/>

⁴ http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

⁵ http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

In questo panel presenteremo le esperienze relative alle questioni della riorganizzazione dei servizi della Biblioteca Digitale di tre Università, Padova⁶, Ca' Foscari Venezia⁷ e Bologna⁸, e di alcuni casi di studio del progetto OpenAIRE 2020 Project⁹ con il CNR di Pisa.

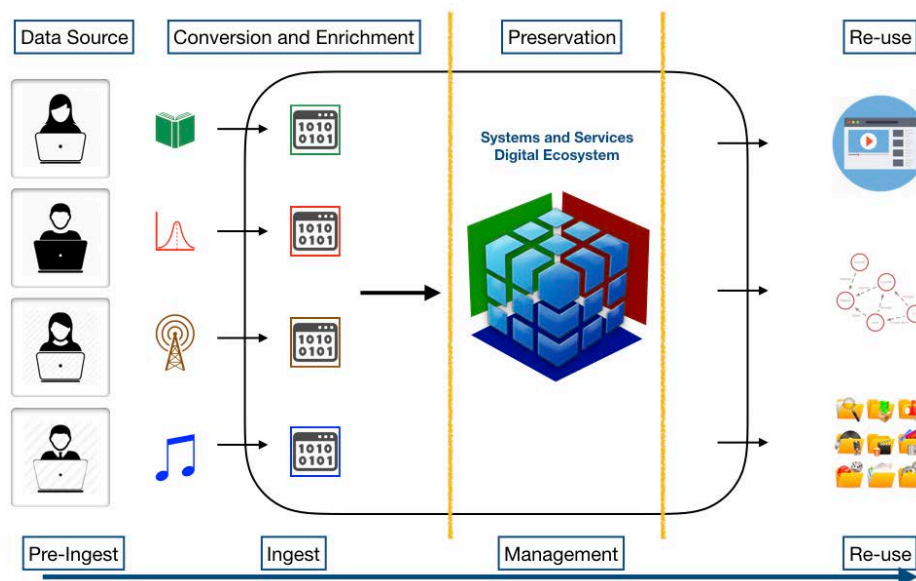


Figura 1 Digital Workflow Mode

In particolare, partendo da un'analisi del *Digital Workflow Model*¹⁰ (mostrato in Figura 1) proposto all'interno del progetto "LEaders Activating Research Networks (LEARN¹¹): Implementing the LERU Research Data Roadmap and Toolkit", discuteremo alcuni punti critici contestualizzandoli nell'ambito delle Digital Humanities e mostrando alcuni risultati di progetti internazionali come EOSCpilot e Digital Research Infrastructure for the Arts and Humanities (DARIAH) per quel che riguarda lo sviluppo di infrastrutture per la preservazione e la ricerca di dati di collezioni provenienti dall'area di ricerca Umanistica. Alcuni dei punti di discussione per questo panel saranno i seguenti:

- la ricerca è sempre più sostenuta da programmi di finanziamento esterni all'organizzazione.
- Il tema dell'Open Science è al centro delle strategie della Commissione Europea volte alla diffusione, alla condivisione della conoscenza e al supporto all'innovazione.
- I finanziatori richiedono che i risultati della ricerca siano *open access by default with some opt out* e che gli Open Research Data siano *as open as possible, as closed as necessary*¹².
- È obbligatorio presentare un Data Management Plan (DMP) entro sei mesi dall'approvazione del progetto.
- È opportuno che un buon progetto di ricerca sia redatto bene e coerentemente con i principi FAIR della EOSC.
- Mancano politiche nazionali legate ai dati della ricerca.
- La gestione dei dati necessita di competenze specializzate e trasversali (data stewardship).
- L'organizzazione deve supportare i ricercatori con servizi che forniscano risposte ai loro bisogni.
- I buoni progetti creano per l'organizzazione una buona reputazione.
- Il fattore tempo: la realizzazione della EOSC diventerà una realtà tangibile nel 2018.

3. INTERVENTI

Il panel avrà quattro interventi da 15 minuti ciascuno più 5 minuti per le domande e/o secondo giro di interventi.

Il chair del panel sarà Giorgio Maria Di Nunzio che farà una breve introduzione sul problema dell'EOSC, contestualizzando il Digital Workflow Model nell'ambito delle Digital Humanities, prima di dare la parola agli invitati.

