Objects, Process, Context in Time and Space: what we can do with the Europeana Data Model (EDM) - and what we cannot do

Providence, Brown University
15 March 2012

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

LibreOffice
The Document Foundation

Overview

- What is the EDM? And what it is not!
- Objects and Context modelled in EDM: Examples
- EDM and LoD
- A few words on RDF
- Potential, Problems and Limitations
- DM2E: Embedding the EDM in a Scholarly semantic Graph

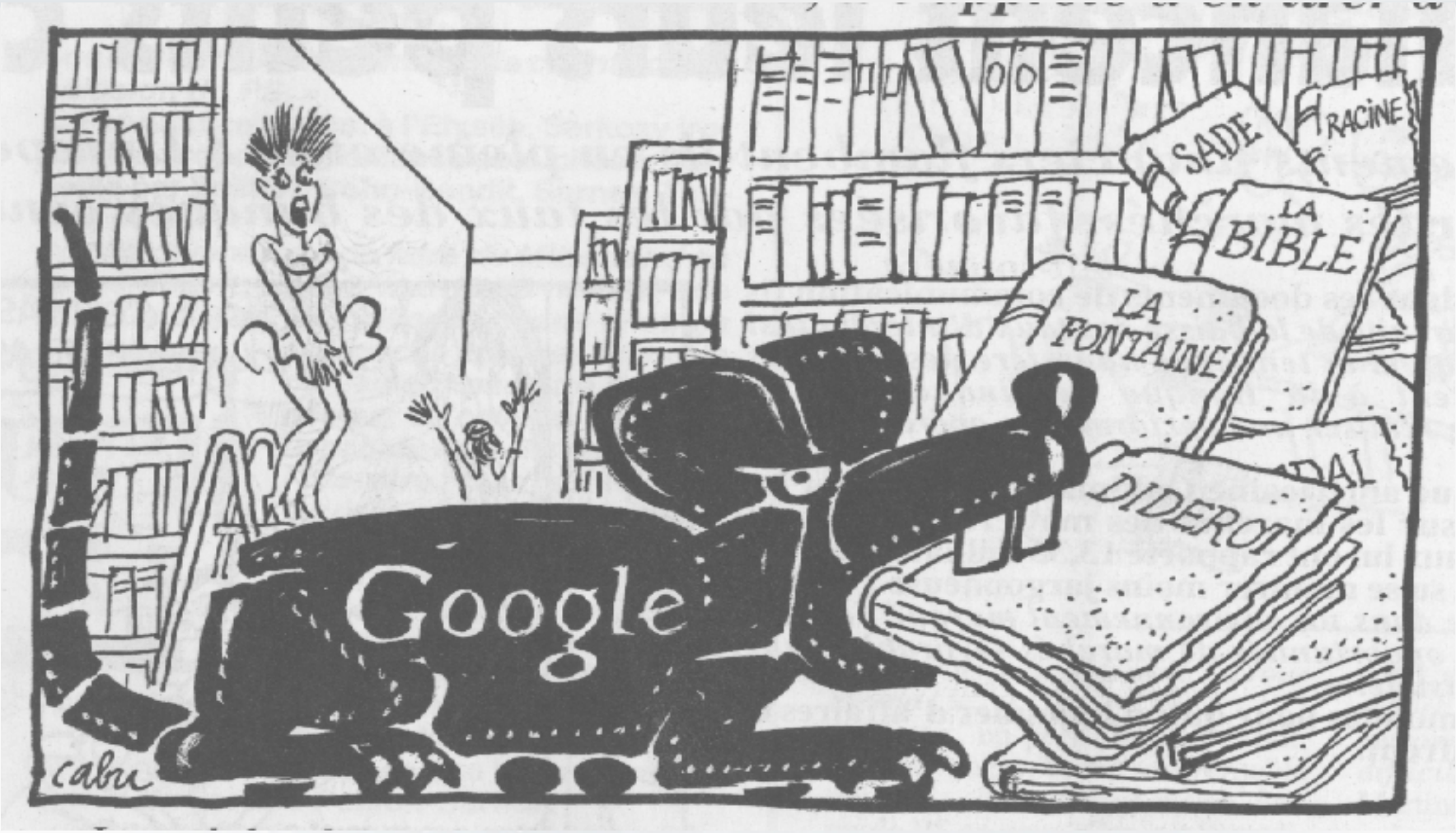
LibreOffice
The Document Foundation

The Europeana Data Model (EDM) ...

CC BY SA

LibreOffice
The Document Foundation

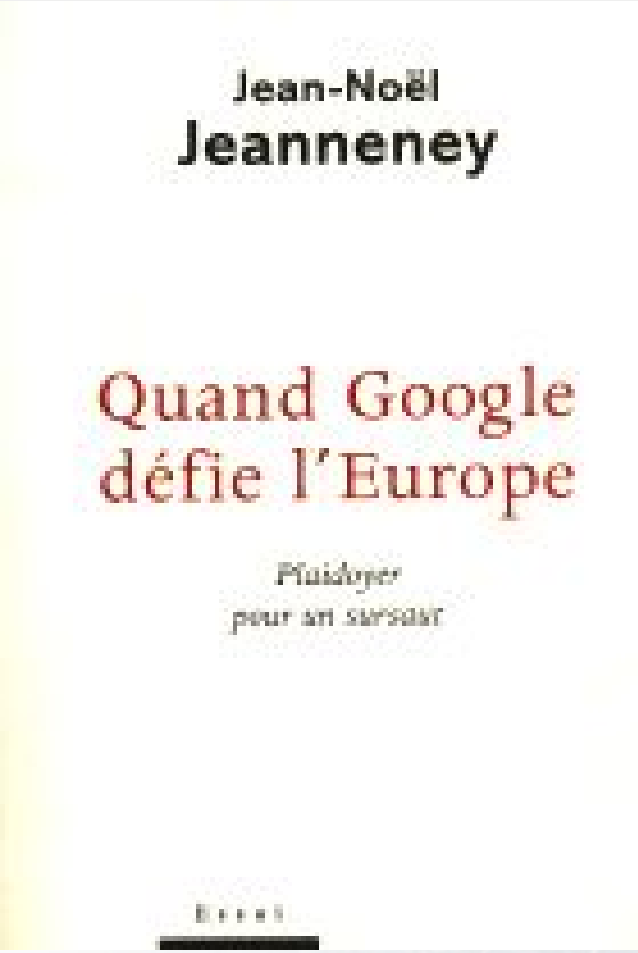
Pre-EDM: 2005



Google Books in 2005 as seen by the Canard Enchainé ...

BY SA

Europeana 2005 - 2011



- … and by Jean-Noel Jeanneney
- → Joint declaration of six European Governments … Chirac and Schröder among those!
- Declaration by Commissioner Reding on the "European Digital Library" flagship as part of i2010.
- Launch (and crash) of a first prototype 20.11.2008
- Constant growth and stable operations since then, > 20 M objects represented to date
- Still basically yet another portal, a big one – but nothing innovative ...

CC BY SA

ESE

"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" :-)

CC BY SA

ESE ...


```
<...xmlns:europeana="http://www.europeana.eu/schemas/ese/"
...
<!--National Library of Poland-->
<record>
  <dc:identifier>urn:repox.bn.org.pl:polona:oai:www.polona.pl:4445</dc:identifier>
  <europeana:uri>http://www.europeana.eu/resolve/record/01406/18A8F5DFCAE75694
  A8AAD18E2791861FBE92589A</europeana:uri>
    <dc:title>[Ośmiu powstańców uzbrojonych w kosy]</dc:title>
    <dc:subject>Powstanie 1863 r. styczniowe - ikonografia</dc:subject>
    <dc:date>[1863-1864]</dc:date>
    <dc:rights>Biblioteka Narodowa</dc:rights>
    <europeana:object>http://193.59.172.16/szzz/IsShownBy.do?
  id=6481</europeana:object>
    <europeana:isShownAt>http://193.59.172.16/szzz/ShowStart.do?
  id=6481</europeana:isShownAt>
</record>
```


CC BY SA

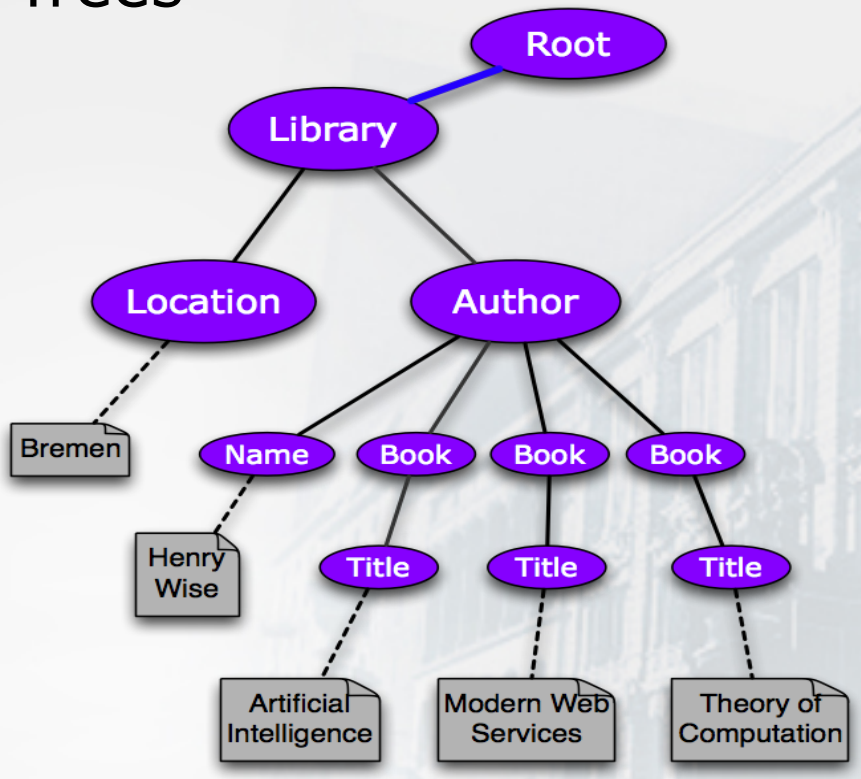
EDM – what is it? And what not?

