Europeana. A digital library for the Humanities?

Prof. Dr. Stefan Gradmann
Humboldt-Universität zu Berlin / School of Library and Information Science
stefan.gradmann@ibi.hu-berlin.de
Overview

- Europeana background: i2010 and the Google challenge
- Putting semantics on the agenda: EC working group on DL interoperability recommendations ...
- ... as taken up in Europeana functional specifications and architecture (D2.5)
- Where do we go from here? Proposals, plans & prospects
  - EuropeanaConnect, Europeana1.0 - Access ...
- Semantics: why, for whom?
  - Digital Humanities Scholars
- Strategic value of Europeana semantic foundations
- Europeana: a digital library for the digital humanities?
How it all started …

2005: Excitement about Google Books (Jeanneney: “Google defie l'Europe” => Chirac, Schröder)

=> EC i2010 agenda with Digital Libraries as one of 3 'flagship initiatives': the setting up of the European Digital Library as a “common multilingual access point to Europe’s distributed digital cultural heritage including all types of cultural heritage institutions”, announced by Commissioner Reding in September 2005

- **2008**: at least 2 million digital objects; multilingual; searchable and usable; work towards including archives.
- **2010**: at least 6 million digital objects; including also museums and private initiatives.

DL Interoperability WG active from January to June 2007 with as main mission to

- Contribute to the short term DL agenda => identify areas for short term action and recommend elements of an action plan (**list of prioritised feasible options**).
DL Interoperability Working Group Composition

- Emmanuelle Bermes (Bibliothèque nationale de France / F)
- Mathieu Le Brun (Centre Virtuel de la Connaissance sur l’Europe / LU)
- Sally Chambers (The European Library Office / TEL)
- Robina Clayphan (The British Library / GB)
- Birte Christensen-Dalsgaard (State and University Library Aarhus / DK)
- David Dawson (The Museums, Libraries and Archives Council / GB)
- Stefan Gradmann (Hamburg University Computing Center / D, Moderation)
- Stefanos Kollias (Technical University of Athens / GR)
- Maria Luisa Sanchez (Ministerio de Cultura / ES)
- Guus Schreiber (Vrije Universiteit Amsterdam / NL)
- Olivier de Solan (Direction des Archives de France / F)
- Theo van Veen (Koninklijke Bibliotheek / NL)
- EC: Pat Manson Chair, Marius Snyders (European Commission, DG INFSO, Cultural Heritage and Technology Enhanced Learning)
- Federico Milani (European Commission, DG INFSO, eContentPlus)
The Interoperability Abstraction Layer Cake

Abstract

- **functional / pragmatic**: based on a common set of functional primitives or on a common set of service definitions
- **semantic**: allowing to access similar classes of objects and services across multiple sites, with multilinguality of content as one specific aspect
- **syntactic**: allowing the interchange of metadata and protocol elements
- **technical/basic**: common tools, interfaces and infrastructure providing uniformity for navigation and access

Concrete

Europeana. A Digital Library for the Humanities?
Towards a Semantic Agenda for Europeana

- DL interoperability group identified semantically enabled functionality as the critical, distinguishing feature of a European Digital Library.

- The final report states:

  “Semantic web technologies can be used to create semantic interoperability in three areas:
  - interoperability of the federated content resources on concept level
  - semantic interoperability of EUDL on user interface level
  - semantic interoperability of EUDL for automated processing
  both for EUDL being plugged in emergent semantically aware WWW services and for integrating such services in the functional scope of EUDL.”
(8) Basic Semantic Interoperability
Make existing metadata and the controlled terminology used therein machine understandable to create a data layer ready for semantic query methods. The method of choice for conversion is SKOS, but use of OWL or RDF may be appropriate in some application scenarios.

(9) Awareness Building regarding Semantic Interoperability
Demonstrate the added value to be gained from semantic interoperability and the short term viability of converting existing controlled terminology in experimentation environments relevant to the EuDL. These environments also to be used to market semantic interoperability functions of EuDL as our unique selling point.