⁶ <http://bibliotecadigitale.cab.unipd.it/>

⁷ <http://www.unive.it/bda>

⁸ <http://www.sba.unibo.it/it/almadl>

⁹ <https://www.openaire.eu/>

¹⁰ http://learn-rdm.eu/wp-content/uploads/red_LEARN_Elements_of_the_Content_of_a_RDM_Policy.pdf

¹¹ <http://learn-rdm.eu/>

¹² http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

3.1 “Data Monitoring Board” Marisol Occioni, Biblioteca Digitale Università Ca’ Foscari Venezia

Nell’era digitale i dati prodotti da ricercatori che operano in strutture pubbliche e private, valgono oro e crescono via via in modo esponenziale. Il livello di preparazione delle organizzazioni rispetto a questo fenomeno è però disomogeneo ed evidenzia anzi gravi lacune¹³. Nel 2015 l’Università Ca’ Foscari Venezia ha identificato nel Data Monitoring Board (DMB)¹⁴ l’organo competente con funzioni consultive cui i ricercatori si rivolgono per sottoporre i propri DMP nel caso di progetti di ricerca che generano dati sensibili o che aderiscono all’azione pilota “Open Access to Research Data” di Horizon 2020.

Il DMB anticipa in parte quanto previsto da EOSC: un servizio di supporto con funzioni consultive, che nel contesto del seminario già citato potrebbe essere identificato in parte come “*single point of entry*”, composto da esperti nel settore informativo e delle telecomunicazioni, della ricerca e della biblioteca digitale.

Il DMB è nominato con decreto rettorale e dotato di regolamento¹⁵ che gli consentono di esprimere pareri sulla corretta gestione, con strumenti informatici, dei dati raccolti e prodotti dai progetti e sull’accesso aperto per ora nel caso di pilot “open access to research data” relativamente ai seguenti aspetti¹⁶:

- la natura dei dati esistenti o che si intendono creare, provenienza, tipologia, ordine di grandezza, nonché la ragione della creazione di nuovi dataset e relativo valore aggiunto.
- La ragione della creazione di nuovi dati, in termini di contenuto e qualità nonché i metadati e la documentazione accompagnatoria dei dataset, con riferimento a standard di disciplina, interoperabilità e normativa.
- I processi adottati per la gestione, documentazione, cura e conservazione dei dati, stima dei costi associati alla loro gestione e mantenimento.
- Le modalità di raccolta, conservazione, protezione e diffusione dei dati personali e sensibili.
- Le politiche di accesso (aperto o no) ai repository, di trasmissione e circolazione dei dati.
- La corretta ed esaustiva formulazione di Data Management Plan dei progetti che lo prevedono.

Per consentire ai ricercatori di poter lavorare al meglio e rispondere quindi ai requisiti degli enti finanziatori, il DMB ha elaborato alcuni template di Data Management Plan.

3.2 “AlmaDL” Francesca Tomasi, Marialaura Vignocchi, Biblioteca Digitale Università di Bologna

L’Università di Bologna sta investendo sul tema dell’open access, confrontandosi con i problemi legati alla valorizzazione, al deposito e alla condivisione dei dati della ricerca. In particolare il centro AlmaDL, che sovrintende alla Digital Library di Ateneo¹⁷, ha avviato alcune iniziative importanti: la definizione di un repository per la preservazione dei dati della ricerca, compresi i dataset prodotti (AMSActa - Institutional Research Repository) e un servizio sperimentale di supporto per la gestione dei dati della ricerca e l’elaborazione di DMP nell’ambito dei progetti europei H2020 (cfr. per esempio Perceive¹⁸). In particolare poi il nostro gruppo di ricerca si è confrontato con la creazione di LOD nel settore dei beni culturali e si è posto il problema di gestire l’intero processo produttivo di dati aperti e interconnessi, per analizzare, e realizzare, l’intero lifecycle dei dati: dalla produzione (o meglio dalla trasformazione) alla curation, dalla disseminazione all’accesso (cfr. il dataset su Acta¹⁹).

3.3 “OpenAIRE” Paolo Manghi, Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo” – CNR, Pisa

La missione dell’infrastruttura OpenAIRE è quella di promuovere e sostenere l’Open Science in Europa, da un lato disseminandone principi e benefici per le organizzazioni accademiche e i ricercatori, dall’altro offrendo servizi tecnologici a supporto del monitoraggio e della pubblicazione scientifica Open Science e dell’impatto della ricerca in Europa. Nel contesto della European Open Science Cloud (EOSC), OpenAIRE svolge il ruolo di referente per le infrastrutture per la ricerca (RI), fornendo la conoscenza e i servizi necessari a completare il ciclo di vita della ricerca che queste supportano in modo “Open Science”. In altre parole OpenAIRE mira a integrare i servizi utilizzati dagli scienziati per produrre prodotti di ricerca (dati, algoritmi, servizi, workflow) con i servizi per la comunicazione scientifica, al fine di rendere il processo scientifico riutilizzabile, riproducibile e valutabile in modo trasparente. OpenAIRE rende disponibili in modalità as-a-service un pacchetto di servizi che permette agli scienziati di una RI di:

