



The Manuscript as Cultural Heritage: Digitisation ++

A Digital Humanities Point of View Not so much a conservation point of view slides on http://www.slideshare.net/gradmans

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





Overview



- Manuscript digitization ...
- ... and after: turning the 'object' into 'heritage' ...
- ... to enable scholarly work: modelling and (re-)aggregation
- Good Practice and Better Practice: Examples
 - COST A32, Discovery, SemLib, Shared Canvas
- How to use digitised manuscripts beyond emulation?
 - Scholarly Primitives
 - Contextualisation
- Towards semantic interoperability
 - RDF, LoD and the EDM!
 - DM2E





Perspective



Digitization and Beyond



- I will not talk about manuscript digitisation ...
- ... but about what it takes to turn digitised manuscripts into cultural heritage
- What is 'cultural heritage' (as opposed to 'cultural objects' and 'cultural witnesses')?
 - Medieval manuscripts are cultural heritage.
 - Digitised manuscripts are not: they are a means to preserve and/or to represent this cultural heritage!
 - Or might the latter not be true: do digitised manuscripts have a potential of becoming 'cultural heritage' in their own right?
- Heritage is related to hermeneutical appropriation and thus to digital humanities scholarship



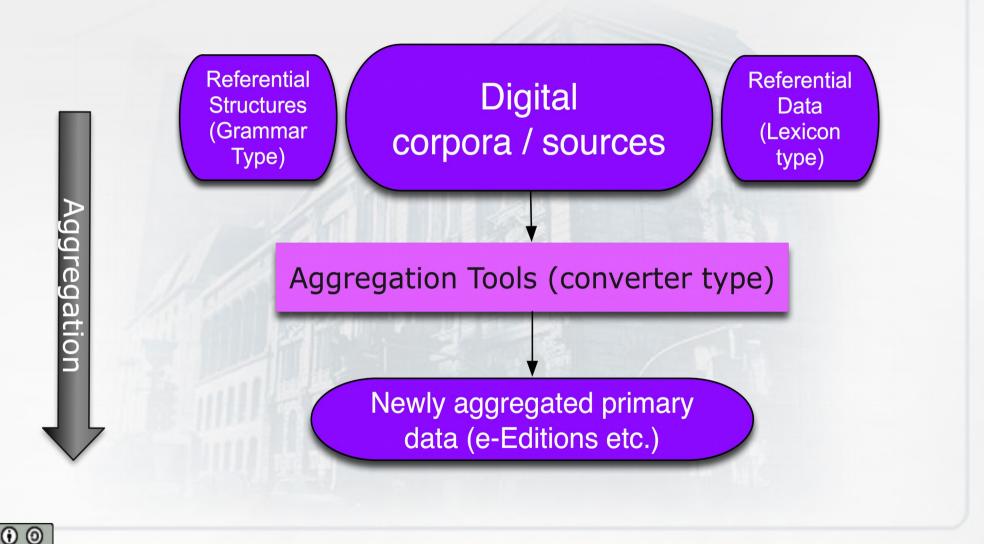


... into scholarly work



Processing of source data in the Humanities: aggregation ...

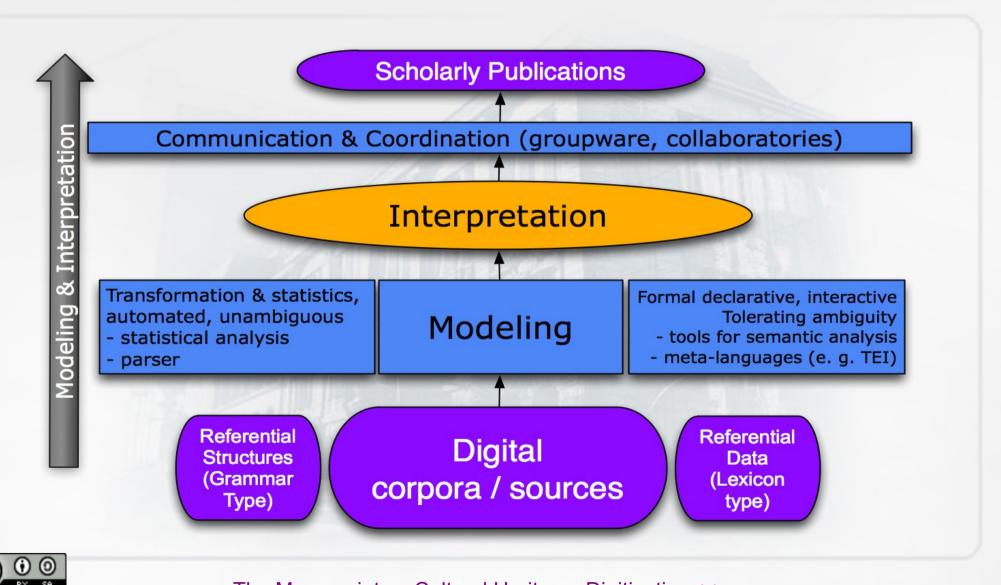






... modeling ...







... and Digital Heuristics?







