Linked Open Europeana: Semantics for the Digital Humanities

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

- 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 Digital Humanities: scholars using Europeana
- Conclusion: on the importance of being 'open'
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:
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-iod-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)
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 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.
Réponse n° 1

Domaine peinture
Type d'objet tableau

Titre PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE
Auteur/executeur LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)

Précision auteur/executeur Vinci, 1452 ; Amboise, 1519
École Italie
Période création/exécution 1er quart 16e siècle
Millénaire création/exécution 1503 entre ; 1505 et

Généalogie œuvre en rapport ; reproduit en gravure
Historique commandé par le fiorentin Francesco de' Giocondo, époux de Mona Lisa entre 1503 et 1505 ; nombreuses copies dont une conservée au Louvre ; gravé par Fauchery, par Filhol, par Landon

Matériaux/techniques peinture à l'huile ; bois
Mesures 77 H ; 53 L

Sujet représenté portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assis, accoudé, loggia, Italien) ; fond de paysage (montagna, rocher, cours d'eau, pont, plaine, route)

Date sujet représenté 1479-1528

Lieu de conservation Paris ; musée du Louvre département des Peintures

Musée de France au sens de la loi n°2002-5 du 4 janvier 2002
Statut juridique propriété de l'État ; musée du Louvre département des Peintures
Anciennes appartenances François Ier ; Couronne de France
Numéro d'inventaire INV 779

Commentaires légère diminution du tableau sur les côtés (environ 7 mm) ; acheté vraisemblablement vers 1519, après la mort de l'artiste

Bibliographie HEYDENRICH 6 ; OTTINO DELLA CHIESA 31 ; VILLOT I 484 ; HAUTECOEUR 1601 ; C.S.I. 1881, P 192

Copyright notice © Musée du Louvre, © Direction des Musées de France, 1999
Crédits photographiques © Réunion des musées nationaux ; © Hervé Lewandowski ; © Thierry Le Mage

commande reproduction et/ou conditions d'utilisation
renseignements sur le musée
000PE025604
Metadata Record in EDM

Aggregation

Digital Representations

Proxy

Object of Interest
Different Semantic Grains

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

**ens:Agent**: persons or organizations

**ens:Place**: spatial entities

**skos:Concept**: entities from KOS

**ens:TimeSpan**: time periods or dates
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:
- EuropeanaLabs:
  - [http://europeanalabs.eu/](http://europeanalabs.eu/)
… and LoD
An Aggregation ...
... some context
... more context
... and the Big Picture:
The Semantic Data Layer

Semantic Network

Networked Object Representations

Linked Open Europeana: Semantics for the Digital Humanities.
Prof. Dr. Stefan Gradmann. THATCamp, Firenze, 24 March 2011
The Semantic Data Layer

Bridging „isles of information“ by connecting objects from different domains via cross-vocabulary links.
EDM and Linked Open Data

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

Europeana Information Space

Linked Open Europeana: Semantics for the Digital Humanities

Prof. Dr. Stefan Gradmann. THATCamp, Firenze, 24 March 2011
'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

http://ex.org/LaJoconde

http://ex.org/ArtisticWork

rdf:type

rdfs:subClassOf
Triple Sets and Reasoning (3)

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

This is a research prototype of Europeana's semantic search engine.

Enter a search term, for example: Egypt, Rembrandt, window.

search

Collections

Search

Europeana

46,036 artworks

Rijksstudio

82,781 artworks

Louvre

11,327 artworks

RDF

Visualised by EOPEI 1.0 beta 2 (18/12/2013)

EOPEI 0.7.3-36-gc5730b0

codi.de/2004
An Example (2)
An Example (3)
An Example (4)
An Example (5)
An Example (6)
An Example (7)

Adam & Eve

Linked Open Europeana: Semantics for the Digital Humanities.
Prof. Dr. Stefan Gradmann. THATCamp, Firenze, 24 March 2011
The Datacloud Behind the Example
Semantic Exploration, Context Discovery and Knowledge Generation: Semantics for Digital Humanities
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
“Cretans are always Liars”
... annotated
Perseus Digital Library

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
600 (28,614 words)
Arabic
1 (256,621 words)
Germanic
955 (68,686 words)
19c. Am.
38 (31,321 words)
Renaissance
6,940 (159 words)
Rich. Times
194 (1,237 words)
Papyrus
4 (4,376 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
“Cretans are always Liars” … in Perseus!
→ Lidell-Scott … and further!

→ Isidore
→ Europeana
→ National Digital Library of Finland
→ Wordnet, OpenCalais, Geonames ...: into context!!!
... 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.
On the Importance of being 'Open' (2)

- 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?