



First steps in building the Eurasian Latin Archive

This is the peer reviewed version of the following article:

Original:

Carbe', E., & Giannelli, N. (2021). First steps in building the Eurasian Latin Archive. QUALITATIVE AND QUANTITATIVE METHODS IN LIBRARIES, 9(4), 578-589.

Availability:

This version is available <http://hdl.handle.net/11365/1123611> since 2021-01-08T11:22:15Z

Terms of use:

Open Access

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license.

For all terms of use and more information see the publisher's website.

(Article begins on next page)

First steps in building the Eurasian Latin Archive

Emmanuela Carbé¹ and Nicola Giannelli²

¹ University of Siena

² University of Siena – QuestIT

Abstract: The aim of the Eurasian Latin Archive (ELA, ela.unisi.it) project is to build an open-access digital platform of Latin and multilingual texts from medieval and early modern times concerning East Asia, which includes a Digital library and tools for textual and semantic analysis. Begun in March 2018 and ended in February 2020, the start-up phase of ELA (DAS-MeMo, www.dasmemo.unisi.it) had been co-financed by Regione Toscana and involves the University of Siena, the IT company QuestIT, specialized in Artificial Intelligence, and the publisher Pacini. This article seeks to explain the workflow of the project, its challenges, the first steps in building the prototype of the platform and adds some notes about further developments.

Keywords: Digital Libraries, TEI, Latin Texts, Multilingual Texts, East Asia, NLP

1. Introduction

Eurasian Latin Archive (ELA) is an ongoing project to build an open-access digital platform of Latin and multilingual texts concerning East Asia from the twentieth to the eighteenth century. It is one of the projects of the Centre for Comparative Studies of the University of Siena, established in 1994 thanks to the medievalist and comparatist I Deug-Su and now under the direction of Francesco Stella. The Centre focuses on Medieval Studies, Korean Studies, International Poetry, and Digital Philology. In addition to ELA, several other projects have been carried out in the field of Digital Humanities since 2007: among them, the Corpus Rhythorum Musicum, an online platform of digital editions of Latin poems from the fourth to the eleventh century set to music, also involving the creation of a printed version with a CD ROM, Stella (2010; 2007a; 2007b); the Archivio della Latinità Italiana del Medioevo (ALIM), and the tool for textual analysis Lexicon.

The theoretical and historical basis of ELA was first presented in 2017 by Francesco Stella at the International Conferences *Medialatinitas*, in Vienna, and *Silk Roads Again II: Eurasian Studies in the Digital Age*, held in the Hanyang University of Seoul.

Started in March 2018 and ended in February 2020, the start-up phase of ELA (DAS-MeMo) was co-financed by the University of Siena (Department of Philology and literary criticism, Department of Information Technology, Engineering and Mathematics), Regione Toscana, the company QuestIT and the publisher Pacini.

The project was presented at the University of Siena during an international meeting in January 2019, *Global Latin. Exploratory workshop*, where a large group of researchers dealt with the topic of Latin Texts of America and East Asia from the Middle Age to the Modern Age.

The DAS-MeMo start-up phase has faced two challenges: building a user-friendly open-access platform, and providing linguistic and semantic tools for the documents. At the beginning of the project the team examined the potential risks using a SWOT Analysis, following among other things the best practices suggested by the King's Digital Lab (Ciula 2019).

2. Background: the ALIM and Lexicon experiences

The Eurasian Latin Archive originates from the Archivio della Latinità Italiana nel Medioevo (ALIM), a project started in the Nineties and aimed at developing a Digital Library, now at its second version. As remarked in Manos (2017), the ALIM project was born to build a database with the aim to integrate the dictionary *Latinitatis Italicae Medii Aevi inde ab a. 476 usque ad a. 1022*, Arnaldi and Smiraglia (2001). ALIM's specific focus is currently to publish Latin texts from the Eighth to the Fifteenth century of Italian Authors. Research units from six Italian Universities contribute to the project, including that of Siena, which currently hosts the platform.

ALIM publishes encoded critical editions as well as born-digital editions. It is possible to search items by using full-text and proximity searches, or sort them by Author, Title, Type of text, Period, Genre, and Collection. The current version of the Digital Library, ALIM2, is managed through Muruca, a framework developed by the company Net7, which also includes some open tools such as Edition Visualization Technology (EVT), a useful suite to manage XML/TEI Editions as explained in Roberto Rosselli del Turco et al. (2017, 2015). Muruca is being used for many projects of Digital Libraries and Digital Editions, for instance Biblioteca Italiana, Burckhardt source, and GramsciProject.