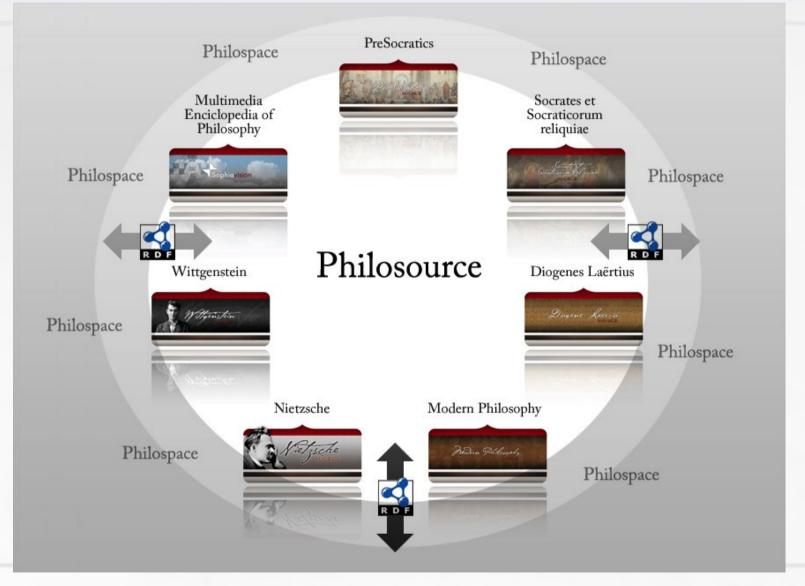


Good Practices and Better Practices COST A32, Discovery, SemLib, Shared Canvas



'Discovery' Corpus: Digitised Manuscripts

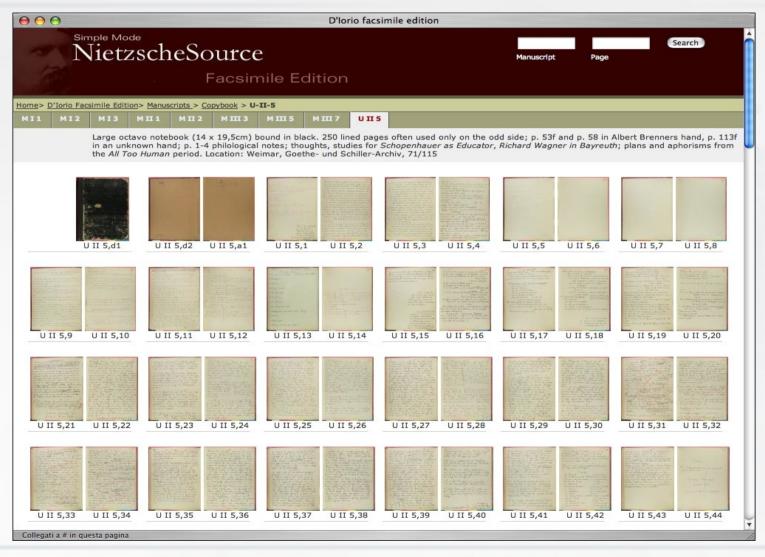






HyperNietzsche: Digitisation, Presentation

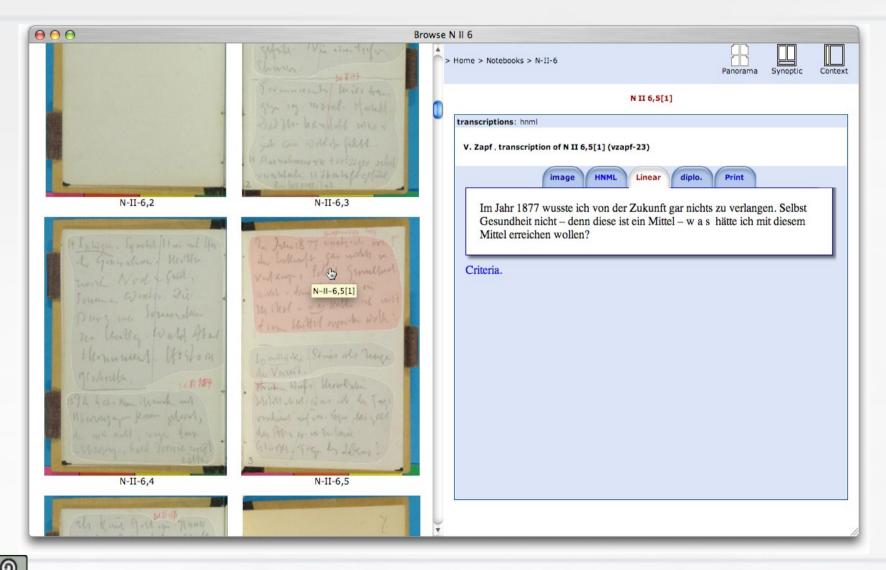






HyperNietzsche: Transcription, Presentation

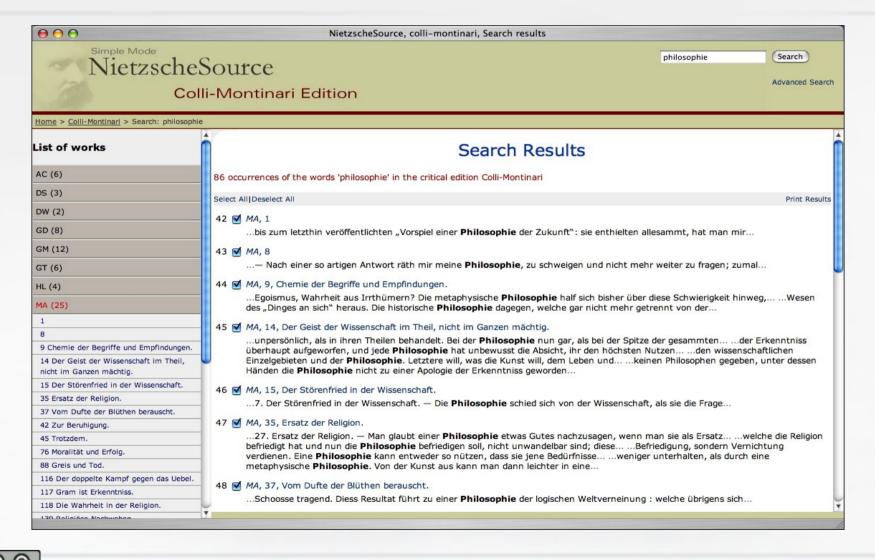






HyperNietzsche: Sources, Editions (synoptic)







Talia: Refactoring Hyper with Semantic Web Technology

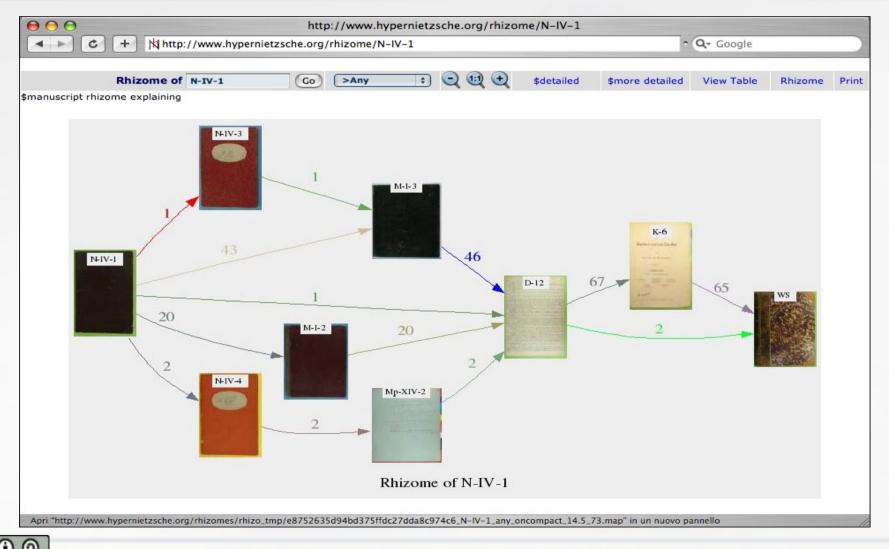


```
N-V-3.71.rdf (modified)
M-V-3,71.rdf (/Users/gradmans/Documents/4-europe/Projects/Discovery/RDF Samples/)
<?xml version="1.0" encoding="UTF-8"?>.
<rdf:RDF
xmlns:taliadom="http://trac.talia.discovery-project.eu/wiki/SampleDomainOntology#" .
xmlns:default="http://default.dummy/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" .
xmlns:xsd="http://www.w3.org/2001/XMLSchema#" .
xmlns:foaf="http://xmlns.com/foaf/0.1/"
 xmlns:local="http://www.talia.discovery-project.org/sources/".
xmlns:dcns="http://purl.org/dc/elements/1.1/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:hyper="http://www.hypernietzsche.org/ontology/"
 xmlns:talias="http://trac.talia.discovery-project.eu/wiki/StructuralOntology#"
xmlns:talia db="http://talia.discovery-project.eu/wiki/DatabaseDupes#"
xmlns:luccadom="http://trac.talia.discovery-project.eu/wiki/LuccaOntology#" .
 xmlns:owl="http://www.w3.org/2002/07/owl#">.
   <rdf:Description about="http://www.talia.discovery-project.org/sources/N-V-3,74">.
     <rdf:type>.
       <rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-schema#Resource"/>.
      <rdf:Description rdf:about="http://www.hypernietzsche.org/ontology/Page"/>
       <rdf:Description rdf:about="http://www.hypernietzsche.org/ontology/Manuscript"/>.
     </re>
     <hyper:position name>71</hyper:position name>.
     <hyper:position>75</hyper:position>.
     <hyper:width>2454.00</hyper:width>
     <talia_db:primary_source>false</talia_db:primary_source>.
     <talia db:workflow state>0</talia db:workflow state>.
     <hyper:part_of>.
       <rdf:Description rdf:about="http://www.talia.discovery-project.org/sources/N-V-3"/>.
     </hyper:part of>.
     <talia db:uri>http://www.talia.discovery-project.org/sources/N-V-3,71</talia db:uri>.
     <dcns:title>N V 3,71</dcns:title>.
     <hyper:height>4080.00</hyper:height>.
  </rdf:Description>.
 </rdf:RDF>.
11.11 All
                                                                (xml,none,UTF-8)- --- U 35/63Mb
```