- EDM is the **metadata model** in the course of replacing the ESE …
- … a model for making statements about **digital representations** of cultural heritage objects
- … a model for **contextualising** such representations
- EDM is **not an object model** (but might be combined with object and process models)!
- EDM is an **RDF based graph model**

CC BY SA

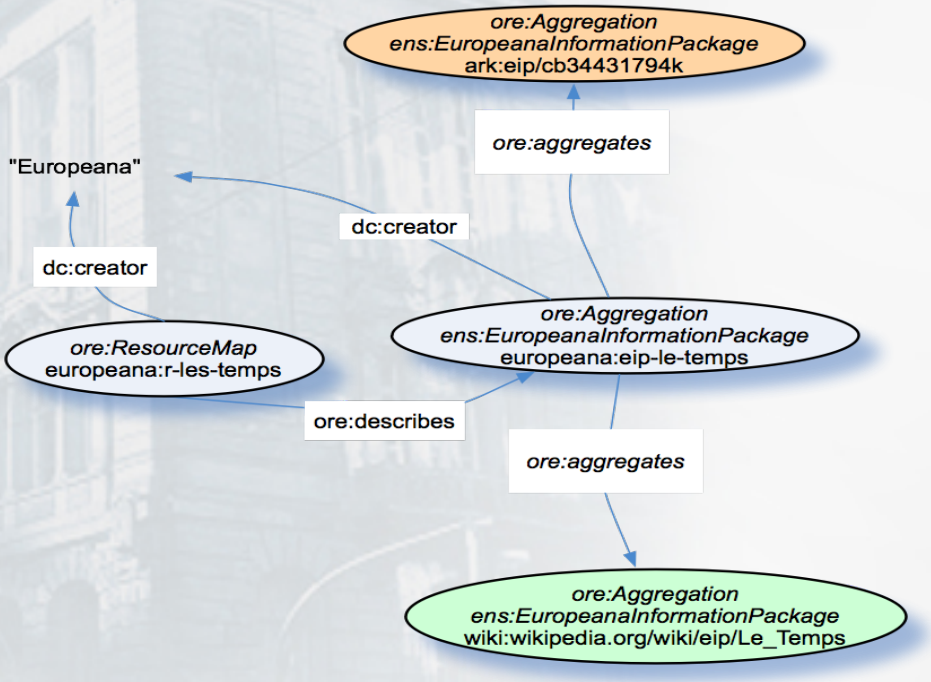
XML vs. RDF:
Trees vs. Graphs, Schema vs. Schema

XML models knowledge as 'Trees'



Schema: Elements & Types
"Validation", prescriptive

RDF models knowledge as graphs



Schema: Classes & Properties, Inheritance, "Inferencing"

LibreOffice
The Document Foundation

EDM: Requirements and Design Principles

- Distinction of real **object** (book, bicture, file, multimedia object) and its digital **representation**
- Distinction between **object properties** and descriptive **metadata** attributes
- **several perspectives of object** must be possible with potentially contradicting statements
- Support for complex **composite objects**
- **Standard** format for **metadata** with **specializing options**
- **Standard** format for **vocabularies** with **specializing options**
- Maximum **reuse** of existing standards

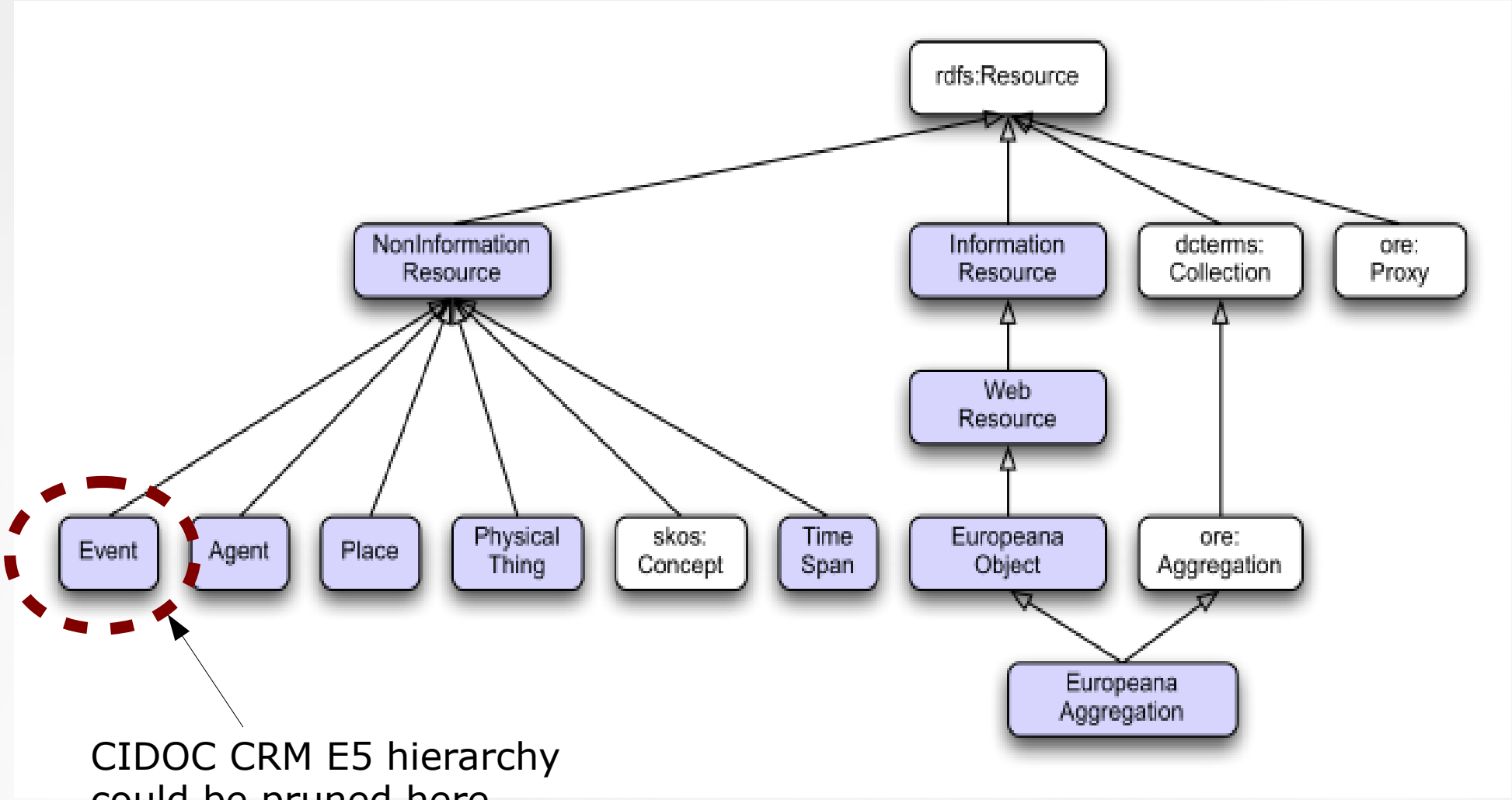
LibreOffice
The Document Foundation

EDM and 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: Description of the provided object from one given perspective.
    - Aggregation: Groups all information pieces together.