- accedere a strumenti per pubblicare (fornire metadati, ottenere DOI e assicurare la preservazione dei file) tutti i loro prodotti di ricerca, come letteratura, dati, software, pacchetti di ricerca, ecc.;
- specificare relazioni semantiche tra tali prodotti, manualmente o sfruttando tecniche di mining avanzate;

¹³ http://www.leru.org/files/publications/API14_LERU_Roadmap_for_Research_data_final.pdf

¹⁴ <http://www.unive.it/dmb>

¹⁵ http://www.unive.it/pag/fileadmin/user_upload/ateneo/norme_regolamenti/regolamenti/didattica_ricerca/Regolamento_DMB.pdf

¹⁶ Tenuto conto dell’attuale normativa e se necessario del parere preventivo della Commissione Etica di Ateneo.

¹⁷ <http://www.sba.unibo.it/it/almadl>

¹⁸ <http://amsacta.unibo.it/5551/>

¹⁹ <http://amsacta.unibo.it/5497/>

- integrare i servizi delle RI per consentire la pubblicazione automatica di metadati e / o payload di prodotti della ricerca;
- accedere a uno grafo informativo dedicato alla RI costituito da prodotti della ricerca attraverso il quale possono non solo condividere qualsiasi tipo di prodotto con altri scienziati, ma anche facilitandone il riuso, la riproducibilità e sensibilizzando nuovi processi di disseminazione scientifica.

Indice degli autori

Albanesi, Davide.....	144	Clini, Paolo	173
Alexelli, Vaso	179	Colaci, Anna Maria	264
Allegrezza, Stefano	256	Connelly, Erin	154
Alvaro, Corrado	183	Conte, Floriana.....	114
Amato, Giuseppe	139	Costabile, Maria	45
Amici, Simone	183	Cristofaro, Salvatore.....	60
Apicella, Alfredo	99		
Ardito, Carmelo	45		
		Danesi, Valeria.....	183
Bajeot, Jade.....	183	Daquino, Marilena.....	65
Balbi, Barbara	99	Dato, Gaetano.....	170
Baldini, Primo.....	256	Dattilo, David.....	144
Bandirali, Luca	196	De Blasi, Deborah	264
Barbuti, Nicola.....	81	De Calisti, Andrea	91
Bartoli, Margherita.....	109	De Felice, Giuliano.....	260
Basile, Pierpaolo	232	De Felice, Irene	193
Belisario, Ernesto	256	De Luca, Maria Rosa	60
Bellandi, Andrea	144, 246	De Matteis, Fabio	114
Bertolacci, Amos	93	De Ruvo, Francesca.....	99
Biernacka-Licznar, Katarzyna	52	De Troia, Alessandro.....	55
Bino, Giovanna	264	Dei Rossi, Stefano	193
Bolpagni, Marcello	74	Del Grosso, Angelo Mario.....	60
Boschetti, Federico	193, 246	Dell'Orletta, Felice.....	193
Brusa, Antonio.....	260	Desolda, Giuseppe.....	45
Bruzelius, Caroline	135	Di Giorgio, Michele	193
Buono, Paolo	45	Di Nunzio, Giorgio Maria.....	237, 275
Burghart, Marjorie	221	Di Pietro, Chiara.....	105
Burri, Julien	205	Dierna, Giovanni Luca	147
Buzi, Paola.....	16	Dollinar, Michael	144
Byzuk, Joanna	187		
		Eide, Øyvind	210
Cafiero, Florian.....	223	Fagioli, Simone	197
Caivano, Danilo	45	Feliciati, Pierluigi.....	173, 256
Calculli, Lorena Maria	70	Felle, Antonio E.	139
Caldarola, Tommaso	81	Ferilli, Stefano.....	81
Calefati, Christopher	268	Fiorio, Antonella	268
Campagnolo, Alberto	154	Frisone, Domenico Matteo	216
Camps, Jean-Baptiste	223	Frontini, Francesca	246
Caputo, Annalina	232	Fruci, Gian Luca.....	268
Caputo, Annalisa.....	260		
Carbé, Emmanuela.....	256	Galiotou, Eleni	25
Carpanese, Irene.....	89	Gerogiannis, Gian Michele	183
Carrino, Annastella	268	Giovannetti, Emiliano.....	60, 144
Casati, Stefano	91, 133	Giovannetti, Francesca	65
Christen, Alessio	205	Grimm, Ludwig.....	127
Cicola, Ilaria	118		
Čiperová, Olga.....	141	Hall, Crystal	97
Ciula, Arianna.....	210	Hasznos, Andrea.....	227