Generating Stemmata based on Inferencing (1)







Generating Stemmata based on Inferencing (2)







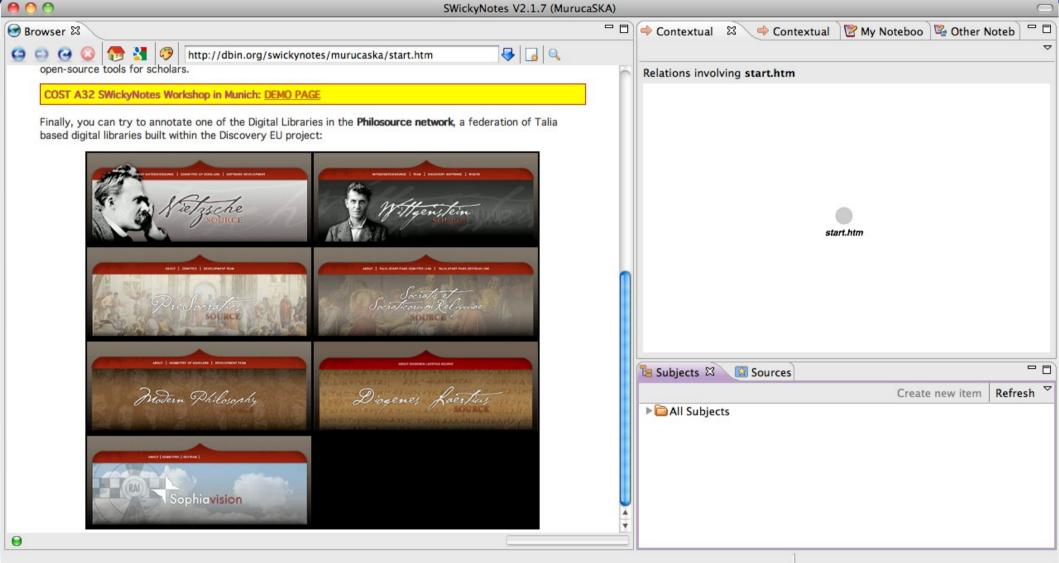


Interpretation: Muruca



SwickyNotes: Ontology Based Annotation as Linked Open Data

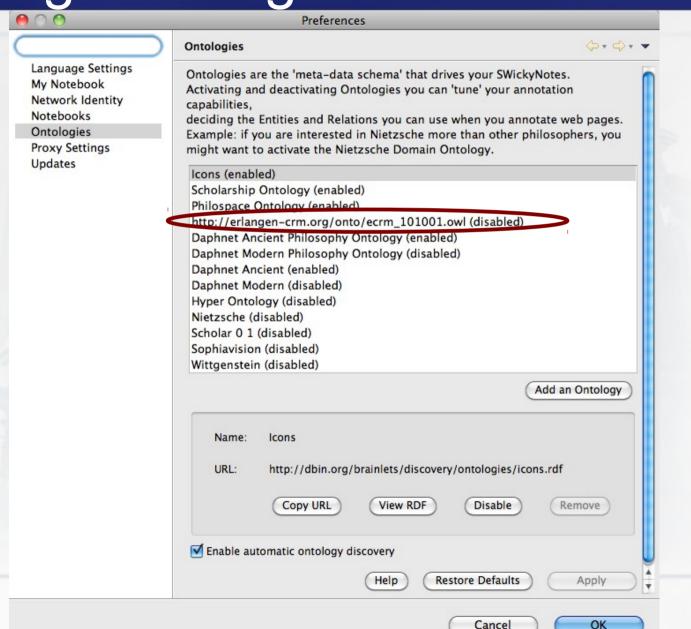






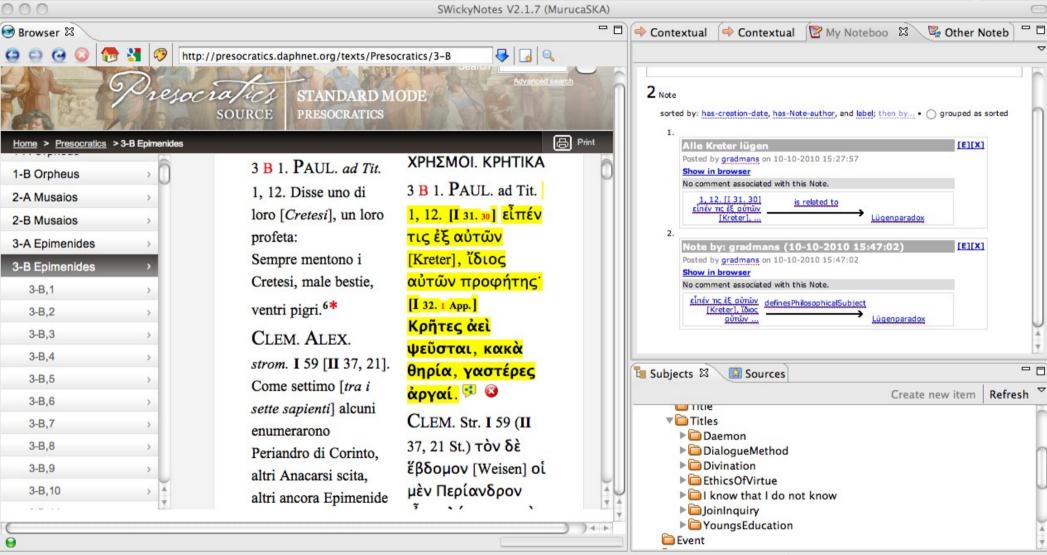
SwickyNotes: Selecting Ontologies





"Cretans are always Liars" ... annotated







SemLib



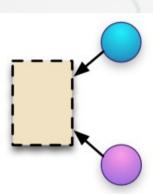
- SemLib (http://www.semlibproject.eu/) is a continuation of Discovery in a EU funded project under FP7 working on
 - A Tool to export existing metadata in RDF and publish it as Linked Data (Web of Data);
 - A Semantic Annotation System, to exploit usergenerated RDF metadata and publish it as Linked Data;
 - A Semantic Recommender System, to use Linked Data to improve searching and browsing in the DLs.



Shared Canvas



Shared Canvas (http://www.shared-canvas.org) is about annotation again – but in a much more sophisticated data model enabling multiple and potentially concurrent layered annotations.