Within ALIM texts are being encoded in XML/TEI, following the guidelines of the TEI Consortium (2019). One of the collections included in ALIM is devoted to documents of Italian authors who wrote about East Asia: Iohannes de Plano Carpini, Odoricus de Portu Naonis, Iohannes de Marignolli, Peregrinus de Castellus, and Andreas de Perugia. This section acted as the starting point for the Eurasian Latin Archive.

The ALIM Digital Library also uses the Lexicon system, developed in 2014 at the University of Siena by Francesco Stella and Luigi Tessarolo, Stella (2018: 54-57). It lemmatizes texts and allows a linguistic analysis of corpora (concordances, collocations, frequencies and linguistic comparisons of texts).

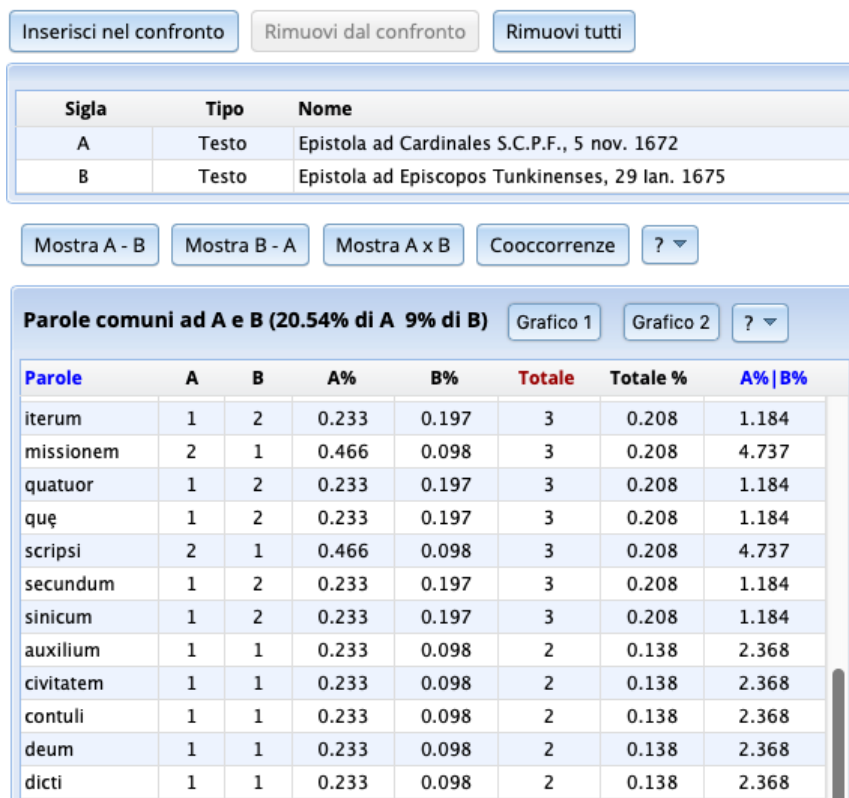


Fig. 1: Comparison of two texts with Lexicon

When ELA started, one of its initial purposes was to be as interoperating as possible with the ALIM Platform, especially as far as the TEI encoding is concerned. Both the design and the architecture of the new project and its workflow have been realised on these grounds.

For this reason, a comparative analysis has been made between the use of metadata and TEI encoding in ALIM and other projects, such as the Corpus Corporum, a database of Latin texts directed by Philip Roelli and developed at the University of Zurich. The Corpus Corporum database, despite being TEI-based, has undergone different choices both in encoding and in metadata management. This is also due to the different mission of the project: while ALIM relies on a manual encoding of the single documents, by providing *editio*

princeps among other outputs, the Corpus Corporum processes large amounts of digitized texts and also provides, in some cases, raw OCR elaboration.

3. The census

Prior to the design of the platform, a census of the documents was performed for a better understanding of how to manage the data. Helpful resources can be found on the web for the purpose of delving into the relationships between Europe and East, as noticed by Morris (2015) on the online magazine “Digital Orientalist”. For ELA the most important resource has been Sinica 2.0, a project of the University of Vienna led by Monika and Georg Lehner. The aim of Sinica 2.0 is to collect data about books published between 1477 and 1939 freely available in more than sixty Digital Libraries. It originated from the bibliography, compiled from 2003 onwards, “Western Books on China in Libraries in Vienna/Austria, 1477-1939”, and is currently a blog based on the WordPress CMS. Every item is associated with bibliographical references and a useful list of Digital Libraries and repositories where the resources are available, including their links and in most cases also the dates when they were digitized, and the provenance of the original resources.