LibreOffice
The Document Foundation

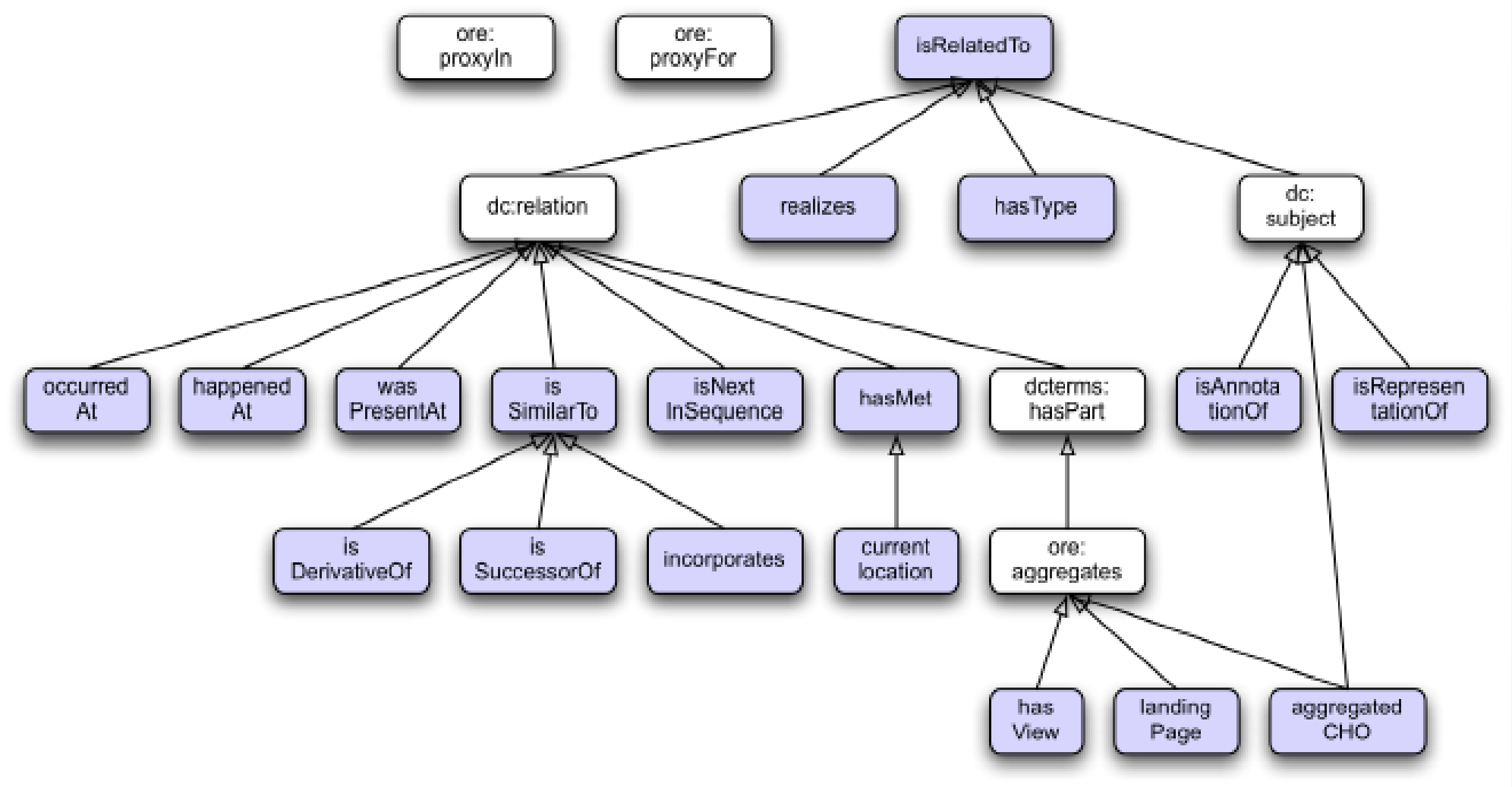
EDM: Classes



CIDOC CRM E5 hierarchy could be pruned here

CC BY SA

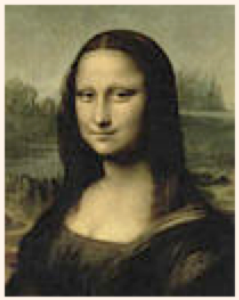
EDM: Properties

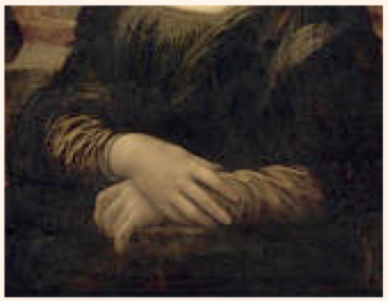


BY SA

Mona Lisa: French Ministry of Culture

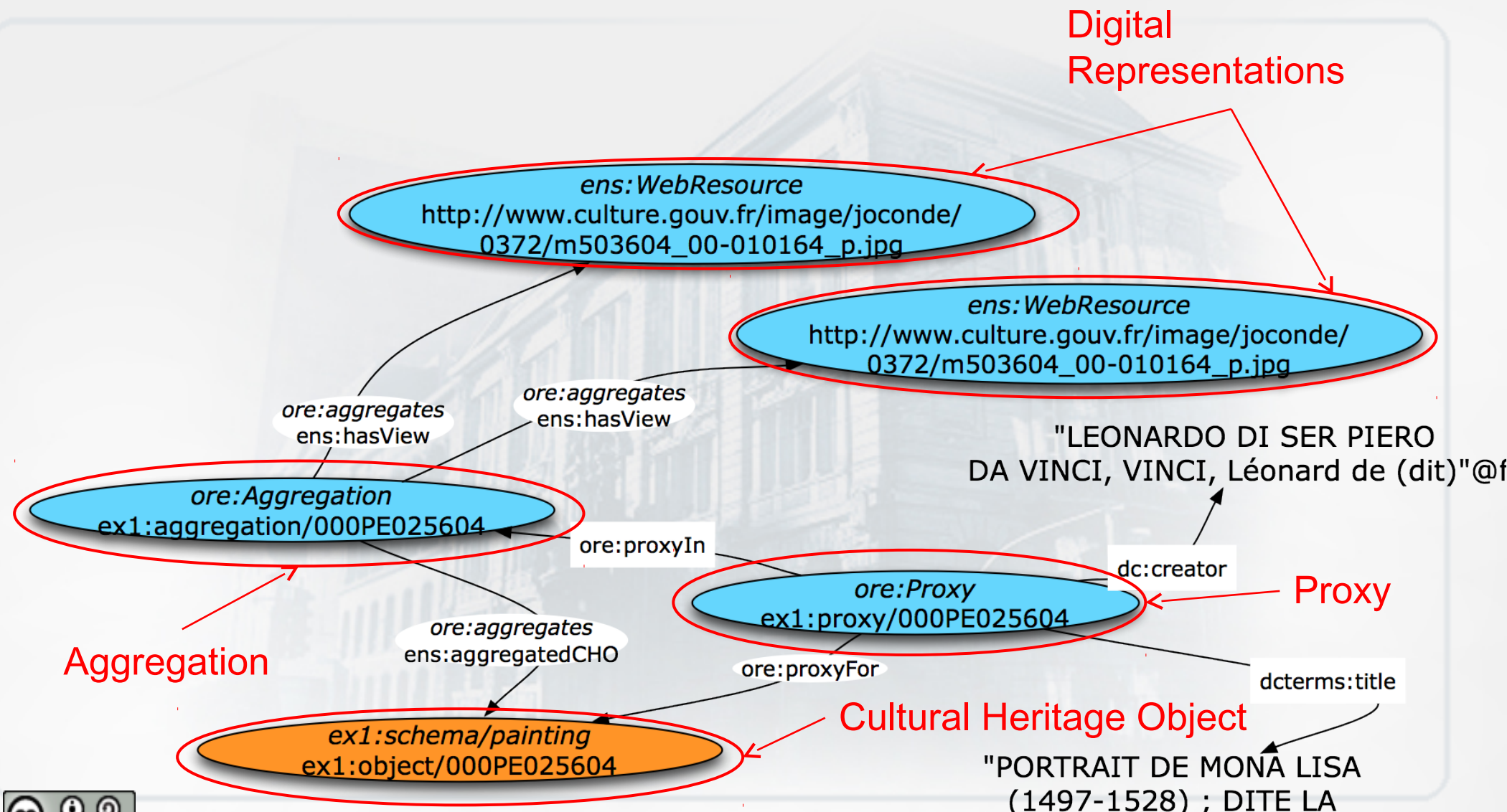
Réponse n° **1**





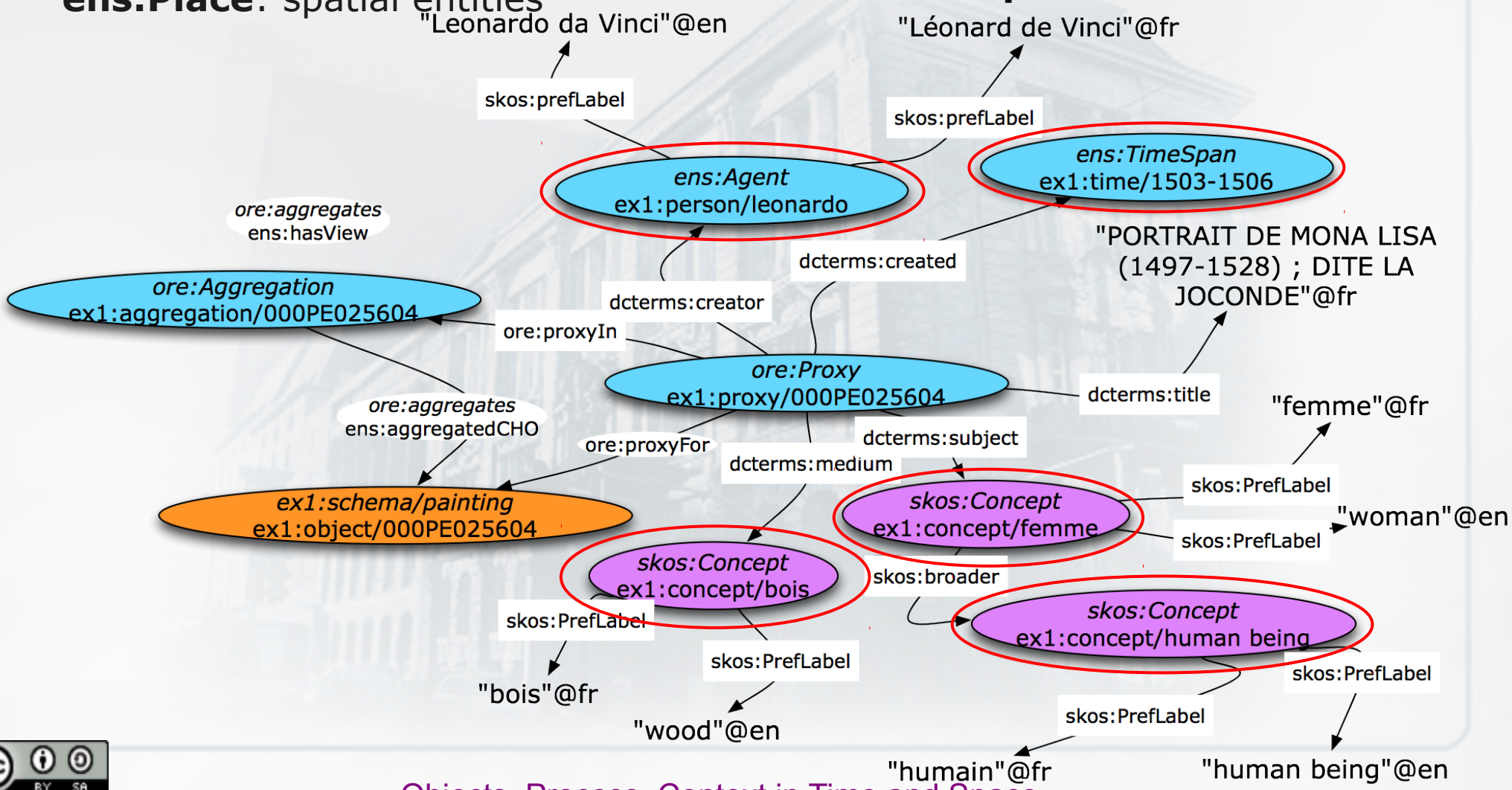
Domaine	peinture
Type d'objet	tableau
Titre	**PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE**
Auteur/exécutant	LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)
Précision auteur/exécutant	Vinci, 1452 ; Amboise, 1519
Ecole	Italie
Période création/exécution	1er quart 16e siècle
Millésime création/exécution	1503 entre ; 1506 et
Genèse	oeuvre en rapport ; reproduit en gravure
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
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 (montagne, 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'Etat ; 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. 1981, 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