- Demo at http://www.shared-canvas.org/impl/demo1/
- Benjamin, you have at least one adopter: DM2E!

- Common traits
 - Moving away from Database-Portal paradigm
 - Use of RDF as underlying technology
 - Emulation of well known functionality on the Web





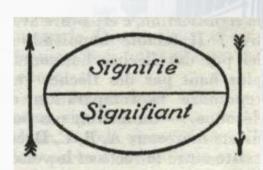
Scholarly Use of Digitisation ... beyond Emulation

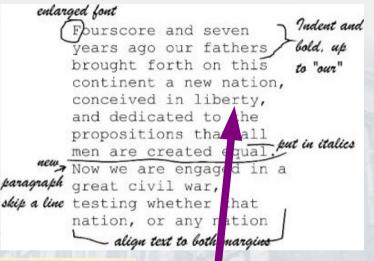


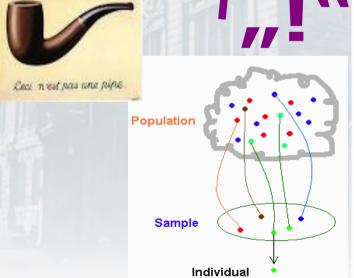
Scholarly Primitives

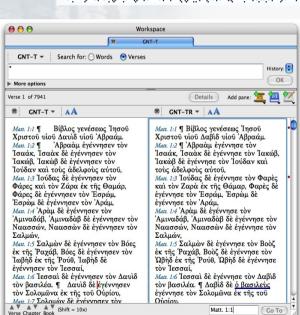


- Discovering
- Annotating
- Comparing
- Referring
- Sampling
- Illustrating
- Representing







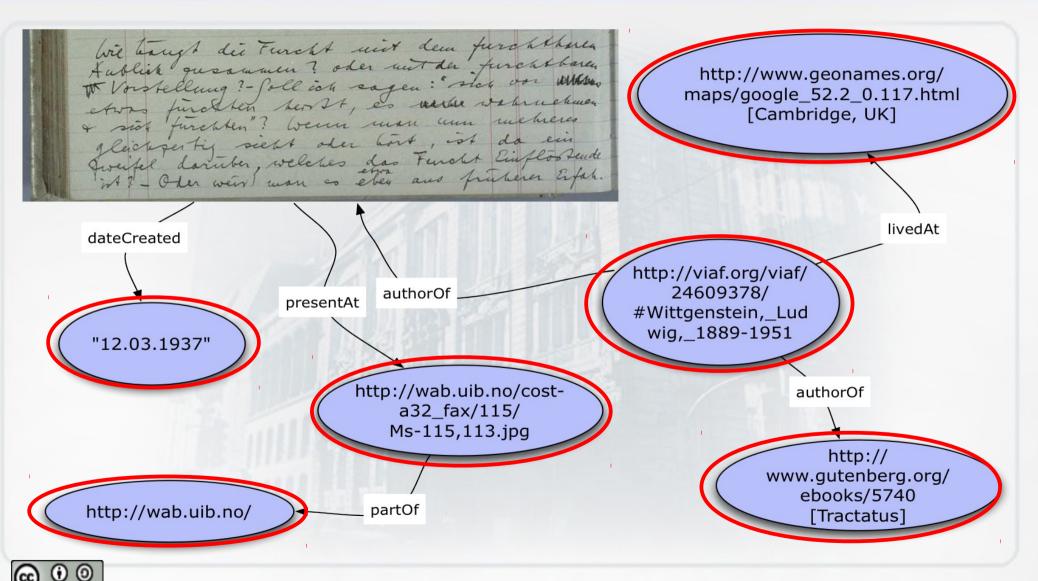






Massive Contextualisation







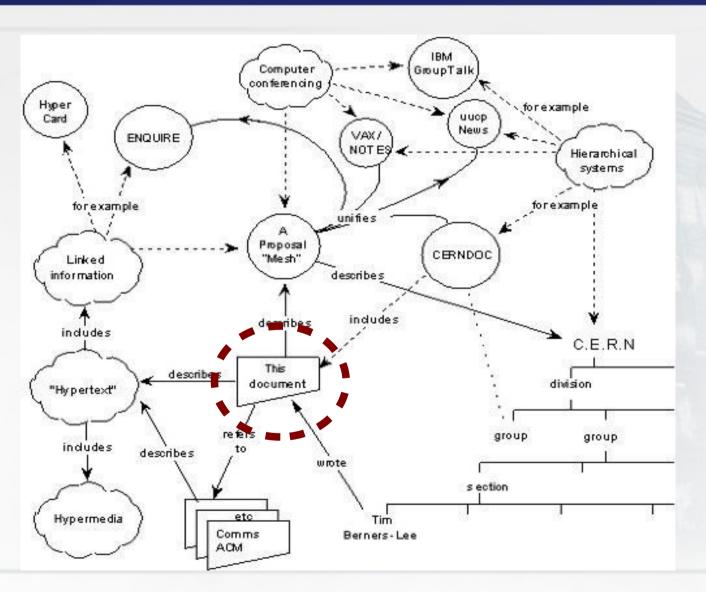


Resource Description Framework (RDF), Linked open Data (LoD) and the Europeana Data Model (EDM)



The Web of Documents





Information Management: A Proposal (TBL, 1989)

The Linked Data Web is a double Extension of this model

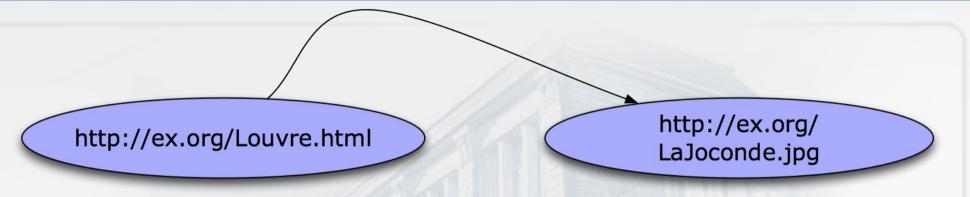
- in **syntax**
- in scope





Resources and Links in the Document Web





b

- 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: RDF (1)



We add a syntax for making statements on resources:

RDF myspace:contains http://ex.org/Louvre http://ex.org/LaJoconde

Or, more generally triples ...