Another remarkable project is the Christian Chinese Texts Database (CCT), of the University of Leuven, based on the open-source software CollectiveAccess which focuses on the creation of catalogues. The CCT-Database, funded by Ad Dudink and Nicolas Standaert, provides the census of sources about Christianity in China, differentiating primary and secondary sources, and includes manuscripts, books and maps. As Sinica 2.0, the CCT-Database makes available links to each Digital Library hosting for item. The metadata of the primary sources include useful references to secondary sources and translations in other languages.

These and other bibliographical repertories allowed the ELA census to be performed. The goal in this phase was to identify as many Latin or multilingual documents as possible. The process was managed through OpenRefine, an open-source tool developed by Google in order to manage, clean, organize and enrich data, with the possibility of connecting the data, as explained in Holland, Verborgh, De Wilde (2013), also by using resources from the web, Williamsom (2017). Each item includes a set of metadata based on Dublin Core elements and is enriched with references to WikiData, Vial, Worldcat, Sinica 2.0, CCT-Database, and a link to the DLs if the resource has already been digitized. The census has been updated continuously by using information related to the ELA processes. Such details encompass the story of the processing of the texts inside ELA: the responsible of the digitation, who is identified with his or her Orcid, as well as the responsible of the OCR processing and the checking of the OCR. The OCR software used, the log of the change history, the evaluation of the source, the dates of the whole process, which type of Licence is used. Documents and its metadata in ELA are published under the Creative Commons BY-NC-SA 4.0 license.

Items in ELA have been encoded in XML/TEI; therefore, the census is not only useful to have a general picture of the ongoing work, but also constitutes the basis to provide a suitable TEI Header for each document.

4. The ELA Framework

The ELA Digital Library has been designed with a requirement analysis and a MoSCoW analysis, so that each phase could be prioritized suitably. Prior to the decision to develop the platform, the possibility to use the Omeka platform – and in particular its semantic edition – had been taken into account, since it had been used successfully in other Digital Libraries, such as Fonte Gaia and Leblanc (2019). However, it appeared that Omeka could not be used for this specific project, because the list of ELA's requirement exceeded its basic capabilities and the team should have spent a considerable amount of time in the development of several plugins to integrate its missing features.

The framework of ELA depends on the type of source: in the case of documents such as ALIM and Corpus Corporum, the codified texts were already available. In other cases, it was necessary to check the copyright licence, and agreements had to be set with Institutions. If texts can be digitized, ELA then prepares transcriptions, or applies OCR and verifies manually the correctness of the process. As already said, items are encoded in XML/TEI. Dates, persons, and places are encoded by using Wikidata, Viaf, and online Gazetteer, or by using Pleiades project as first choice, in order to give consistency to the data, and the geographical database GeoNames.

The Digital Library system was developed in Java EE and relies on an ElasticSearch for its search capabilities. The architecture is based on modules, which allow future development of the project and guarantee the scalability of the Digital Library, Xie and Matukiak (2016: 330).

In terms of users, the prototype tentatively manages only Digital Library System Administrator, Content Creator and Content Users, with a minimal front-end for the team to upload, modify and manage documents. However, with an eye to future implementations, the platform contemplates different user roles, in relation to different possibilities and permissions. In the current ELA Platform, Content Users are able to discover and acquire items in different formats, and visualize and analyze them. In the future, the Authenticated User should also be made able to create his/her personal shelf with sub-corpora. The Content Creators will be provided with different access levels (Head of the research units, Reviewers, Editors), to give the opportunity to create, submit, annotate, validate and update items and collections.