Henrot, Geneviève	237	Nicolosi, Anika	85
Iodice, Antonio	216	Occioni, Marisol.....	275
Iurlano, Giuliana	264	Orio, Nicola.....	38
Jaouen, Loïc.....	205	Paffrath, Heidi.....	109
Khan, Fahad.....	246	Palmieri, Federico	268
Kiesling, Brady	230	Pansini, Antonella	183
Klimek, Tomáš	141	Papanagiotou, Christos	211
Kozicz, Gerald.....	127	Passaro, Lucia	193
La Marca, Chiara	183	Patruno, Giovanna.....	70
Labanca, Nicola	193	Pecchioli, Alessandra	144
Lacriola, Michele	216	Pellegrino, Bruno	205
Lanzilotti, Rosa.....	45	Petolicchio, Marco.....	74
Lenci, Alessandro	193	Piccinno, Antonio.....	45
Liberotti, Giovanna.....	183	Pinto, Carmine	268
Liestøl, Gunnar	225	Piperno, Clelia.....	144
Liuzzo, Pietro.....	139	Pizzoli, Enrico.....	183
Lo Duca, Angelica	130	Podjed, Dan.....	218
Lucci Baldassari, Guido	89	Pop, Liviu.....	252
Machì, Alberto.....	147	Pozzo, Riccardo.....	18
Maggetti, Daniel	205	Pretnar, Ajda	218, 250
Mamdoh, Ahmed Kamal	227	Psohlavec, Tomáš.....	141
Manghi, Paolo.....	275	Puren, Marie.....	203
Marazzini, Claudio.....	193	Quattrini, Ramona	173
Marchetti, Andrea	130	Rak, Michele	17
Marchi, Simone.....	60	Ralli, Angela	25, 179, 211
Marotta, Daniele	93	Rebora, Simone.....	161
Marras, Cristina	210	Redavid, Domenico	81
Martignano, Chiara	105	Reilly, Michele.....	49
Massimo Salgaro,	161	Ria, Demetrio.....	264
Matera, Maristella.....	45	Ricci, Vito.....	55
Mazzaro, Alessandra Monica	99	Rivoal, Marion	205
Micheletti, Andrea	38	Rizzi, Rosanna	70
Michelone, Francesca.....	190	Roeder, Torsten.....	28
Milicia, Maria Teresa.....	268	Rosselli Del Turco, Roberto	105
Monachini, Monica	85	Rozas Arceo, Alba.....	159
Monella, Paolo.....	21	Rubini, Pietro	216
Montanari, Roberto	99	Rybicki, Jan.....	52
Montemagni, Simonetta	193	Sahle, Patrick	210
Moretti, Giovanni.....	77, 207	Salice, Giampaolo	164
Mortellaro, Isidoro Davide	260	Salvatore, Francesca.....	264
Moscati, Paola.....	41	Salvemini, Biagio.....	70
Moscone, Daniele	183	Scacchi, Alessia.....	168, 176
Mugelli, Gloria	246	Semeraro, Giovanni.....	232
Nava, Beatrice.....	123	Seminara, Graziella	60
Nicoletti, Edoardo	216	Serdar, Ali.....	33
		Sonetti, Silvia.....	268

Spadini, Elena.....	205	Vezzani, Federica.....	237
Spampinato, Daria.....	11, 60	Viazzi, Federica.....	133
Sprugnoli, Rachele.....	77, 207, 242	Vignocchi, Marialaura.....	275
Stefanini, Alberto.....	85	Villani, Claudia.....	260, 268
Štefcová, Petra.....	214	Vitolo, Paola.....	135
Tambassi, Timothy.....	156	Wacha, Heather.....	154
Thompson, Santi.....	49	Weston, Paul Gabriele.....	256
Tomasi, Francesca.....	65, 275	Williams, Joseph.....	135
Tonelli, Sara.....	77, 207	Woźniak, Monika.....	52
Tsimpouris, Charalambos.....	179, 211	 	
Tutumlu, Reyhan.....	33	Žagar, Lan.....	218, 250
Tzanavara, Tonia.....	211	Zenzaro, Simone.....	93
 		Zilio, Daniel.....	38
Valach, Jaroslav.....	214	Zinni, Martina.....	183
Vassallo, Salvatore.....	256	Zoppi, Franco.....	139
Venturi, Giulia.....	193	Zupan, Blaž.....	218