Predicate / Property

Subject / Resource

Object / Value

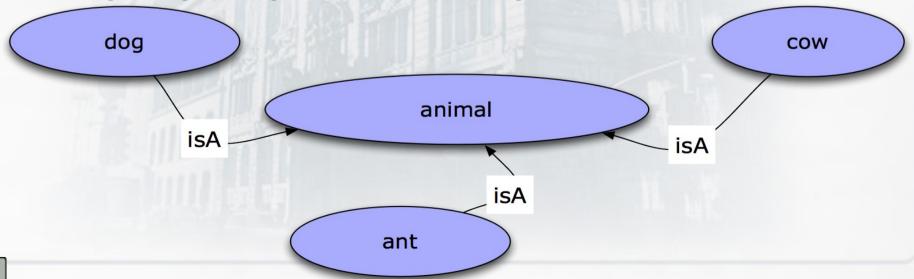
... where S and P are web resources (identified using URIs) and O is either a web resource or a literal



Syntactically Extending the Document Web: RDF (2)



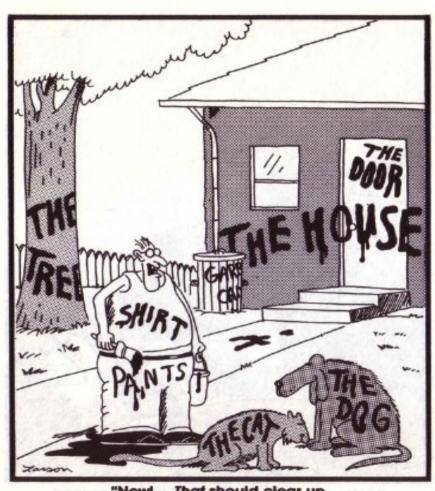
- We add a schema language (RDFS) with elements such as
 - classes (chair' as instance of chairs),
 - hierarchies of classes and properties (chairs are a subclass of furniture, 'teaches' is a sub-property of 'communicates')
 - inheritance (communication based on language → teaching also is)
 - support for basic inferencing.
- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:





The Web of Things ... Somewhat Mistaken





"Now! ... That should clear up a few things around here!"

What's wrong with this picture?

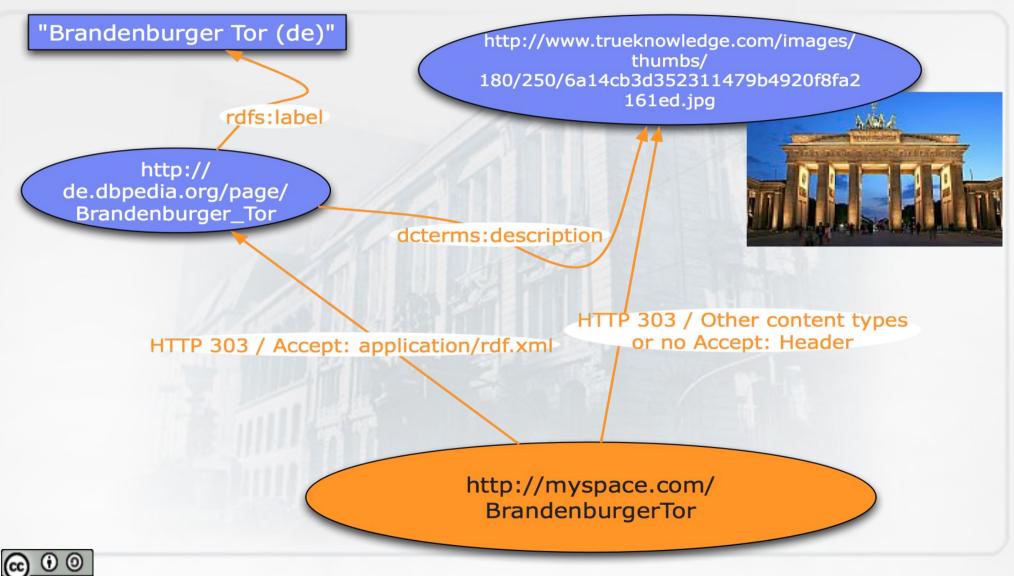
Taken from Ronald Carpentier's Blog at

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

- I. Use URIs Standard Identifiers
- 2. Use HTTP URIs Standard Pointers
- 3. Serve useful information using SPARQL, RDF standards
- 4. Mention URIs of related objects

Standards for **Queries** and **Statements**

Link to Context

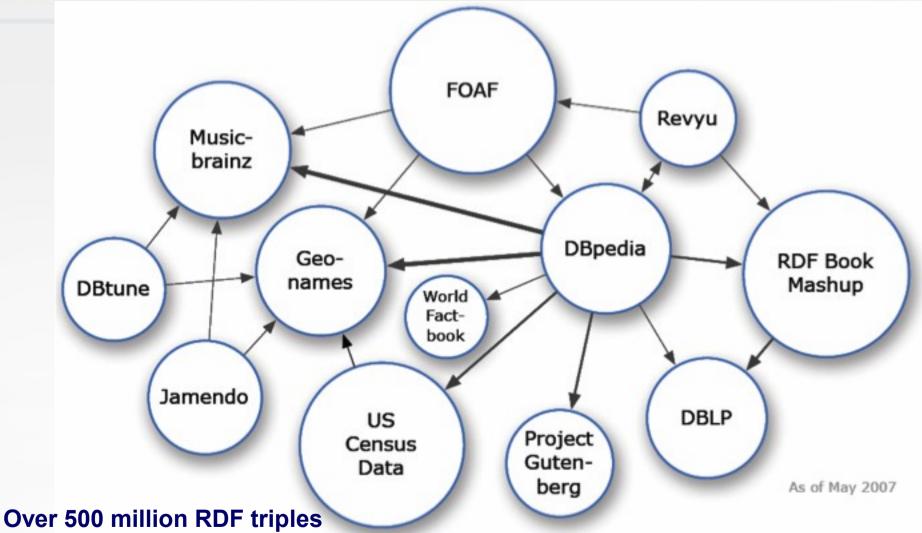
Copyright © 2008 W3C (MIT, ERCIM, Keio)

http://www.w3.org/2008/Talks/0617-lod-tbl/#(4)