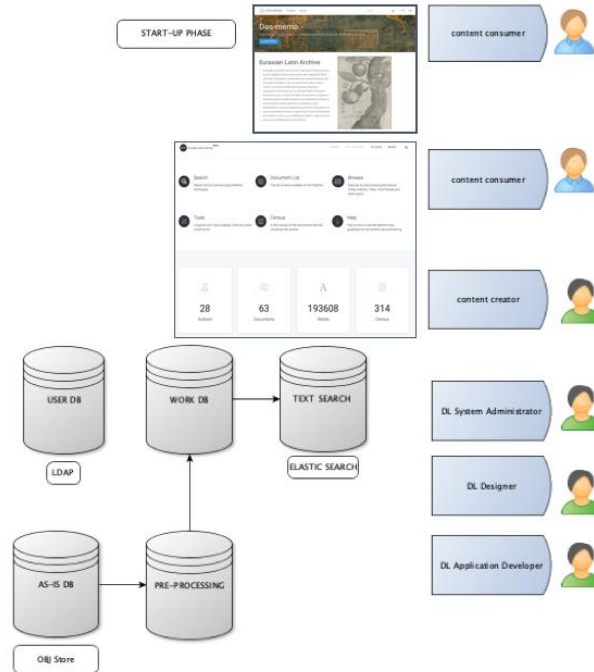


Fig. 2 The ELA architecture

5. The prototype

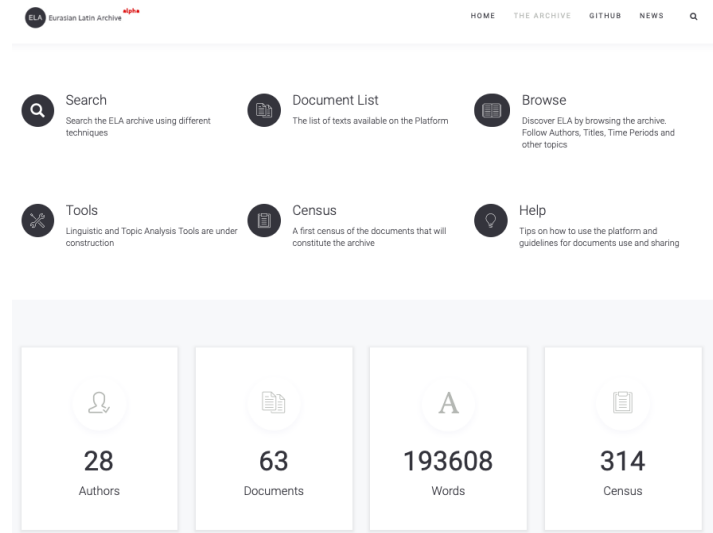


Fig. 3 Home Page of the ELA Digital Library

The prototype of the platform consists of a website embedding the Digital Library. As previously chosen for ALIM, the site of the Eurasian Latin Archive was developed in WordPress: this makes the management of dynamic site pages, such as the blog and the news, easier for the research units. The census is also available, rendered by a specific plugin.

Census

Show entries Search:

ID	Author	VIAF	WIKI ID	Author's dates	Co-Author	Title
1	Giovanni, da Pian del Carpine, Archbishop of Antivari, -1252	http://www.viaf.org/86860272	https://www.wikidata.org/wiki/Q315149	1190-1252		Historia Mongalorum
2	Odorico, da Pordenone, 1265?-1331	http://www.viaf.org/106967420	https://www.wikidata.org/wiki/Q379733	1265?-12331		Relatio
3	Johannes de Marignollis gest. 1358/59	http://www.viaf.org/362145856886222920387	https://www.wikidata.org/wiki/Q1362183	1290-1358		Relatio
4	Giovanni da Montecorvino 1247-1328 b	http://www.viaf.org/88989381	https://www.wikidata.org/wiki/Q372219	1247-1328		Epistula Secunda

Fig. 4 The census

The Digital Library is integrated with customs Javascript and JQuery. It can be explored in three different ways: through a document list, a search panel and a browse panel.

In the search panel, queries can be made on a full-text basis by using the Lucene Query Syntax and/or by short lemmas. Specific queries are available through search filters based on the following tags: Author, Date, Title of the document, and Places, Persons or Dates mentioned within the text; when filters are used, the interface suggests possible completions of partially entered words. If words are searched for, these will be highlighted within the texts, and the number of occurrences will be shown.

Text

Lorem Ipsum Dolor Sit Amet

Lemmas

x frater

Filters

Has

Place name

VALUE Kathay

Has

Document Date

VALUE 1307/1400

Search

Author	Title	Date	
Iohannes de Monte Corvino	Epistola (2)	1305	
Iohannes de Monte Corvino	Epistola (3)	1306	

Showing 1 to 2 of 2 entries

Previous **1** Next

Fig. 5 The Search Panel

The browse panel allows for the search of text through lists. Documents can be browsed by Author, Languages, Geographic references, Time period of the text, Source (i.e. primary source, digital edition), Type of document, (i.e. prose), Genre (i.e. *epistolae*, *itineraria*, *adnotationes*), as well as Places and People mentioned within the texts.