... as taken up in Europeana

Draft functional specification (D2.5) statements:

- **A semantic interface for Europeana**
  “A central principle for building Europeana is that a network of semantic resources will be used as the primary level of user interaction.”

- **Interaction with data providers**
  “Aggregators and other content providers need to provide identifiers, metadata files, **vocabularies in SKOS form, links to semantic nodes**, licensing and rights information and access to the original digital objects.”

- **Terminology mapping**
  “The work to turn this [collection of Europeana terminologies] in a 'European Ontology' and more specifically the **mapping of these concept schemes** cannot be done in the context of Europeana alone but must be made be part of the wider EC research agenda. However, Europeana will have to contain instruments that can be used to produce such mappings and to promote best practices.” → Europena.Connect WP1

- **Make Europeana**
  a network of inter-operating object surrogates enabling semantics based object discovery and use.
Europeana is not
Yet Another Digital Library
Contextualisation of Europeana
Surrogate Aggregations

Europeana. A Digital Library for the Humanities?
... and further developed in projects planned and proposed for EC funding

- **Europeana V1.0 (eContent+ TN, implementation)**
  - Providing Europeana infrastructure components
  - Building parts of the core functionality of Europeana
  - Building the Europeana organisation

- **EuropeanaConnect (eContent+ BPN, implementation)**
  - Integrate work from different (mostly EU funded) development activities to provide additional core and advanced functionality
  - WP1 being about “Creating the Europeana Semantic Layer” including ontology matching, ontology mapping and mapping evolution

- **ASSETS (FP7 IP proposal, Research)**
  - Multimedia modeling and retrieval
  - Generalised surrogate building machine
  - Surrogate long term preservation
  - Not to mention Europeana Local, Athena, European Film Gateway, Apenet & all the rest ...

Europeana. A Digital Library for the Humanities?
Europeana 1.0 Big Picture

Data Suppliers
- Aggregators: national or per domain
- Consortial Providers
- Individual Contributors

Europeana
- OAI PMH + I/O API
- Surrogate Database
- User Functionality / API
- Europeana Portal
- Any other Portal

Projects & Proposals
- Europeana V1.0 (org, data, tech)
- EDLnet (org, data, tech)
- EDLlocal (data)
- EuScreen (data)
- Europeana Connect (tech)
- ROSE (tech, research)

User

Europeana. A Digital Library for the Humanities?
Semantics: for whom?

- Citizens

- Machines
  - Make European cultural heritage part of future global semantic processing networks

- Digital Humanities Scholars
  - Humanities scholars always have been concerned with reaggregation and interpretation of cultural heritage corpora (literature, music, artwork – all kinds of cultural artefacts)
  - Europeana enabling automated semantic operations over large cultural heritage corpora creates entirely new opportunities for the digital humanities!
Scholarship in the digital era
Processing of source data in the Humanities: *aggregation* ...
... modeling ...

Scholarly Publications

Communication & Coordination (groupware, collaboratories)

Interpretation

Modeling

Transformation & statistics, automated, unambiguous
- statistical analysis
- parser

Digital corpora / sources

Referential Data (Lexicon type)

Referential Structures (Grammar Type)

Formal declarative, interactive
- tools for semantic analysis
- meta-languages (e.g. TEI)

Europeana. A Digital Library for the Humanities?
... and Digital Heuristics?
Digital Document Value Add-On
(c) J.-C. Meister

Complexity

Semantic Web Linking
Hyperlinks
Semantic MarkUp

Collocation
Variant Comparison
Semantic Profiling (Z-Score)

Concordance
Search & Retrieval
Pattern Recognition

Parsing, Character Substitution

Hermeneutical Richness

Europeana. A Digital Library for the Humanities?
Open Scholarly Communities on the Web partly (c) Paolo d'Iorio / Michele Barbera

- **Objective**
  - Create social networks of specialists in a humanities research area
  - Enable 'web scholarship'
  - Similar to academies in the 17\textsuperscript{th} century