A few Bubbles: 5/2007





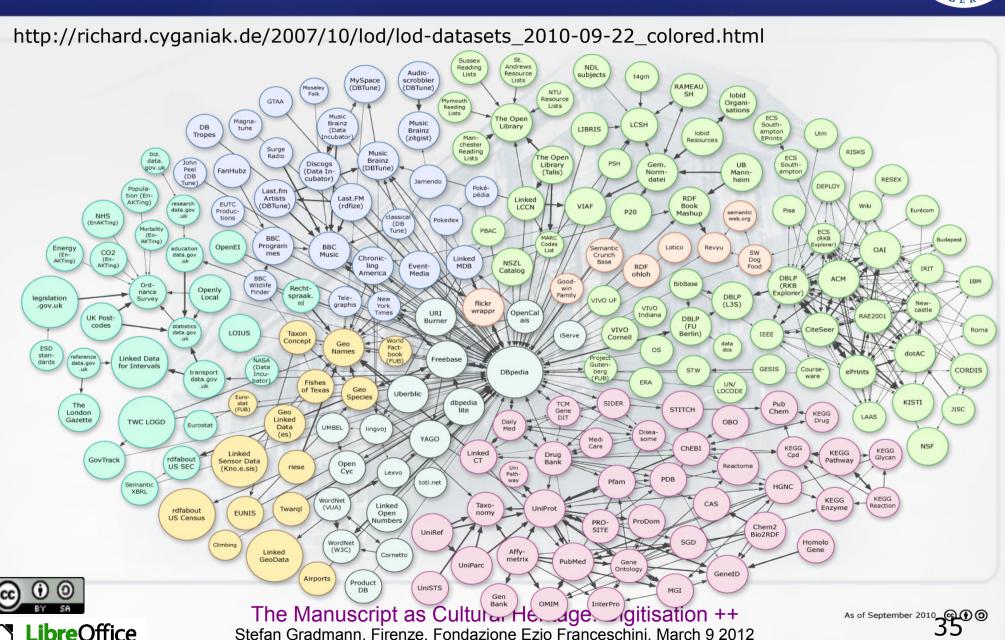


© Richard Cyganiak



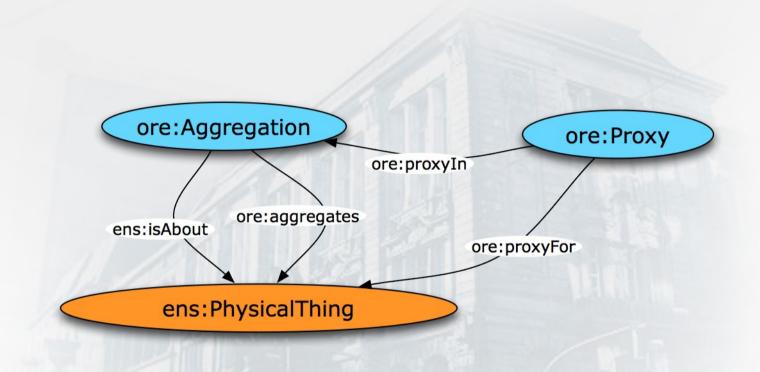
A Whole Lot of Bubbles: 09/2010





An EDM Aggregation ...