Browse by		Author	Title	Date	
Genre		Bonaventura Ibañez	Brevis Relatio	1688	 
Adnotationes		Gulielmus de Rubruquis	Itinerarium	XIII	 
Catalogi		Iohannes de Marignollis	Relatio	1355-1359	 
Catechismus		Iohannes de Plano Carpini	Historia Mongalorum	1250	 
Confucius		Julianus OFM	Epistula de vita Tartarorum	1237	 
Epistolae		Showing 1 to 5 of 9 entries			
Informationes		Previous 1 2 Next			
Itineraria et Relationes					
Passio					

Fig. 6 The Browse panel

The document viewer – an aspect which still needs improvement –, shows a heading with the Title, the Author (linked to a Vial and/or WikiData), and the edition of the texts when available (linked to WorldCat). An “info panel” displays document-related metadata, including usage rights. It is possible to disable the page/folio break indications by using the menu, in order to read the text without interruption. Underlined text in the document indicates the presence of locations, names and dates. A pop-up window shows the types of tags (Place/Date/Person) with a normalized form. Texts can be downloaded in TXT, PDF and XML TEI formats.

Info

Publisher: ELA

Date: 1356

Language: Latin

License: CC BY-NC-SA 4.0

Epistola

Andreas de Perusio 12...-1332

In: Sittica Franciscana, I: Itinera et relationes fratrum Minorum saeculi XIII et XIV, a cura di A. Van Den Wyngaert, Firenze, Quaracchi, 1929, 373-377

page: 373

Fr. Andreas de Perusio de Ordine Minorum fratrum divina permissione vocatus Episcopus, reverendo patri fratri Guardiano Perusini Conventus salutem et pacem in Domino sempiternam. Et infra sequitur.

1. Nara propter immensam terrarum mariumque distantiam, inter me et vos interiectam, vix sperare possum, quod littere ad vos per me transmissae ad manus vestras valeant pervenire. Et infra subsequitur. Noveritis me itaque cum bone memorie fratre Peregrino Coegiscopo, et mee peregrinationis individuo comite, post multos labores et langores, inedia, variaque incomoda, atque pericula in terra pariter et in mari, ubi faimus rebus omnibus et etiam tunicis et habitibus spo

page: 374

Place: Kumbaly

liati, demum, Deo iuvante, ad Cambalensem civitatem, que sedes est imperii magni Chanis anno dominice incarnationis millesimo CCC. XXIII, ut credo, pervenisse; ubi secundum mandatum a Sede Apostolica nobis datum, Archiepiscopo consecrato, moram ibi per quinquaginta ferme contraximus. Infra quod temporis spatium procuravimus a lafa ab imperatore magifico pro victu et vestitu octo personarum. Est autem alpha impense, quas Imperator tribuit munitis magnatorum, oratoribus, bellatoribus et diversarum artium artificibus et loculatoribus, pauperibus et diversis diversarum conditionum personis; que impense plurimum latinorum Regum introitus expensasque transcendunt.

Menu

Hide page breaks

Search results

Tags

Background:

Default

Download TXT

Download XML

Download PDF

Fig. 7 The document viewer

6. ELA Tools

The ELA Team has completed the development of the toolset available on GitHub also in form of Jupyter notebooks. The framework, as shown in Carbé and Giannelli (2019), extracts information from XML documents and performs operations such as automatic TEI tagging, clean-up, normalization of documents and text analysis. The main tool, devoted to extracting linguistic and semantic data by using Natural Language Processing methods, is based on CLTK – The Classical Language Toolkit, funded in 2014 by Kyle P. Johnson (it is realized for historical languages and also includes Chinese); 2. NLTK – The Natural Language Toolkit, developed by Steven Bird and Edward Loped at the University of Pennsylvania, used in ELA for language-agnostic operations such as calculation of frequencies, concordances, collocations, and N-grams. This tool was integrated into the platform to perform statistics (such as number of words, and Latin words, if the document is multilingual; number of Latin lemmas and types, maximum/minimum average of Latin types, Type/Token Ratio), lists of lemmas, type and stop words, with its frequencies; N-Grams; collocations. Additionally, the tool extracts the list of persons and places found in the texts, associating the latitude and longitude based on Pleiades and GeoNames database. The interface of the platform also has maps for each documents, to show the geolocalization of place names on maps using the JavaScript library Leaflet.

