Linked Open Europeana: Semantics for the Citizen

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

Re: David: From 'oral' to 'written' (Plato) – from 'written' to 'digital' (David) – changing cultural techniques are threatening

- Linked Open Data: What is it, how does it work?
- How does it relate to the Semantic Web?
- Linked Open Europeana: how will it work in Europeana
- ... for the citizen: scholars, pupils and teachers, tourists, politicians, entrepreneurs
- Conclusion: on the importance of being 'open'
Apologies first: I am aware of speaking in a country that named Tim Berner-Lee for its first Millennium Technology Prize as early as 2004.

Delivering this talk in such a place is courageous ...

It may be equivalent of
  - ... selling snow to eskimos
  - ... carrying coals to Newcastle
  - ... bringing beer to Munich

And I've done it before: some of you may remember me delivering a talk on Semantic Web issues here in Helsinki back in 2004!

→ Thank you for the invitation, be humble, and quote Tim Berners-Lee excessively

Tarja Halonen & TBL
The Web of Documents

Information Management: A Proposal (TBL, 1989)
Resources and Links in the Document Web

- We have HTTP URIs to identify resources and links between them – but we are missing a few things!
- What kinds of resources are 'Louvre.html' and 'LaJoconde.jpg'?
  - A machine cannot tell.
  - Humans can: we recognize implied context!
- How exactly do they relate to each other?
  - A machine cannot tell.
  - Humans can: again we recognize implied context!
Syntactically Extending the Document Web (1)

- We add a syntax for making statements on resources: RDF

- Or, more generally triples ...

- ... where S and P are web resources (identified using URIs) and O is either a web resource or a literal
Syntactically Extending the Document Web (2)

- We add a schema language (RDFS) with elements such as:
  - classes,
  - hierarchies of classes and properties,
  - inheritance
  - support for basic inferencing.

- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:

![Diagram showing class relationships: dog, cow, animal, ant. Each isA relationship connects them to the animal class. Dog and cow are also connected to the animal class. Ant is connected to the animal class through the class ant.](diagram.png)
The Web of Things ... Somewhat Mistaken

What's wrong with this picture?

Taken from Ronald Carpentier's Blog at http://carpentier.wordpress.com/2007/08/08/1-2-3/
... and the Way we extend the Web in scope to make it a 'Web of Things'
And we get … Linked Data

**Linked Data essentials**

1. Use URIs
2. Use HTTP URIs
3. Serve useful information using SPARQL, RDF standards
4. Mention URIs of related objects

Copyright © 2008 W3C (MIT, ERCIM, Keio)

http://www.w3.org/2008/Talks/0617-lod-tbl/#(4)
LoD … and the Semantic Web?

- “Semantic Web done right” (TBL, http://www.w3.org/2008/Talks/0617-lod-tbl/#(3))
- → What was wrong about the Semantic Web in 2007?
  - Artificial Intelligence heritage (agents, heavy logic)
  - Mostly corporate, inhouse applications
  - Little visibility on the WWW (“Where's the Web in the SW?” Frank van Harmelen, 2006)
  - Misuse of the attribute 'semantic'
- „I called this graph the Semantic Web, but maybe it should have been Giant Global Graph!“ (TBL, http://dig.csail.mit.edu/breadcrumbs/node/215)
- => Linked Open Data extends the Web of documents in syntax and scope without falling back into the mistakes of Artificial Intelligence. Future extensions may well grow into a truly 'semantic' web ... (≠Web 3.0)
The Europeana Data Model:
Making Europeana Part of Linked Open Data

Partially based on Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Herbert Van de Sompel: The Europeana Data Model (IFLA 2010)
Pre-EDM

This made V. Reding promise a „European Digital Library“ in 2005
“Europeana Semantic Elements” (ESE)
- Created for 2008 version of Europeana
- enforces interoperability by converting datasets to a Dublin-Core like “flat” representation
- “simple and robust” but:
  - original metadata is not visible anymore
  - no specializations to finer-grained models
  - no connections to external (open data) resources
- Probably shouldn't have been called “semantic” :-)

"Europeana Semantic Elements" (ESE)
- Created for 2008 version of Europeana
- enforces interoperability by converting datasets to a Dublin-Core like “flat” representation
- “simple and robust” but:
  - original metadata is not visible anymore
  - no specializations to finer-grained models
  - no connections to external (open data) resources
- Probably shouldn't have been called “semantic” :-)

“Europeana Data Model” (EDM)
- destined to replace ESE with the 2011 release of Europeana
- ESE “application profile” of EDM (backwards compatibility)
- preserves original data while still allowing for interoperability
- allows for Semantic Web representation
EDM and other standards

- **Simple Knowledge Organization System (SKOS)**
  - Models the KOSs in the Semantic Data Layer of Europeana.
  - Allows for matching between KOSs.