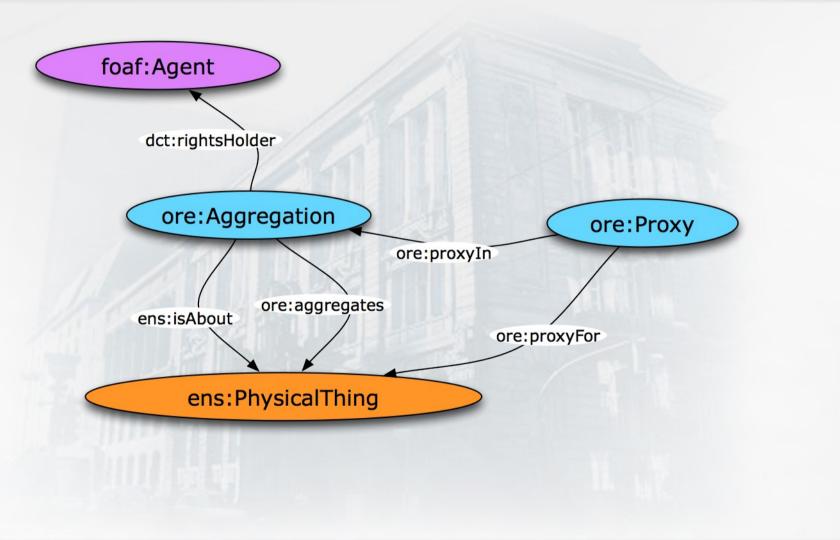






... some context

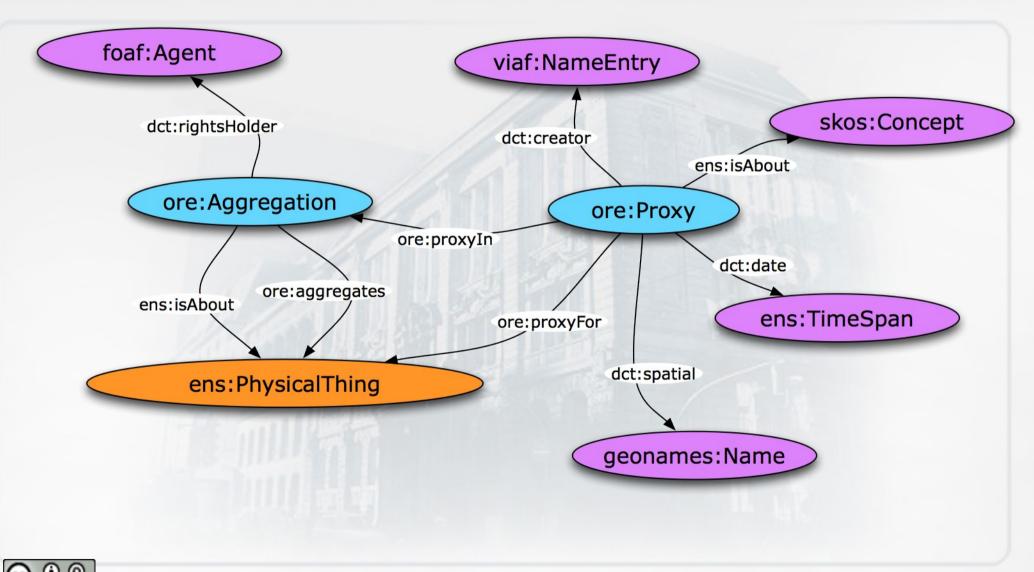






... more context



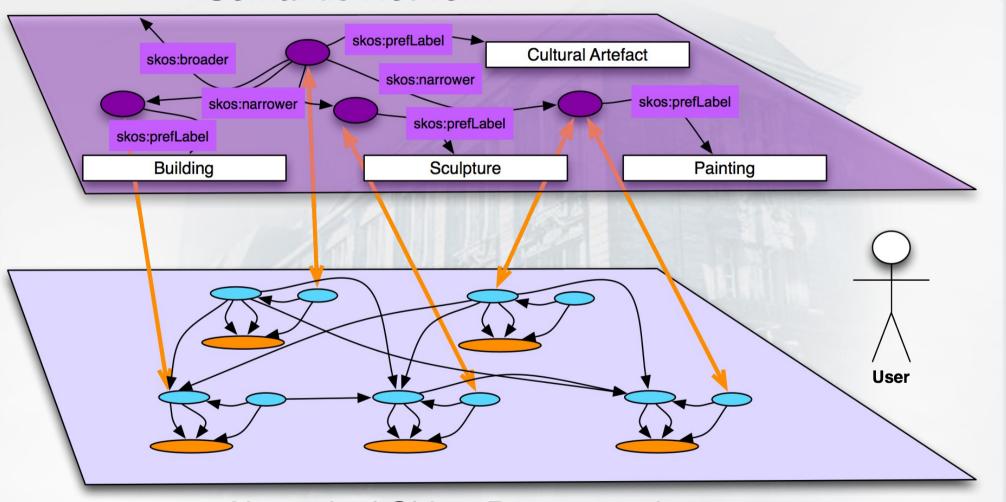




... and the Big Picture: The Semantic Data Layer



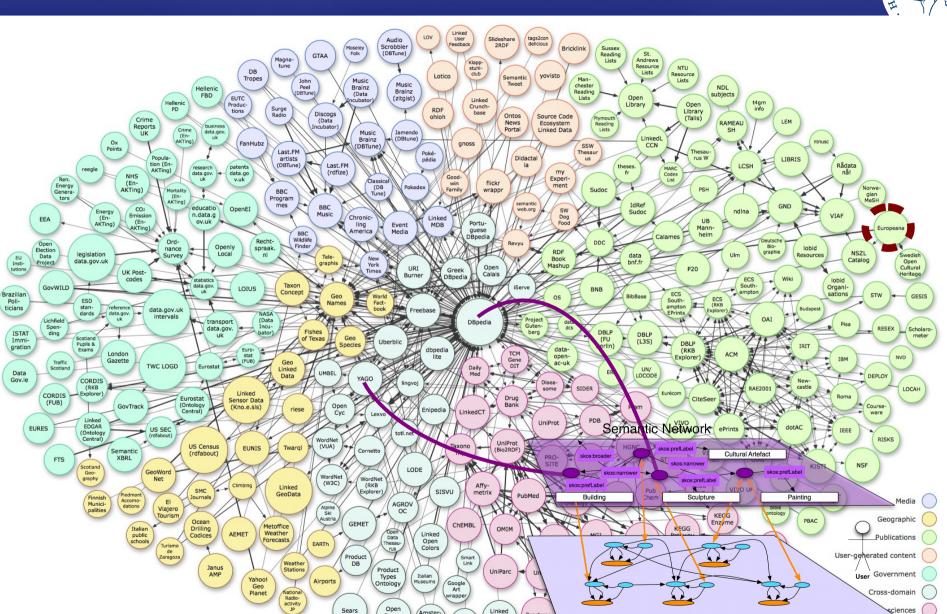




Networked Object Representations



EDM and Linked Open Data



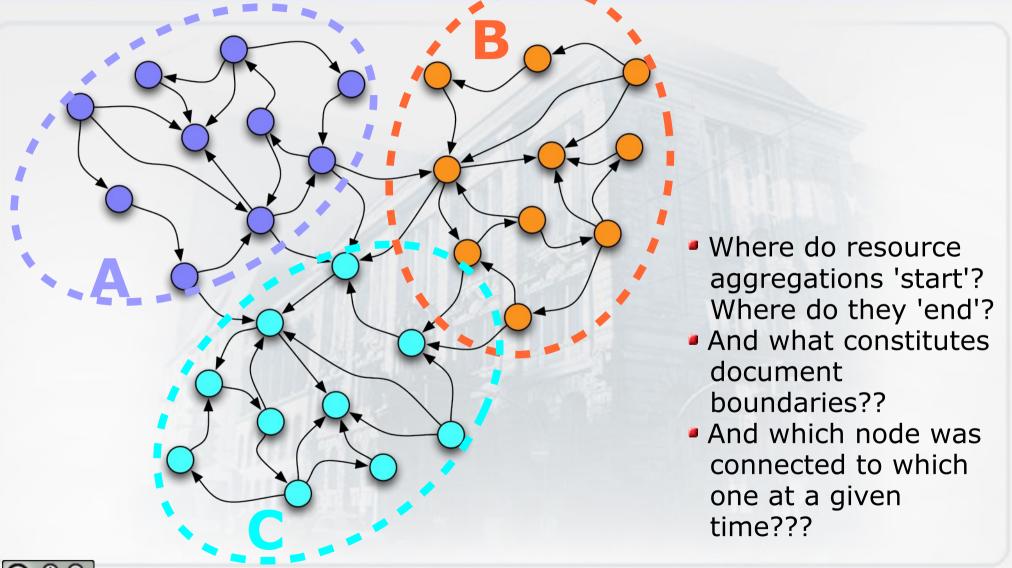




Networked Object Representations

'Beyond Catalogues and Records' generates new questions!

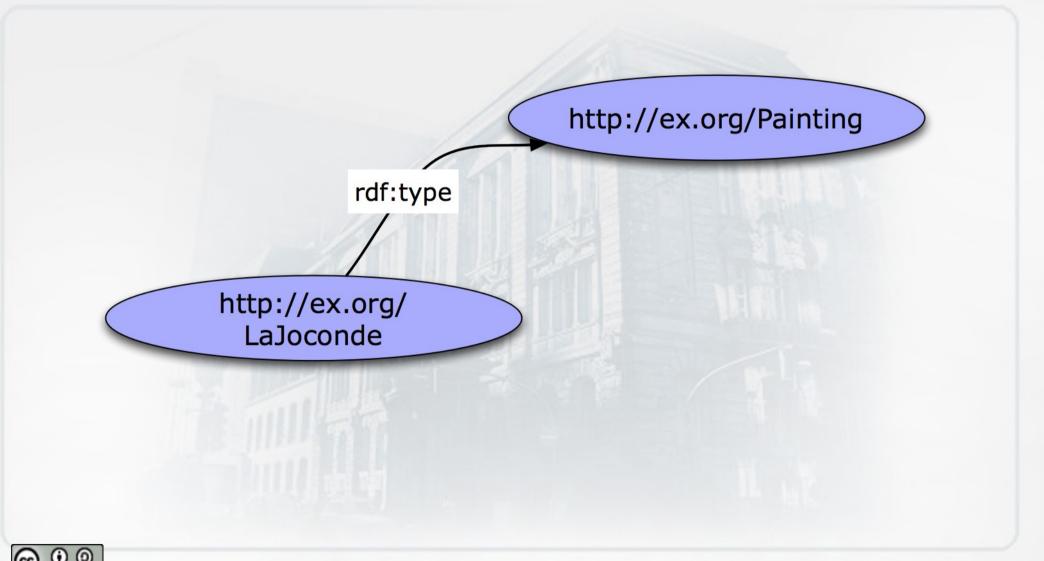






... and new opportunities: Triple Sets and Reasoning (1)

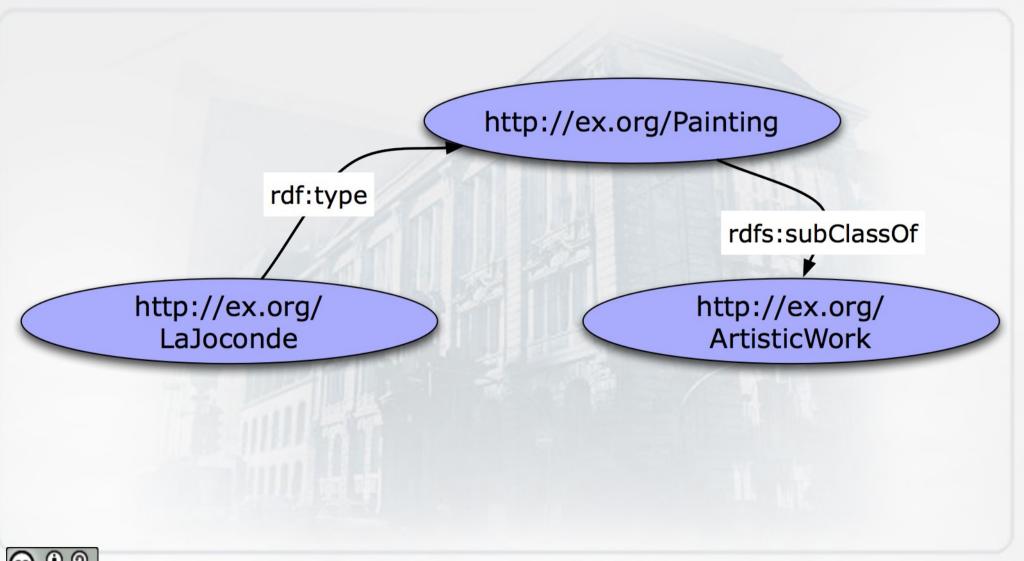






Triple Sets and Reasoning (2)