- **Barriers**
  - Scholars lack digital literacy
  - Public institutions tax digital access to public domain
  - Publishers protect old business models
  - Lack of content, lack of open sources
Would you go online more often if:

- There were more primary resources / digital facsimiles of texts relevant to your area of interest or research:
  - Yes: 74%
  - No: 14%
  - I don't know: 9%
  - No Response: 3%

- More of your colleagues' papers were available online:
  - Yes: 65%
  - No: 23%
  - I don't know: 10%
  - No Response: 2%

- There was an online community devoted to your area of interest/field of research where you could engage with colleagues:
  - Yes: 53%
  - No: 26%
  - I don't know: 17%
  - No Response: 4%
Some Work Done / Ongoing

- **HyperNietzsche**

- COST A32: “Open Scholarly Communities on the Web” (13 countries)
  - April 2006 - December 2010
  - to establish and foster the growth of Scholarly Communities on the Web
  - to create a digital infrastructure for the humanities
  - to define an appropriate legal, economic and social framework

- eContent+ project: “Discovery. Digital semantic corpora for virtual research in philosophy” (6 partners)

- Key partners in all scenarios are **CNRS ITEM** (Paolo d'Iorio) and **NetSeven** (Michele Barbera)

- I am leading A 32 WG 2 (Technology) together with Michele Barbera
Components Created to Work on this Corpus

- **Hyper**
  - “... a web application created to help scholars in accessing primary sources (like manuscripts) and to share the result of their researches.”
  - pre-Discovery (PHP, PostgreSQL)

- **Talia**
  - “... a semantic web digital library system, designed to help philosophy researchers and scholars in their work with digital content and provide them with all the resources needed for their work.”
  - Discovery (Ruby) => http://www.talia.discovery-project.eu/

- **Philospace**
  - “... is an innovative client application that allows users to browse philosophical content, published in the Talia platform, leveraging the power of semantic knowledge associated to such contents.”
  - Discovery (Dbin 2.0, => http://dbin.org)
## Discovery Corpus: Digitised Manuscripts

<table>
<thead>
<tr>
<th>Philosophy</th>
<th>Literature</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedrich Nietzsche</td>
<td>Gustav Flaubert</td>
<td>Fernand Braudel</td>
</tr>
<tr>
<td>CNRS-ITEM, Paris</td>
<td>CNRS-ITEM, Paris</td>
<td>MSH-Paris</td>
</tr>
<tr>
<td>Ancient / Modern Philosophy</td>
<td>Marcel Proust</td>
<td></td>
</tr>
<tr>
<td>CNR-ILIESI, Roma</td>
<td>CNRS-ITEM, Paris</td>
<td></td>
</tr>
<tr>
<td>Arthur Schopenhauer</td>
<td>Paul Valéry</td>
<td></td>
</tr>
<tr>
<td>Universities Pisa / Mainz</td>
<td>CNRS-ITEM, Paris</td>
<td></td>
</tr>
<tr>
<td>Contemporary philosophy</td>
<td>Virginia Woolf</td>
<td></td>
</tr>
<tr>
<td>RaiNet, Roma</td>
<td>Leicester / London</td>
<td></td>
</tr>
<tr>
<td>Ludwig Wittgenstein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAB, Bergen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“To work towards making all good things part of the common good and all things free to those who are free”
Hyper: Digitisation, Presentation (1)
Europeana. A Digital Library for the Humanities?
Hyper: Transcription, Presentation (1)

Europeana. A Digital Library for the Humanities?
Hyper: Transcription, Presentation (2)

Europeana. A Digital Library for the Humanities?
Hyper: Sources and Editions (synoptical)

Europeana. A Digital Library for the Humanities?
Europeana. A Digital Library for the Humanities?
Talia: Refacturing Hyper using Semantic Web (1)
Manuscript Stemmatia: A Specific Kind of Inferencing (1)
Manuscript Stemmata: A Specific Kind of Inferencing (2)
Talia/PhiloSpace: Annotations and Semantic Enrichment

Philoso-space and Talia(1)

- Philospace can be loaded with one or more Domain Ontologies.
- Content providers use Philospace to add semantic annotations to the contents (already published in Talia).
  - E.g., “An essay contradicts/cites an other one”, “A sentence of an essay (fragment) defines the concept of Nature”, etc.
- End-users (e.g. scholars) browse Talia contents and, contextually, semantic annotations related to them.