- **DCMI Metadata Terms**
  - Used for a core of semantically interoperable properties for descriptive metadata about an object.
  - Ensures backwards compatibility to ESE.

- **Open Archives Initiative Object Reuse & Exchange (OAI ORE)**
  - Organizes the metadata about an object in Europeana:
    - Provided Object: Represents the described object of interest.
    - Digital Representation: Some digital view of the object.
    - Proxy: The description of the provided object from one given perspective.
    - Aggregation: Groups all information pieces together.
Mona Lisa: French Ministry of Culture

<table>
<thead>
<tr>
<th>Réponse n° 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Domine peinture</td>
<td></td>
</tr>
<tr>
<td>Type d’objet tableau</td>
<td></td>
</tr>
<tr>
<td>Titre PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE</td>
<td></td>
</tr>
<tr>
<td>Auteur/exécutant LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)</td>
<td></td>
</tr>
<tr>
<td>Précision auteur/exécutant Vinci, 1452 ; Amboise, 1519</td>
<td></td>
</tr>
<tr>
<td>École Italie</td>
<td></td>
</tr>
<tr>
<td>Période création/exécution 1er quart 16e siècle</td>
<td></td>
</tr>
<tr>
<td>Millénaire création/exécution 1503 entre ; 1506 et</td>
<td></td>
</tr>
<tr>
<td>Genèse ouvre en rapport ; reproduit en gravure</td>
<td></td>
</tr>
<tr>
<td>Historique commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 ; nombreuses copies dont une conservée au Louvre ; gravé par Fauchery, par Filhol, par Landon</td>
<td></td>
</tr>
<tr>
<td>Matériaux/techniques peinture à l’huile ; bois</td>
<td></td>
</tr>
<tr>
<td>Mesures 77 H ; 53 L</td>
<td></td>
</tr>
<tr>
<td>Sujet représenté portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assis, accoudé, loggia, italien) ; fond de paysage (montagne, rocher, cours d’eau, pont, plaine, route)</td>
<td></td>
</tr>
<tr>
<td>Date sujet représenté 1479-1528</td>
<td></td>
</tr>
<tr>
<td>Lieu de conservation Paris ; musée du Louvre département des Peintures</td>
<td></td>
</tr>
<tr>
<td>Statut juridique propriété de l’État ; musée du Louvre département des Peintures</td>
<td></td>
</tr>
<tr>
<td>Anciennes appartenances François 1er ; Couronne de France</td>
<td></td>
</tr>
<tr>
<td>Numéro d’inventaire INV 779</td>
<td></td>
</tr>
<tr>
<td>Commentaires légère diminution du tableau sur les côtés (environ 7 mm) ; acheté vraisemblablement vers 1519, après la mort de l’artiste</td>
<td></td>
</tr>
<tr>
<td>Bibliographie HEYDENRICH 6 ; OTTINO DELLA CHIESA 31 ; VILLOT 1 484 ; HAUTECOEUR 1601 ; C.S.I. 1981, P 192</td>
<td></td>
</tr>
<tr>
<td>Copyright notice © Musée du Louvre, © Direction des Musées de France, 1999</td>
<td></td>
</tr>
<tr>
<td>Crédits photographiques © Réunion des musées nationaux ; © Hervé Lewandowski ; © Thierry Le Mage</td>
<td></td>
</tr>
<tr>
<td>commande reproduction et/ou conditions d’utilisation</td>
<td></td>
</tr>
<tr>
<td>renseignements sur le musée</td>
<td></td>
</tr>
<tr>
<td>000PE025604</td>
<td></td>
</tr>
</tbody>
</table>
Metadata Record in EDM

Linked Open Europeana: Semantics for the Citizen. Prof. Dr. Stefan Gradmann
Different Semantic Grains

- Keep data expressed as close as possible to original model.
- Using mappings to more interoperable level: the EDM.