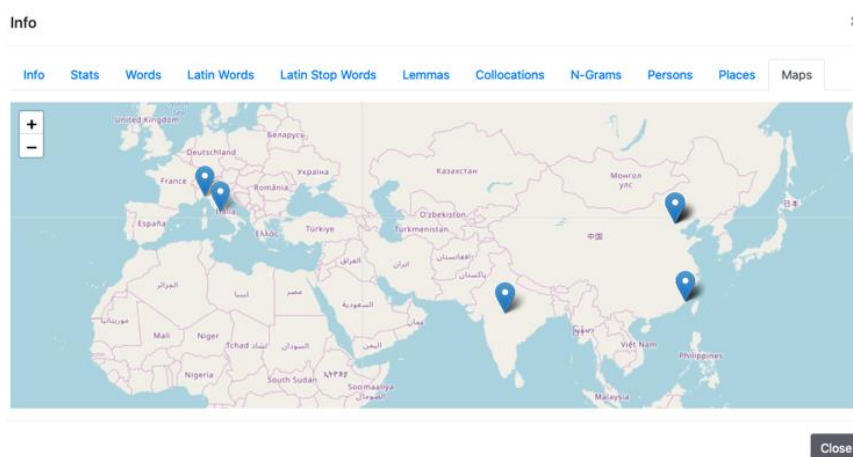


Fig. 8 Map of PlaceName mentioned in the document

The team started to analyze the entire corpus by developing the tool GlobalStats, available on a Jupyter notebook of the ELA GitHub and based on matplotlib widget, with the intention of performing some word clouds for the statistics of frequencies and some experiments of network analysis. This first network is based on authors and people mentioned in the corpus. People are nodes, and edges connect authors to the people mentioned.

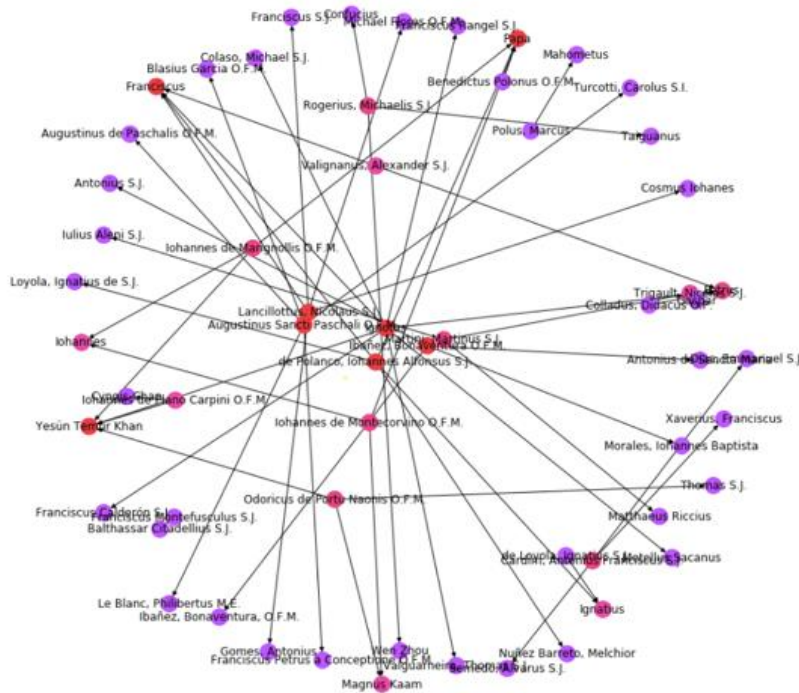


Fig. 9 First experiments of network analysis on the ELA corpus

7. Conclusions and next steps

In the two-year start-up phase of the Eurasian Latin Archive projects, the team built the prototype of a digital library of Latin and multilingual documents concerning East Asia, and tools for linguistic and semantic data. The team has faced many challenges since the very beginning, from the constitution of the corpus to the data model and the managing of a large amount of different type of texts.

The project has the unique goal to manage a wide multilingual corpus which mostly comprises texts in Latin from the twentieth to the eighteenth century. If, on the one hand, this means a great opportunity to understand better the use of the Latin language through time and space, on the other hand, difficulties in the development of a NLP framework, starting from the lemmatization issues, are evident.

For the purpose of creating a digital platform, it is essential to interoperate with other projects and respond to the necessity of sustainability of the project itself. In Latin related DH analyses many different methods and approaches can be found, as confirmed by “LiLa: Linking Latin”, a project established in 2018 to connect and harmonize linguistic resources, Passarotti et al. (2019).

The next steps of the project concern the improvement of the platform and ELA Tools: qualitative analysis of the results is being carried out, especially to establish a comparison with Lexicon's tools and other lemmatizers, Passarotti and Mambrini (2019), Eger, Brück and Mehler (2015). Another goal is to improve the automatic system of the Named-Entity Recognition and increase the datasets of Places and Names. An increment of the corpus, now comprising 70 documents, is regularly in progress.