Semantic Enrichment

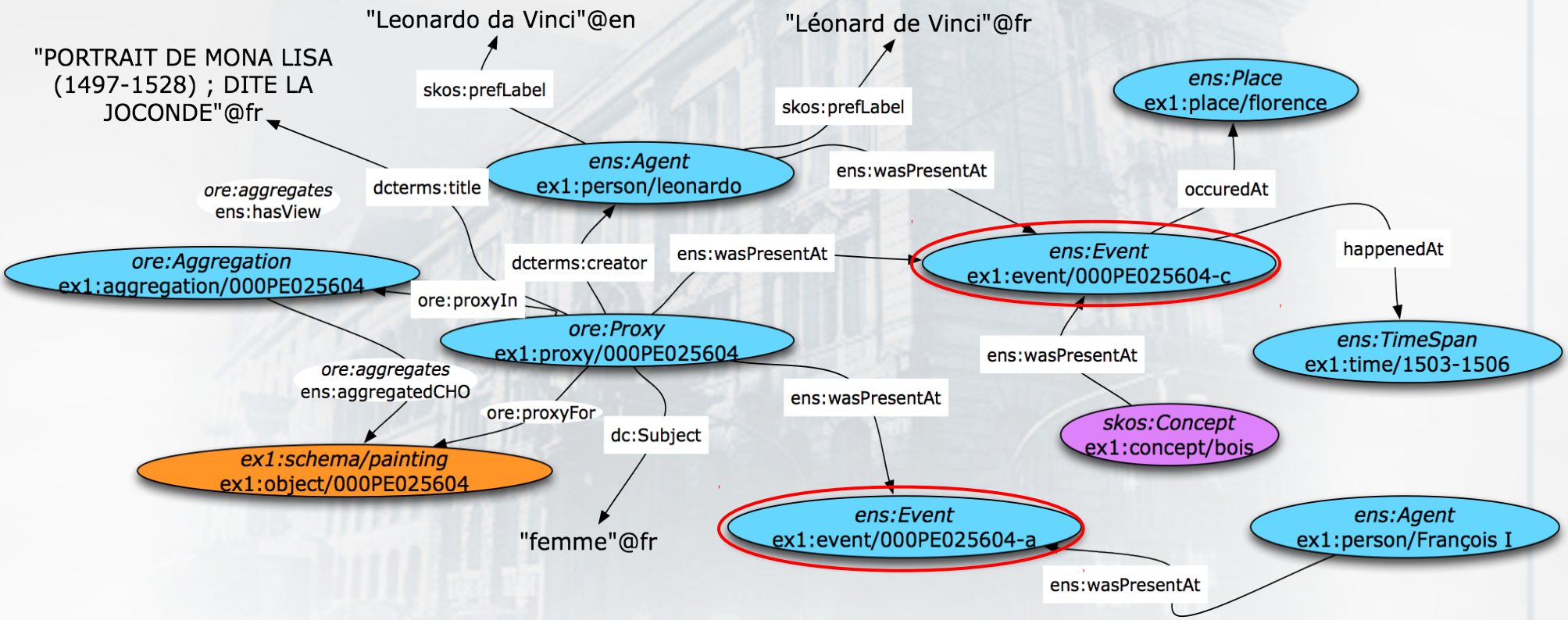
ens:Agent: persons or organizations
ens:Place: spatial entities
ens:TimeSpan: time periods or dates
skos:Concept: entities from KOS



CC BY SA

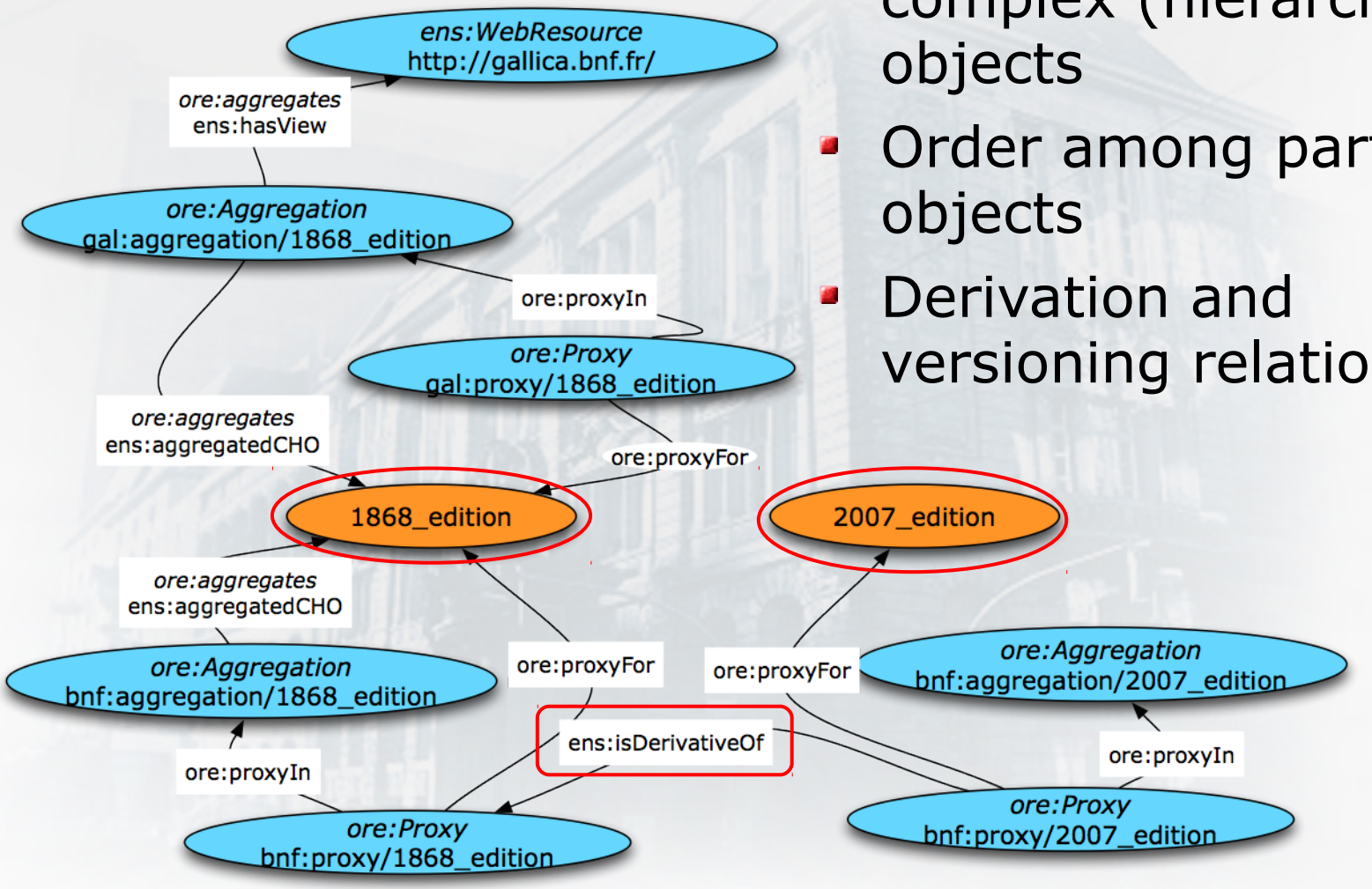
Event-Centric Modeling

Preserving and exploiting original data also means being compatible with descriptions beyond simple object level (→ CIDOC CRM!)



LibreOffice
The Document Foundation

Complex Objects



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

BY SA

Les Fleurs du Mal: UNIMARC

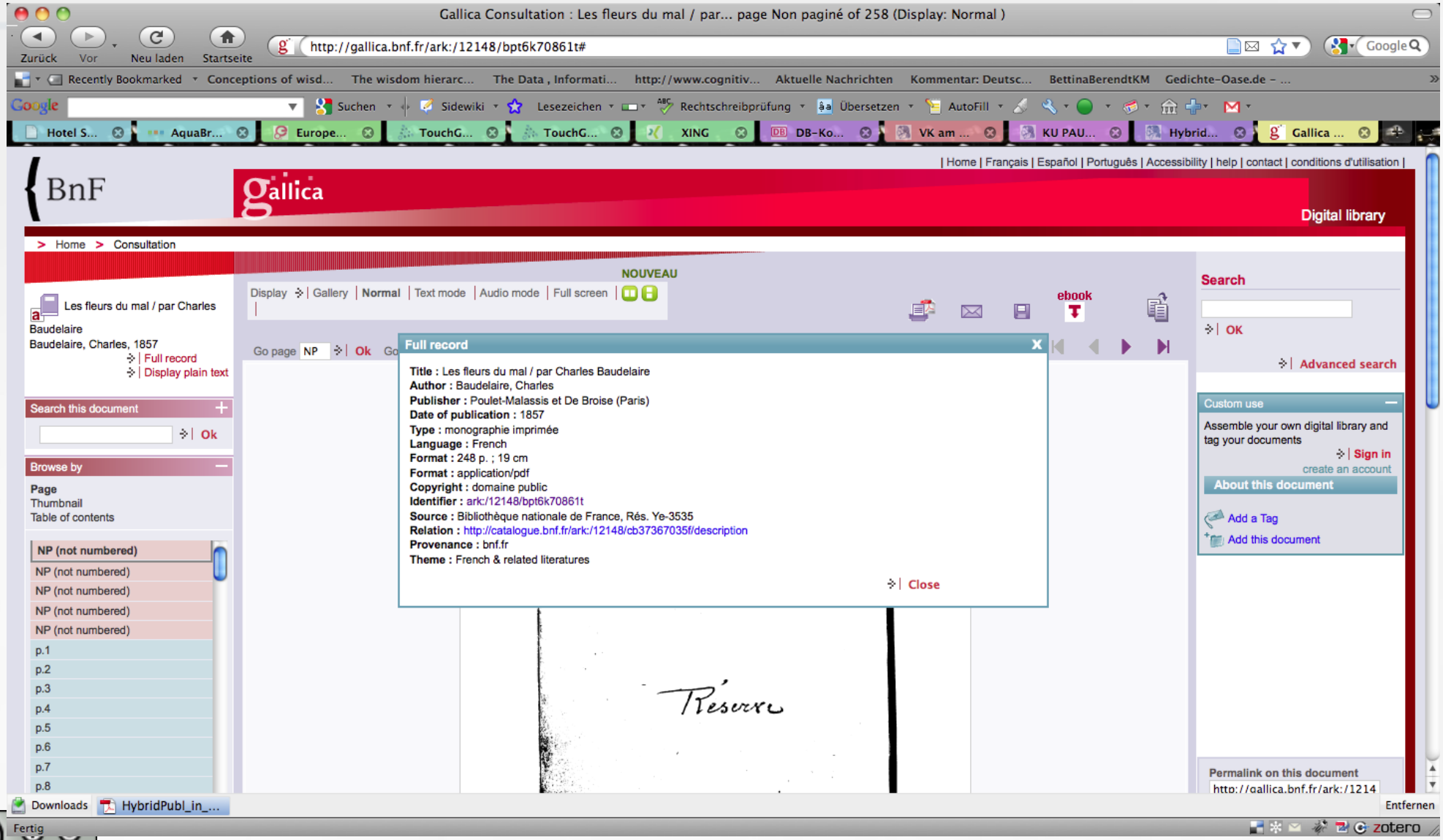
http://catalogue.bnf.fr/ark:/12148/cb37367035f