Europeana. A Digital Library for the Humanities?
Semantic enrichment in practice

Semantic Enrichment Example (1)
Semantic enrichment in practice

Stable URL:
http://piccini-source/sources/Bio/xpointer(id('123'))
Semantic enrichment in practice

Example:

Stable URL:
http://piccini-source/sources/Bio/xpointer(id('123'))

http://piccini-source.com/sources/Bio/xpointer(id('123'))

"This is partially true because...."
Semantic enrichment in practice

Example (4)

Stable URL
http://piccini-source/sources/Bio/xpointer(id('123'))
Semantic enrichment in practice

Example:

Stable URL
http://piccini-source/sources/Bio/xpointer(id('123'))

Puccini Source

Example:

"This is partially true because..."

http://puccini-source.com/sources/Bio

part of

http://puccini-source.com/sources/Bio

comment

RDF/XML
Hyper/Talia: Bidirectional Links

Bidirectional Links

Essay A quotes B

Quoted Essays

Essay B is quoted by A

Quoting Essays
Hyper/Talia: Dynamic Contextualisation (1)

Primary Sources

Dynamic Contextualisation

Quoted Essays

Essay

Quoting Essays

Europeana. A Digital Library for the Humanities?
Dynamic Contextualisation

Quoting Contributions

- Editions
- Translations
- Essays
- Paths
- Multimedia

Aphorism
... a nice basis for reasoning / digital heuristics!
Conclusion (1):
It's the semantics, stupid!

- Converging agendas
  - Platforms such as Talia and communities built around those need to be pluggable into Europeana
  - Such platforms and communities must be capable to display Europeana resources as part of their proper functional context

- For the first two interoperability requirements to be viable at all a strong semantic data layer as part of Europeana and an appropriate API for specialised reasoning as a basis for digital humanities heuristics are vitally required

- It remains an open issue to what extent we need to support advanced reasoning features that may not be available on SKOS / RDFS level: do we need to invest in OWL ontologies – and if so: using which resources?
Conclusion (2): It's the semantics, stupid!

- We thus **do not propose to fight Google** (which would be a silly idea anyway) – but rather to provide an **added value** Google doesn't offer (but may of course be working on, in secret ...)

- In such a perspective, a strong profile characteristic for Europeana would – paradoxically – result from what may be perceived as a specific European weakness: the **scattered, heterogeneous and multilingual nature** of our cultural resources requiring semantic foundations for conceptual interoperability!

- A semantics aware Europeana delivers what has been asked for in the “Cyberinfrastructure for the Social Sciences and Humanities” report commissioned by the ACLS: a resource we could call “**Our Cultural Commonwealth**”
Conclusions (3):
It's the humanities, stupid!

- Interoperability of Technology and Scholarship:
  - “The often institutionally enforced disconnect between technician and scholar is a microcosm of the much larger failure to consider the ends to which our powerful machine is put. [...] Unfortunately we are still in some instances thinking that the non-technical scholar specifies the end in mind, whereupon the technician implements it. In that circumstance both lose.” (Willard McCarty, Humanist Discussion Group 22.053)

- Europeana needs to
  - build a strong semantic data layer at the foundation of its large digital corpora
  - provide users with clearly defined APIs to support specialised functions for reasoning and digital heuristics
  - work with the humanities computing community

- In that circumstance both win!

And to finally answer the initial question: Europeana will be a prime resource for the Digital Humanities – but much more than a Digital Library. We thus shouldn't use this term.
Conclusions (4): from 'Connecting' to 'Thinking'

Co-operation of a semantically based Europeana and of Digital Humanities communities enables an interesting transition from

Thank you for your patience and attention!