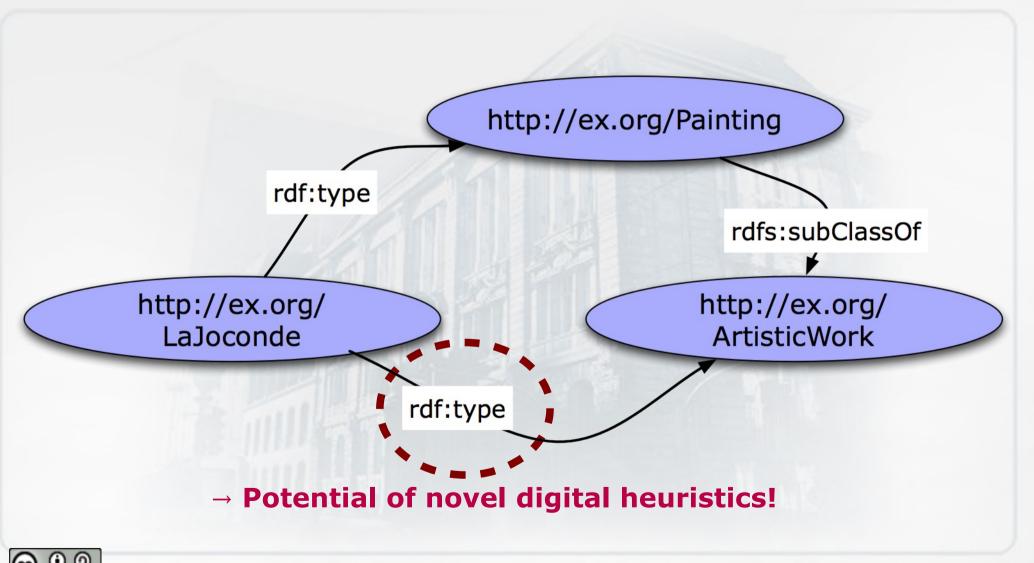






Triple Sets and Reasoning (3)



















... and Europeana?







Digitised Manuscripts to Europeana (DM2E): What?



- WP1: Provide substantial amounts of digital content to Europeana with a focus on digitised manuscripts
- WP2: Integrate existing technical building blocks
 - from Europeana development
 - as well as from generic LoD oriented development
 - into a generic production chain for migrating data from various sources to the EDM as well
 - as for the contextualisation of the object representations.
- WP3: Explore usage scenarios of such content in a specialised platform for humanities research generating digital heuristics and making data as well as heuristics available to specialised visualisation or reasoning environments.



Digitised Manuscripts to Europeana (DM2E): Who (1)?

Content Providers

- European Association for Jewish Culture (Judaica)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Österreichische Nationalbibliothek (Google)
- Staatsbibliothek zu Berlin (Kalliope)
- University of Bergen (Wittgenstein)
- CNRS ITEM (Nietzsche)
- National Library of Israel (Judaica)
- Berlin Brandenburgische Akademie (German Text Archive)
- Humboldt-Universität zu Berlin (Polytechnisches Journal)

Technology Providers

- ExLibris (Aleph, MARC sources management)
- Freie Universität Berlin (dBPedia, LoD, LoD2, D2R, SILK)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Net7 S.r.l. (Muruca)
- National Technical University of Athens (MINT)



Digitised Manuscripts to Europeana (DM2E): Who (2)?

Digital Humanities Community

- Dr. Tobias Blanke (King's College, London)
- Sally Chambers (The European Library / DARIAH-D)
- Prof. Dr. Stefan Gradmann (HUB)
- Prof. Dr. Gerhard Lauer (Göttingen Center for the Digital Humanities)
- Dr. Alois Pichler (UIB)
- Dr. Jürgen Renn (MPIWG)
- Prof. Dr. Susan Schreibman (Trinity College Dublin)
- Dr. Claire Warwick (University College, London)

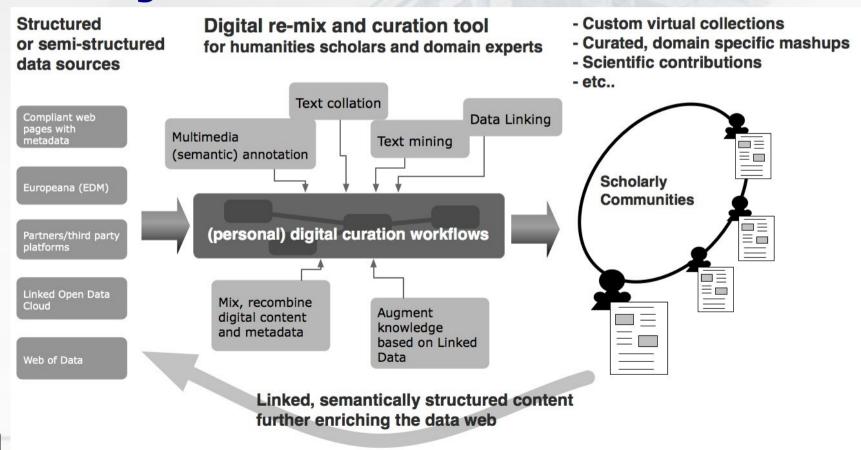
Community Building

- Open Knowledge Foundation (OKFN)
- Coordination, Management & Information Science
 - Humboldt-Universität zu Berlin (HUB)



WP3: Digital Humanities Requirements and Related Engineering - Context

Goal: lower the barriers for **digital content curation** by providing an integrated, flexible, semantic based environment targeted to **digital humanities scholars**





Digital Humanities Functionality (WP3)



- How to enable digital scholarship building on combined EDM metadata and digital surrogates ...
- ... building on an ontological, granular representation of John Unsworth's scholarly primitives (or their successors according to Blanke/Hedges 2011) ...
- and resulting in a increasingly complex social semantic
 scholarly graph containing RDF statements such as
 - Version_A isSuccessorOf Version_B
 - Statement₁ contradicts Statement₂
 - Scribe_y copiedFrom Scribe_z
- building scholarly heuristics on inferences on this graph ...



Conclusion (after-lunch wake-up)



- Consider digitised manuscripts as cultural heritage in their own right!
- Leave behind the walled gardens: 'databases' are a point of departure, not a target!
- RDF and its future extensions is key technology!
- Needs of scholarly users should be the driving force
 - We need to better understand what scholars are actually doing
- Try to avoid a few terms librarians are so fond of such as 'Catalogue' or 'Holdings': they are deeply rooted in the walled garden paradigm!
- Have a look at http://dm2e.eu from mid-next week on!



Selected Reading



- Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, Herbert van de Sompel (2010): The Europeana Data Model. IFLA 2010 (Gothenburg). Session on "Libraries and the Semantic Web". http://www.ifla.org/files/hq/papers/ifla76/149-doerr-en.pdf
- Stefan Gradmann (2010): Knowledge = Information in Context: on the Importance of Semantic Contextualisation in Europeana. Europeana White Paper 1.
 - http://www.scribd.com/doc/32110457/Europeana-White-Paper-1
- John Unsworth (2000): Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this? In the seminar on Humanities Computing, King's College, London.
 - http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html
- Tobias Blanke, Mark Hedges (2011): Scholarly primitives. Building institutional infrastructure for humanities e-Science, Future Generation Computer Systems, In Press (Science) http://www.sciencedirect.com/science/article/pii/S0167739X11001178#br000085





Thank you!