000 nam 22 450
001FRBNF373670350000003
009http://catalogue.bnf.fr/ark:/12148/cb37367035f
039 $oGEA$a000288182
100 $a19920409d1857 m y0frey50 ba
1010 $afre
102 $aFR
105 $a||||z 00|||
106 $ar
2001 $a^Les ‰fleurs du mal$bTexte imprimé$fpar Charles Baudelaire
210 $aParis$cPoulet-Malassis et De Broise$d1857
215 $a248 p.$d19 cm
676 $a841.8$v22
686 $a840$2Cadre de classement de la Bibliographie nationale française
700 |311890582aBaudelaire$bCharles$4070
801 0aFRbBNF$c19920409$gAFNOR$2intermrc

cc BY SA

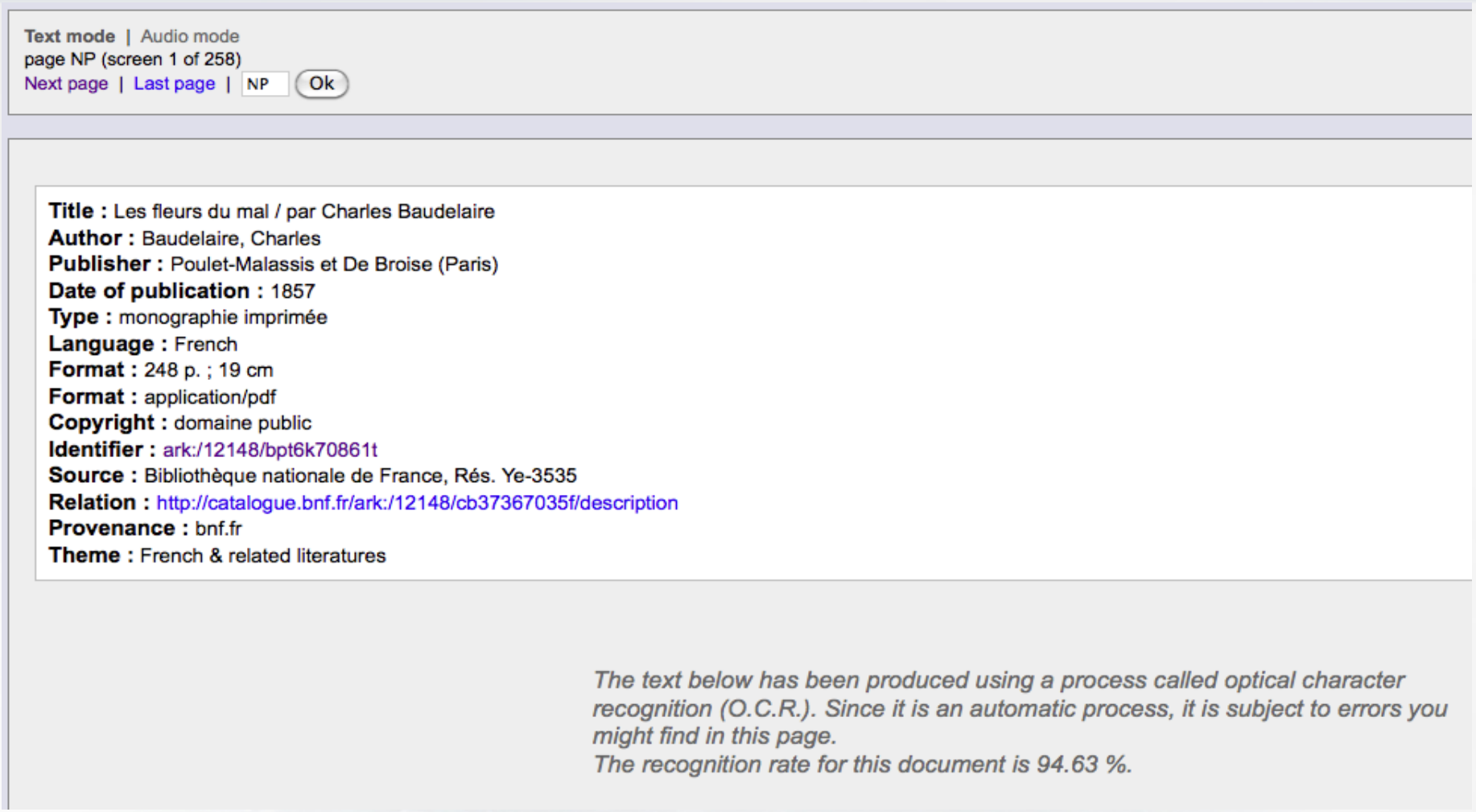
Les Fleurs du Mal: Gallica

http://gallica.bnf.fr/ark:/12148/bpt6k70861t



Les Fleurs du Mal: Digitised

http://gallica.bnf.fr/ark:/12148/bpt6k70861t.textePage.f1



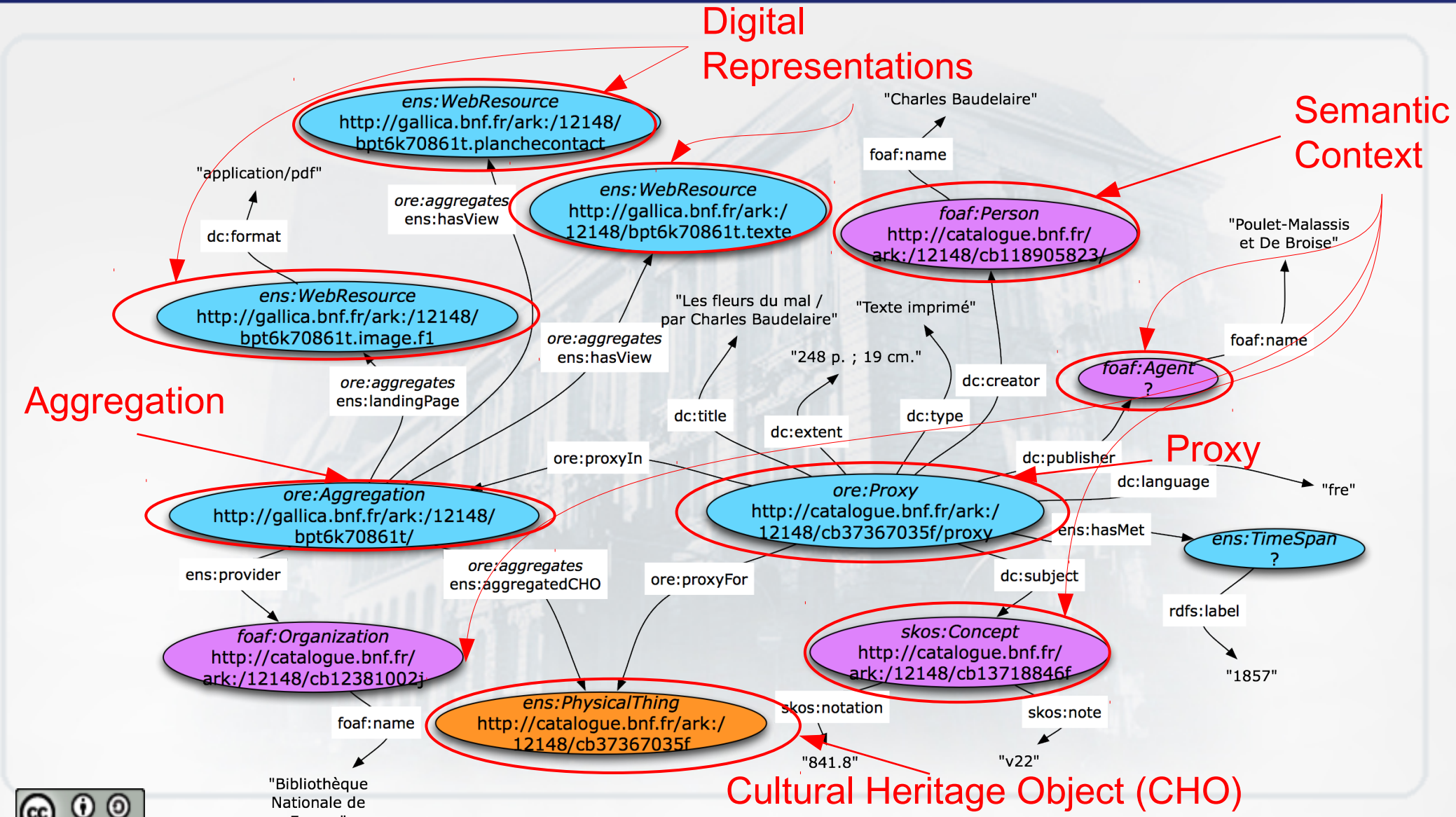
Text mode | Audio mode
page NP (screen 1 of 258)
Next page | Last page | NP Ok

Title : Les fleurs du mal / par Charles Baudelaire
Author : Baudelaire, Charles
Publisher : Poulet-Malassis et De Broise (Paris)
Date of publication : 1857
Type : monographie imprimée
Language : French
Format : 248 p. ; 19 cm
Format : application/pdf
Copyright : domaine public
Identifier : ark:/12148/bpt6k70861t
Source : Bibliothèque nationale de France, Rés. Ye-3535
Relation : http://catalogue.bnf.fr/ark:/12148/cb37367035f/description
Provenance : bnf.fr
Theme : French & related literatures

The text below has been produced using a process called optical character recognition (O.C.R.). Since it is an automatic process, it is subject to errors you might find in this page.
The recognition rate for this document is 94.63 %.