"PORTRAIT DE MONA LISA (1497-1528) ; DITE LA JOCONDE"@fr

"LEONARDO DI SER Piero DA VINCI ; VINCI Léonard de (dit)"@fr

"1503, entre"@fr

"1506, et"@fr

"bois"@fr

"François Ier"@fr
Semantic Enrichment

g:Agent: persons or organizations

g:Place: spatial entities

"Leonardo da Vinci"@en

"Léonard de Vinci"@fr

g:TimeSpan: time periods or dates

"PORTRAIT DE MONA LISA (1497-1528); DITE LA JOCONDE"@fr

"woman"@en

"bois"@fr "wood"@en

"humain"@fr "human being"@en

\textit{Linked Open Europeana: Semantics for the Citizen. Prof. Dr. Stefan Gradmann}


\textit{20}
Event-Centric Modeling

Preserving and exploiting original data also means being compatible with descriptions beyond simple object level.
Complex Objects

- Part-whole links for complex (hierarchical) objects
- Order among parts of objects
- Derivation and versioning relations
Current State of EDM

- Confirmed feasibility in community workshops (archives, libraries, audiovisual archives, museums).
- The EDM is now closely articulated with the 'Danube' requirement process (Europeana release 2011).
- We're in the course of prototyping on a larger scale.
- EDM Specifications and Primer:
  - http://version1.europeana.eu/web/europeana-project/technicaldocuments/
- EuropeanaLabs:
  - http://europeanalabs.eu/
… and LoD
An Aggregation ...
... some context
... more context
... and the Big Picture:
The Semantic Data Layer

Semantic Network

Networked Object Representations
The Semantic Data Layer

Bridging „isles of information“ by connecting objects from different domains via cross-vocabulary links.

Networked Object Representations

library

archive

museum
EDM and Linked Open Data

Context Data
- DBpedia
- PND and SWD (prototype)
- Geonames
- LCSH
- …

Semantic Network

Networked Object Representations

Europeana Information Space

Linked Open Europeana Semantics for the Citizen, Prof. Stefan Gradmann
'Beyond Catalogues and Records' generates new questions!

- Where do resource aggregations 'start'? Where do they 'end'?
- And what constitutes document boundaries??
- And which node was connected to which one at a given time???
... and new opportunities: Triple Sets and Reasoning (1)
Triple Sets and Reasoning (2)

http://ex.org/Painting

rdf:type

http://ex.org/LaJoconde

rdfs:subClassOf

http://ex.org/ArtisticWork
Triple Sets and Reasoning (3)

→ Potential of novel digital heuristics!
An Example (1)
An Example (2)

This is a research prototype of Europeana's semantic search engine. Enter a search term, for example: Egypt, Rembrandt, window.

Collections Thesauri

powered by CitoPatria 1.0 beta 2 (18/12/2008) SWI Prolog 5.5.6-g507ab80 eculture.cs.vu.nl

OpenOffice.org
An Example (3)
An Example (4)
An Example (5)
An Example (6)
An Example (7)

Adam & Eve

Linked Open Europeana: Semantics for the Citizen. Prof. Dr. Stefan Gradmann
The Datacloud Behind the Example
Semantic Exploration, Context Discovery and Knowledge Generation: Semantics for the Citizen
What if ...

- Europeana EDM is fully implemented and data migrated.
- Data originating from public bodies (PSI) are available as LoD
  - As in http://data-gov.tw.rpi.edu/wiki:

And we combine all this with the rest of the Cloud: what use to make of all this?
SwickyNotes: Ontology Based Annotation as Linked Open Data

COST A32 SwickyNotes Workshop in Munich: DEMO PAGE

Finally, you can try to annotate one of the Digital Libraries in the Philosource network, a federation of Talia based digital libraries built within the Discovery EU project.

http://dbin.org/swickynotes/murucaska/start.htm
“Cretans are always Liars”

... annotated
Perseus Digital Library
Gregory R. Crane, Editor-in-Chief
Tufts University

Browse the Collections

Greek and Roman Materials
Primary and secondary sources for the study of ancient Greece and Rome

Art & Archaeology Artifact Browser
Look through a massive library of art objects, sites, and buildings. The library's catalogs document 1305 coins, 1909 vases, 2003 sculptures, 179 sites, 140 gems, and 424 buildings. Each catalog entry has a description of the object and its context; most have images. Descriptions and images have been produced in collaboration with many museums, institutions, and scholars. Catalog information and keywords have been taken from standard sources, which are cited in the entries for each object.