Acknowledgements

“GiovaniSi” (www.giovanisi.it), a project of Regione Toscana based on POR-FSE 2014-2020, co-financed a Research Fellowship for the start-up phase of the Eurasian Latin Archive, from March 2018 to February 2020.

References

- Archivio della Latinità Italiana nel Medioevo. Access date 16.06.2020. Available at: <http://en.alim.unisi.it>
- Arnaldi, F. and Smiraglia, P., ed. (2001). *Latinitas Italicae Medii Aevi lexicon (saec. V ex – saec. XI in.). Editio altera addendis quae confecerunt L. Celentano, A. De Prisco, A.V. Nazzaro, I. Polara, P. Smiraglia, M. Turriani*, Firenze: Edizioni del Galluzzo. ISBN: 88-8450-020-6.
- Biblioteca italiana. Access date 16.06.2020. Available at: <http://www.bibliotecaitaliana.it>
- Bibliotheca Sinica 2.0. Access date 16.06.2020. Available at: <https://www.univie.ac.at/Geschichte/China-Bibliographie/blog/>
- Carbé, E. and Giannelli, N. (2019). A Digital Platform for the “Latin Silk Road”: Issues and Perspectives in Building a Multilingual Corpus for Textual Analysis. Abstract for the Digital Humanities Conference 2019, Utrecht, 9-12 July. Access date 16.06.2020. Available at: <https://dev.clariah.nl/files/dh2019/boa/0947.html>
- CCT – Chinese Christian Texts Database. Access date 16.06.2020. Available at: <https://www.arts.kuleuven.be/sinologie/english/cct>
- Center for Comparative Studies I Deug-Su. Access date 16.06.2020. Available at: <http://www.centroideugsu.unisi.it/en>
- Ciula, A. (2019), KDL Checklist for Digital Outputs Assessment (version 2.0). Access date 16.06.2020. Available at: [10.5281/zenodo.3361580](https://zenodo.org/record/3361580)
- CLTK – The Classical Language Toolkit. Access date 16.06.2020. Available at: <http://cltk.org>
- CollectiveAccess Software. Access date 16.06.2020. Available at: <https://www.collectiveaccess.org>
- Corimu – Corpus Rhythmorum Musicum. Access date 16.06.2020. Available at: <http://www.corimu.unisi.it>
- Corpus Corporum: repositorium operum Latinorum apud universitatem Turicensem. Access date 16.06.2020. Available at: <http://www.mlat.uzh.ch/>
- DAS-MeMo – Data-mining e analisi statistica su fonti testuali storiche del periodo medievale e moderno. Access date 16.06.2020. Available at: <http://www.dasmemo.unisi.it>
- Eurasian Latin Archive (ELA). Access date 16.06.2020. Available at: www.ela.unisi.it
- Eurasian Latin Archive's Tools. Access date 16.06.2020. Available at: <https://github.com/EurasianLatinArchive/ELA>
- Eger, S., Brück, Tim vor der, and Mehler, A. (2015), Lexicon-assisted tagging and lemmatization in Latin: A comparison of six taggers and two lemmatization methods. In