LibreOffice
The Document Foundation

Les Fleurs du Mal: EDM



CC BY SA

Current State of EDM

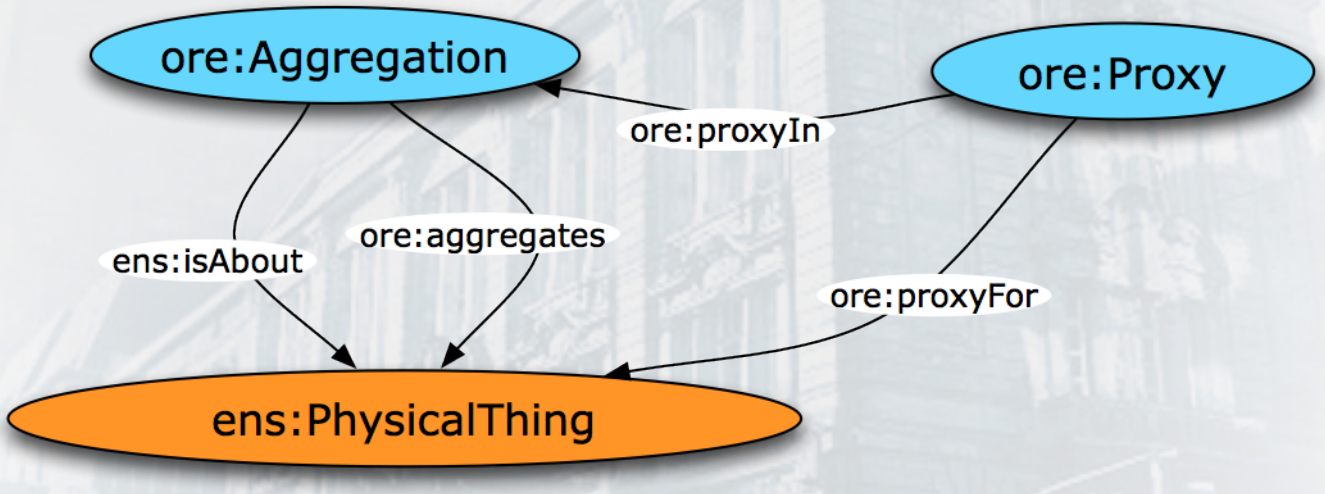
- Confirmed feasibility in community workshops (archives, libraries, audiovisual archives, museums)
- EDM is already deployed in the Europeana Linked open Data pilot at data.europeana.eu
- EDM Specifications and Primer
  - http://www.pro.europeana.eu/web/guest/edm-documentation
- Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, Herbert van de Sompel: The Europeana Data Model. IFLA 2010 (Gothenburg).
  - http://www.ifla.org/files/hq/papers/ifla76/149-doerr-en.pdf
- Development documentation + Ontology @ EuropeanaLabs:
  - http://europeanalabs.eu/

CC BY SA

… and LoD: into contextualisation

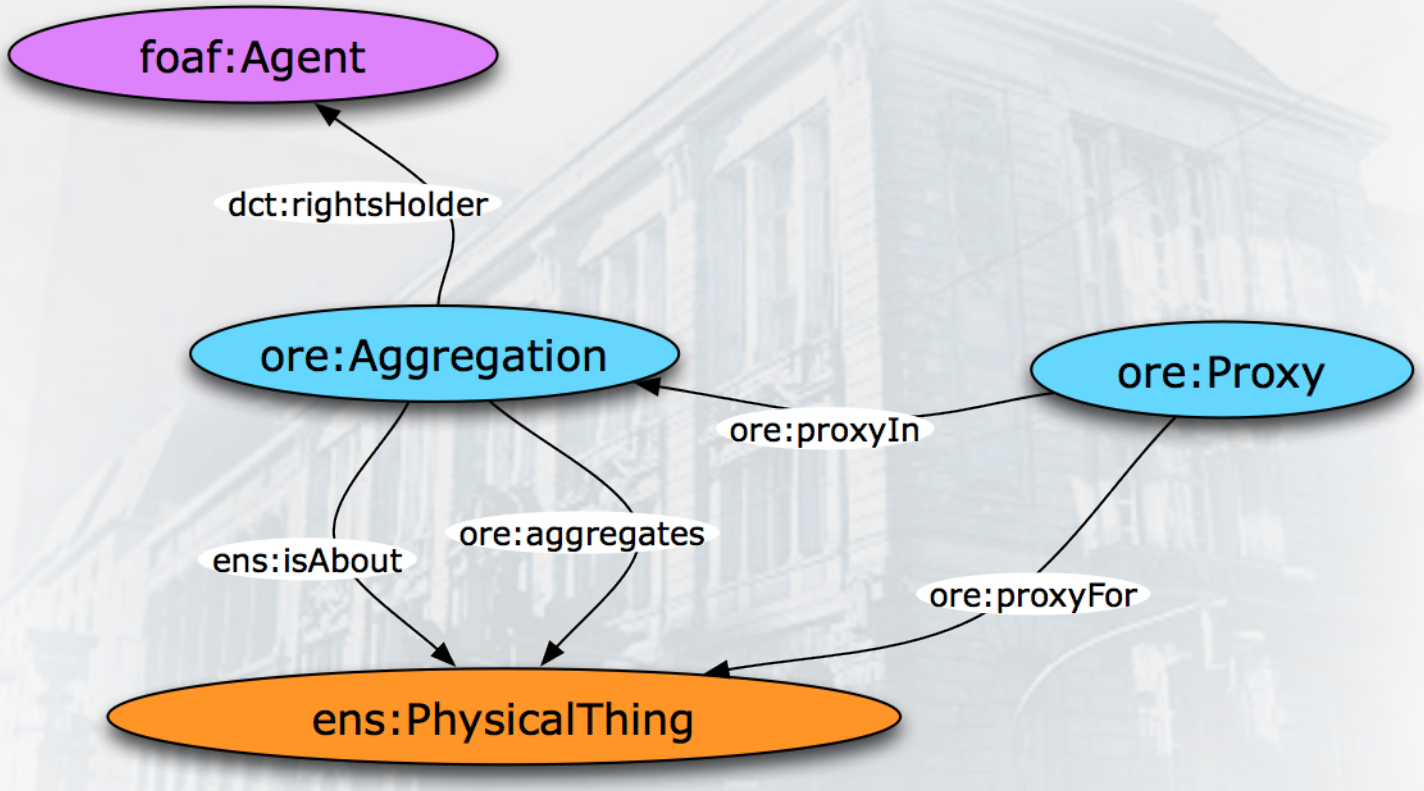
CC BY SA

An Aggregation ...