Arabic Materials
Arabic language documents

Germanic Materials
Materials for the study of the Germanic Peoples

19th-Century American
Sources on the history of the 19th-century United States.

Renaissance Materials
Primary and secondary sources in early modern English literature

Richmond Times Dispatch
Issues of the Richmond Times Dispatch.

Word Counts by Text Collection

- Classics (60,278,054 words)
- Arabic (1,066,623 words)
- Germanic (355,696 words)
- 19c. Am. (58,312,291 words)
- Renaissance (5,940,158 words)
- Rich. Times (19,413,298 words)
- Papyri (4,737,070 words)

View a map of the most frequently mentioned places in the Perseus Digital Library.

Or by collection:
- Greek and Roman Materials
- 19th-Century American
- Renaissance Materials
- Richmond Times Dispatch

Linked Open Europeana: Semantics for the Citizen. Prof. Dr. Stefan Gradmann
“Cretans are always Liars”
… in Perseus!
→ Lidell-Scott … and further!

- → Isidore
- → Europeana
- → National Digital Library of Finland
- → Wordnet, OpenCalais, Geonames …: into context!!!
Tourists and the like


- Berlin: I'd like to spend a day in places connected to Alexander von Humboldt → Europeana LoD + http://www.fahrinfo-berlin.de/Fahrinfo/bin/ (LoD) + http://dbpedia.org/page/Alexander_von_Humboldt

- Berlin: Take me to places that were important in the context of the public uprise 17th of June 1953 (by bus, please. And mind all the streets that have been renamed after 1990!) → Europeana + http://www.fahrinfo-berlin.de/Fahrinfo/bin/ (LoD) + http://www.17juni53.de/karte/berlin.html (LoD)

- Italian perspective: show me everything the French have taken away from us (not just La Joconde) → Europeana LoD (provenance info!)
Teachers and pupils

- Find surrealist paintings from France, Spain and Italy – is there a significant difference between the artworks produced in those countries? → *Europeana Portal (country facet)*
- Document the post-war history of your hometown → *relevant local PSI resources + Europeana LoD*
- Find jazz music by bands from London under CC SA → *(BBC Music, MusicBrainz + Europeana LoD)*

Business people

- Is there a significant market for access to audiovisual objects of art in France? → *Europeana LoD + PSI from INA*
Politicians

- To what extent did funding from eContent+ stimulate content digitisation activities over the past 10 years in EU member states? → *Europeana Portal (provenance information)*

- Does it actually save money after all to close down the Historic Museum in Hamburg Altona? → *Europeana (Logfiles) + relevant local PSI resources (relating cultural resources to financial data)*

- What is the Finnish contribution to Europeana – and what degree of attention did this contribution generate for our country? → *Europeana Portal + Logfiles*
... and the political bit
On the Importance of being 'Open' (1)

- "Openness (allowing access) is separate question." (TBL, http://www.w3.org/2008/Talks/0617-lod-tbl/#(22))
- Does Linked Data work **without being 'open'**?
  - Technically speaking: yes (cf. pharma industry or biomedical data)
  - But it gets horribly expensive that way ...
  - ... much too expensive, probably, for Europeana to afford!
  - And much of its 'semantic' charms would be lost in such a setting, anyway.
This has a number of **implications**
- No control over data usage
- No income to be generated from data access and use
- Innovative and (commercially) attractive services can be built on LoD

→ Do not repeat mistakes we are very familiar with from the Open Access debates of the past 10 years!
  - 'open' vs. 'free',
  - 'freen' vs. 'commercial'

→ Do not exclude commercial reuse for Europeana metadata!

→ What is the actual **value** of context (in business terms)?
Selected Reading


- A lot more mercifully skipped ...
Questions?