- Zervanou, K., Erp, M. van, Alex, B. (eds.) *Proceedings of the 9th SIGNUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, Beijing, China, July 30, 2015, pp. 105-113. Access date 16.06.2020. Available at: <https://www.aclweb.org/anthology/papers/W/W15/W15-3716/>
- Edition Visualization Technology (EVT). Access date 16.06.2020. Available at: <http://evt.labcd.unipi.it/> and <https://github.com/evt-project/evt-viewer/>
- GeoNames. Access date 30.02.2020. Available at: <https://www.geonames.org>
- Gonçalves, M. A. (2008). Towards a digital library theory: A Formal digital library ontology. In *International Journal on Digital Libraries*, 8(2), pp. 91-114. DOI: 10.1007/s00799-008-0033-1
- Gramsciproject. Access date 16.06.2020. Available at: <http://quaderni.gramsciproject.org>
- Hooland, S. van, Verborgh, R. and De Wilde, M. (2013). Cleaning Data with Open Refine. *The Programming Historian*, 2. Access date 16.06.2020. Available at: <https://programminghistorian.org/en/lessons/cleaning-data-with-openrefine>
- Leaflet. Access date 16.06.2020. Available at: <https://leafletjs.com>
- Leblanc, E. (2019). Which Services for User Participation? Representing Cooperation and Collaboration in a Participative Digital Library. *Poster Session of the Digital Humanities Conference, Utrecht, July 2019*. Access date 16.06.2020. Available at <https://dev.clariah.nl/files/dh2019/boa/0243.html>
- Lexicon. Access date 16.06.2020. Available at: www.lexicon.unisi.it
- Manos, T. (2018). ALIM: Archivio della Latinità Italiana del Medioevo, *Digital Medievalist*, 11(1): 4. DOI: 10.16995/dm.79
- Morris, J.H. (2015). Beyond “Laures Kirishitan Bunko”: Digital Repositories for Studying 16th and 17th Century Japanese Christianity. In *Digital Orientalist*. Access date 16.06.2020. Available at: <https://digitalorientalist.com/2018/12/18/beyond-laures-kirishitan-bunko-three-digital-repositories-for-studying-16th-and-17th-century-japanese-christianity/>
- Muruca Framework. Access date 16.06.2020. Available at: <http://www.muruca.org>
- NLTK – The Natural Language Toolkit. Access date 16.06.2020. Available at: <http://nltk.org>
- OpenRefine. Access date 25.07.2019. Available at: <http://openrefine.org>
- Passarotti, M. et al. (2019). Lila: Linking Latin – A Knowledge Base of Linguistic Resources at NLP Tools. In T. Declerck, T. and McCrae, J. P. (eds), *Proceedings of the Poster Session of the 2nd Conference on Language, Data and Knowledge (LKD-PS 2019)*, Leipzig, May 2019, pp. 20-23. Available at: <http://ceur-ws.org/Vol-2402/paper2.pdf>
- Passarotti, M. and Mambrini, F. (2019). Harmonizing Different Lemmatization Strategies for Building a Knowledge Base of Linguistic Resources for Latin. In Friedrich, A. and Zeyrek, D. (eds), *Proceedings of the 13th Linguistic Annotation Workshop (LAW XII)*, Association for Computational Linguistics, Florence, 2019, pp. 71-80. ISBN: 978-1-950737-28-2. Available at: <https://sigann.github.io/LAW-XIII-2019/pdf/W19-4009.pdf>
- Pleiades Gazetteer. Access date 16.06.2020. Available at: <https://pleiades.stoa.org>
- Rosselli del Turco, R. et. al. (2019). Visualisation with EVT: Simplicity is Complex. *Poster session of the Digital Humanities Conference 2019, Utrecht*, July 2019. Access date 16.06.2020 available at <https://dev.clariah.nl/files/dh2019/boa/0416.html>
- 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*, 8: 1-21. DOI: 10.400/jtei.1077.
- Stella, F. (2018). *Testi letterari e analisi digitale*. Roma: Carocci, 2019.

- Stella, F. (2017). From Medieval Latin to Global Latin: The Eurasian Digital Archive. *Silk Roads Again II: Eurasian Studies in the Digital Age, December 8-9, 2017* (pp. 245-269). Seoul: Hanyang University, Asia-Pacific Research Center.
- Stella, F. (2013). Encoding Issues in Philological critical editions. *Horizons. Seoul Journal of Humanities*, 4 (2): 239-254.
- Stella, F. (2010). The Versification Database of the Digital “Corpus Rhythmorum Musicum”. In Selaf, L., Noel Aziz Hanna, P. and van Driel, J. (eds.), *Formes strophiques simples / Simple Strophic Patterns* (pp. 17-36). Budapest: Akademiai Kiadó.
- Stella, F., dir. (2007a). *Corpus Rhythmorum Musicum Saec. IV-IX. Songs in non-Liturgical Sources / Canti di tradizione non liturgica*. Firenze: Sismel. ISBN: 978-88-8450-251-3.
- Stella, F. (2007b). Digital philology, medieval texts, and the Corpus of Latin Rhythms, a digital edition of music and poems. In Ciula, A. and Stella, F. (eds.), *Digital Philology and Medieval Texts* (pp. 223-249). Pisa: Pacini.
- TEI Consortium (2019). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.5.0. Last updated on 29th January 2019. Access date 16.06.2020 available at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- Williamson, E. P. (2017). Fetching and Parsing Data from the Web with Open Refine. In *The Programming Historian*, 6. Access date 16.06.2020. Available at: <https://programminghistorian.org/en/lessons/fetch-and-parse-data-with-openrefine>
- Xie, I. and Matukia, K. K. (2016). *Discover Digital Library. Theory and Practice*. Amsterdam: Elsevier.