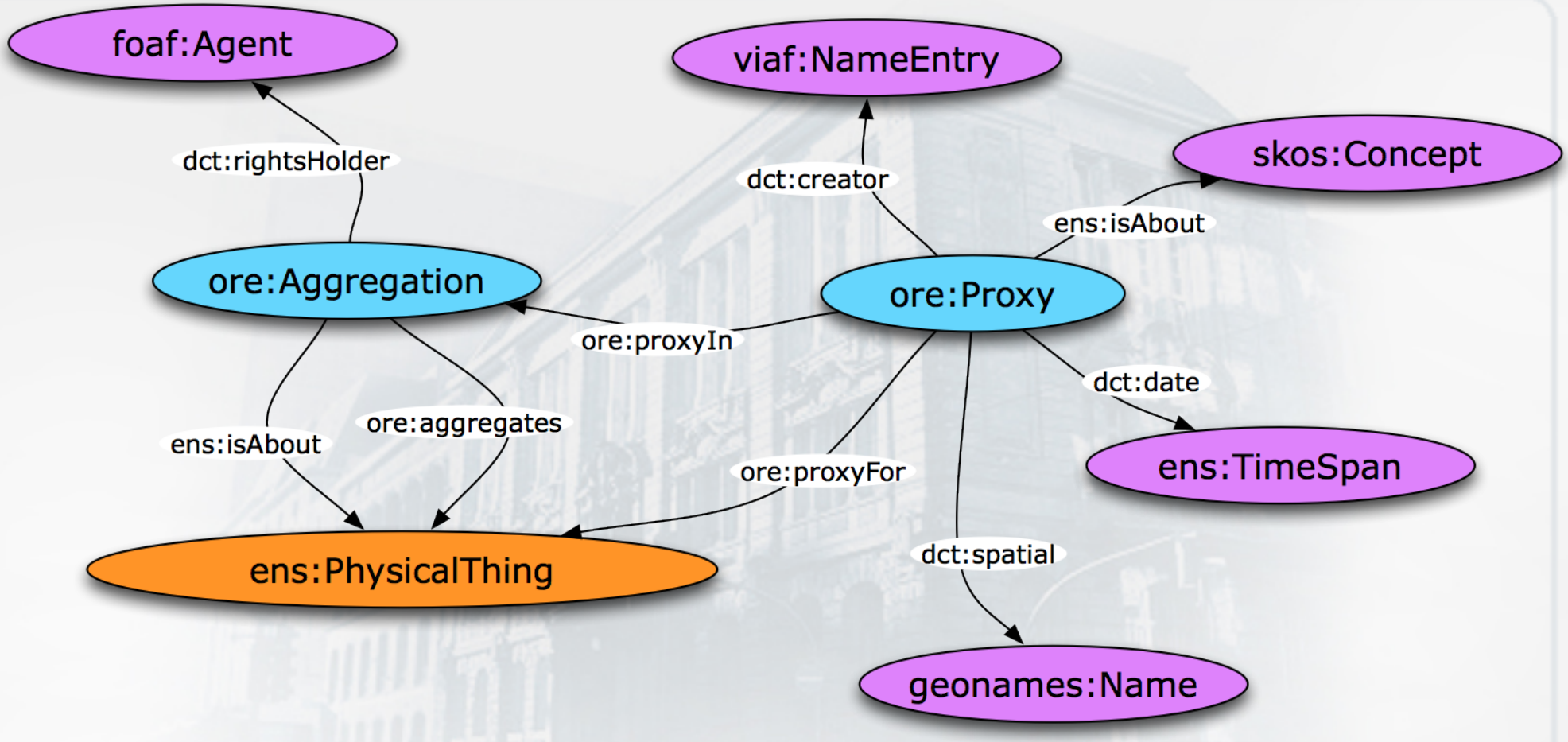
BY SA

… some context



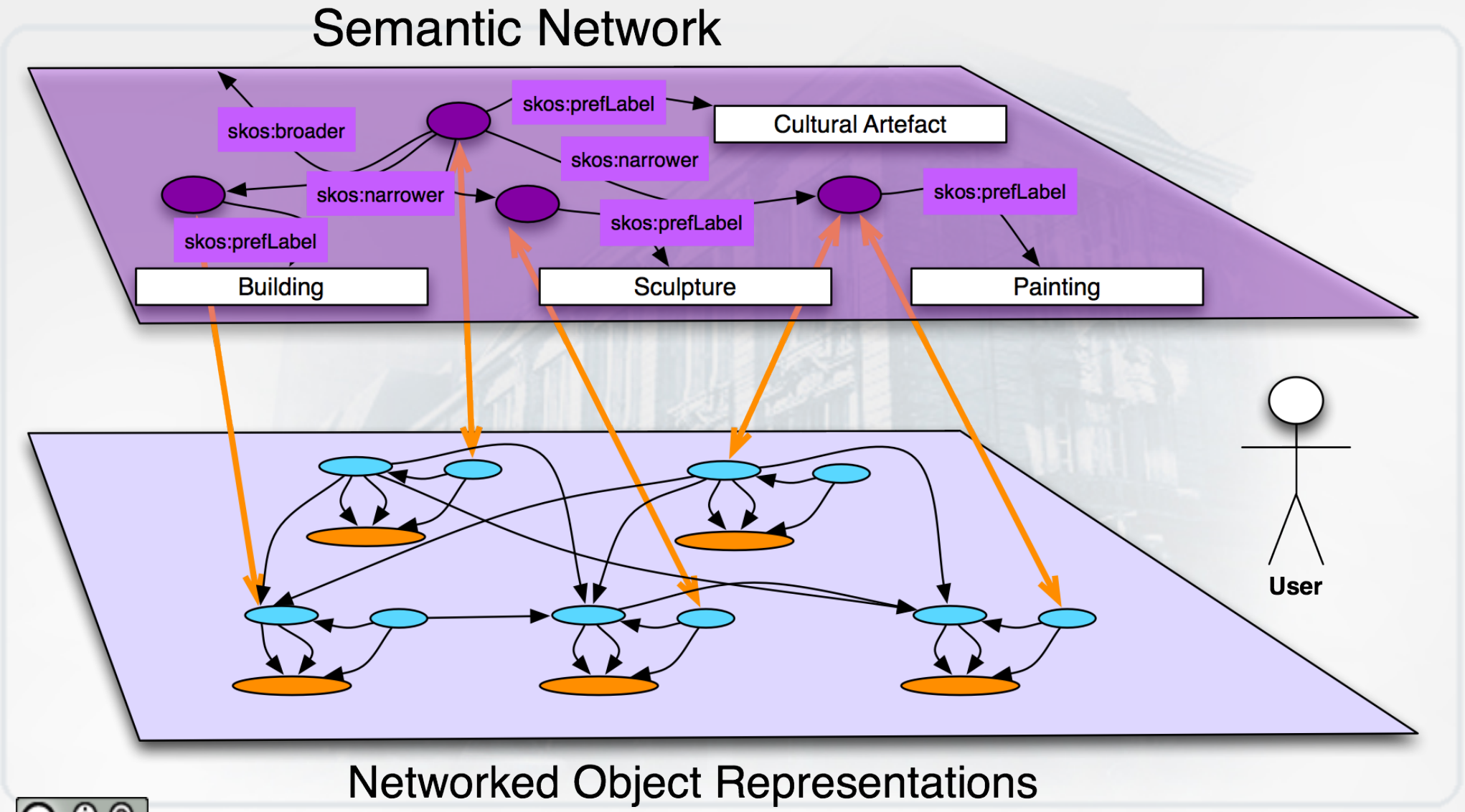
LibreOffice
The Document Foundation

… more context



LibreOffice
The Document Foundation

… and the Big Picture: The Semantic Data Layer

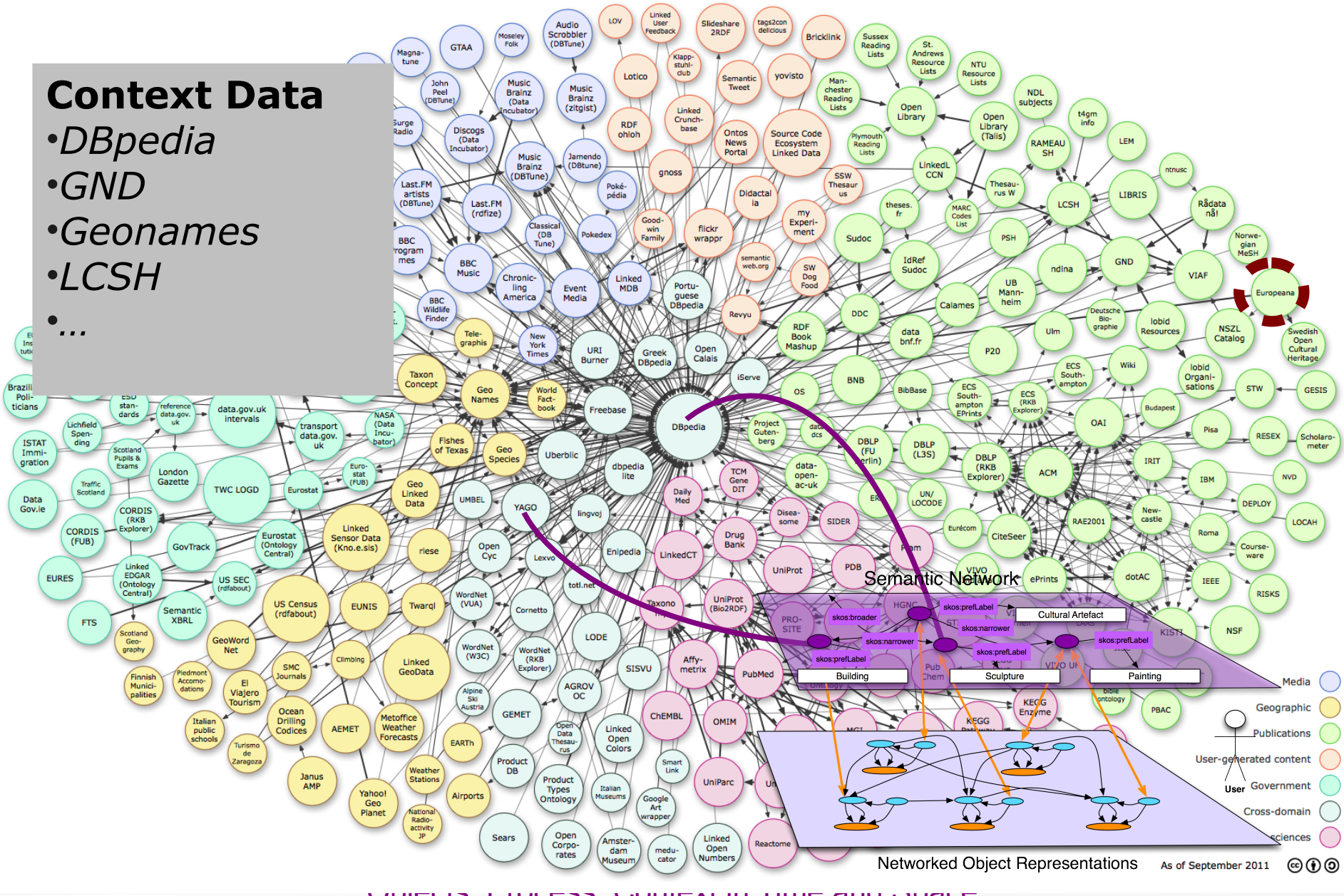


LibreOffice
The Document Foundation

EDM and Linked Open Data

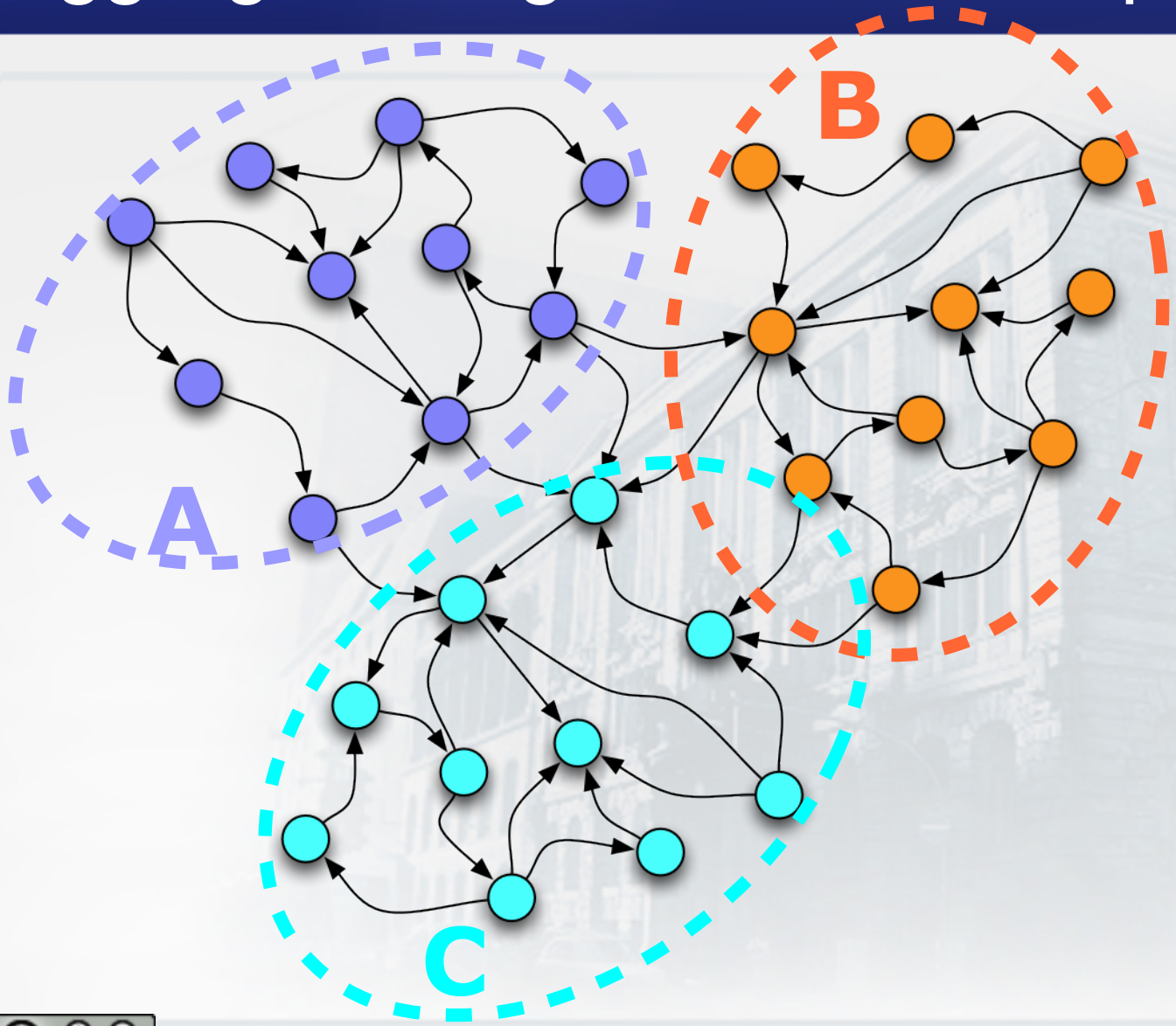
Context Data

- *DBpedia*
- *GND*
- *Geonames*
- *LCSH*
- *...*



Objects, Process, Context in Time and Space

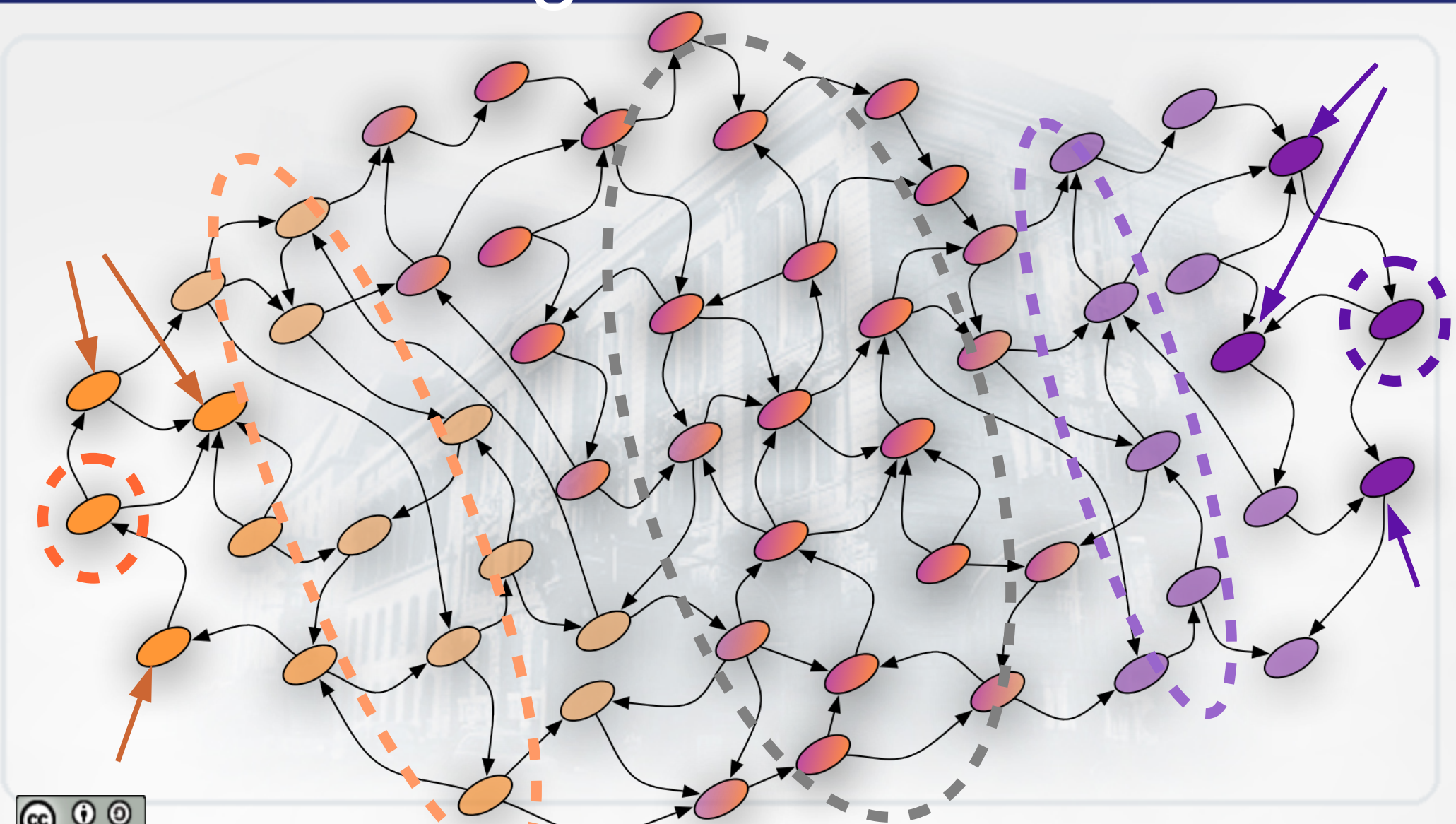
Modelling Object Representations as RDF Aggregations generates new questions ...



- 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???

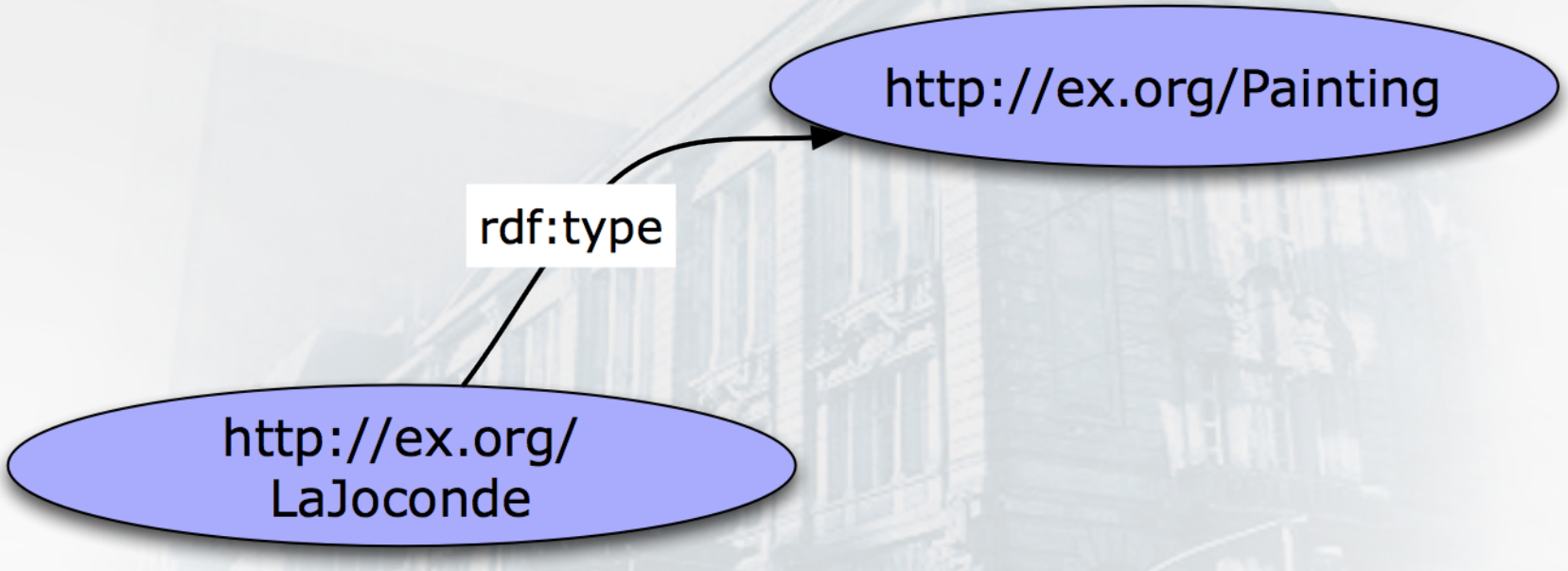
BY SA

Aggregations and Context: Calculating Closeness



LibreOffice
The Document Foundation

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

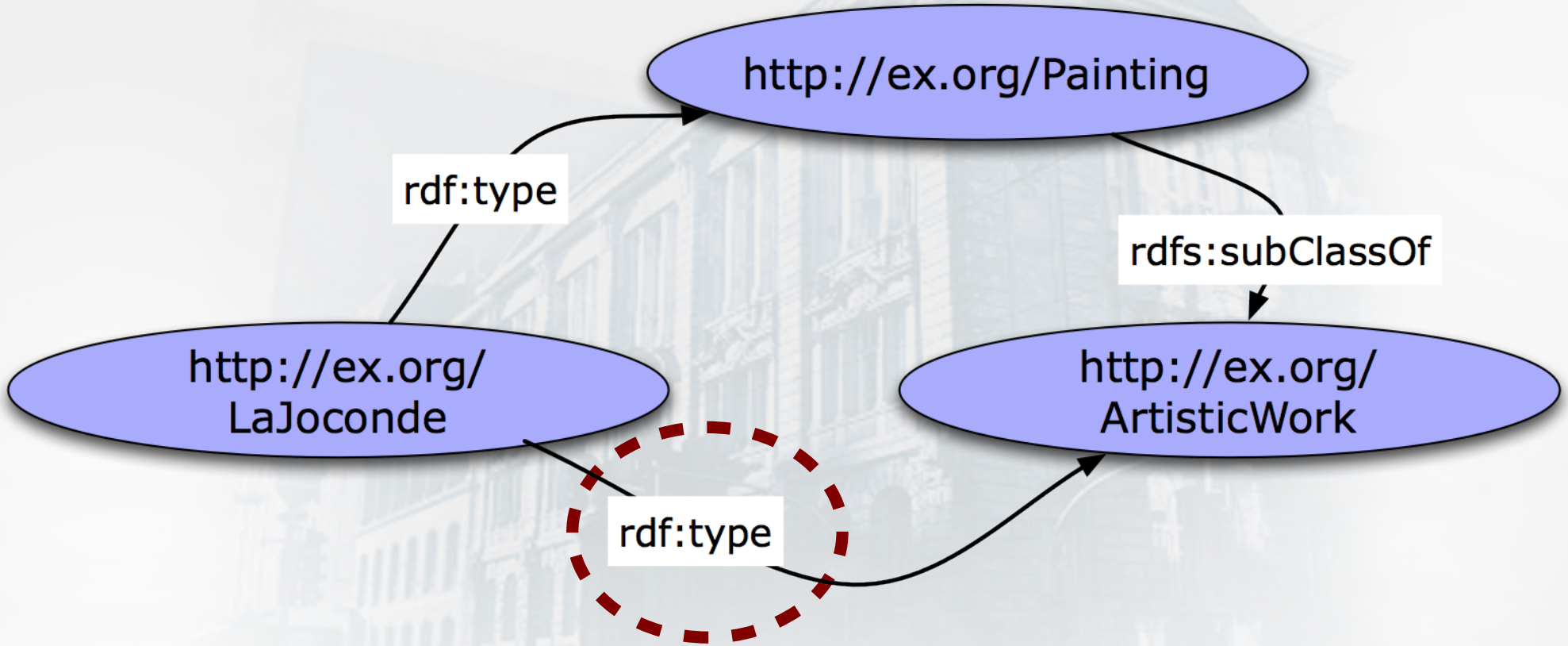


LibreOffice
The Document Foundation

Triple Sets and Reasoning (2)

LibreOffice
The Document Foundation

Triple Sets and Reasoning (3)



→ simple, deterministic 'reasoning' may have some potential of enabling novel digital heuristics!

BY SA

… and shares the more general pros and cons of RDF

- **Pros**: Simple, lightweight, robust, atomistic and scalable
- Limitations of triple syntax and problems of expressing
  - provenance
  - versioning
- Hence W3C work on 'Named Graphs' to **avoid systematic reification**
- How to deal with '**similarity**' and '**ambiguity**'?
- Lacking support for complex signification modes **beyond mere denotation**
- … and the **data quality issues** of the LoD cloud

BY SA

→ **WP3 experiment**

LibreOffice
The Document Foundation

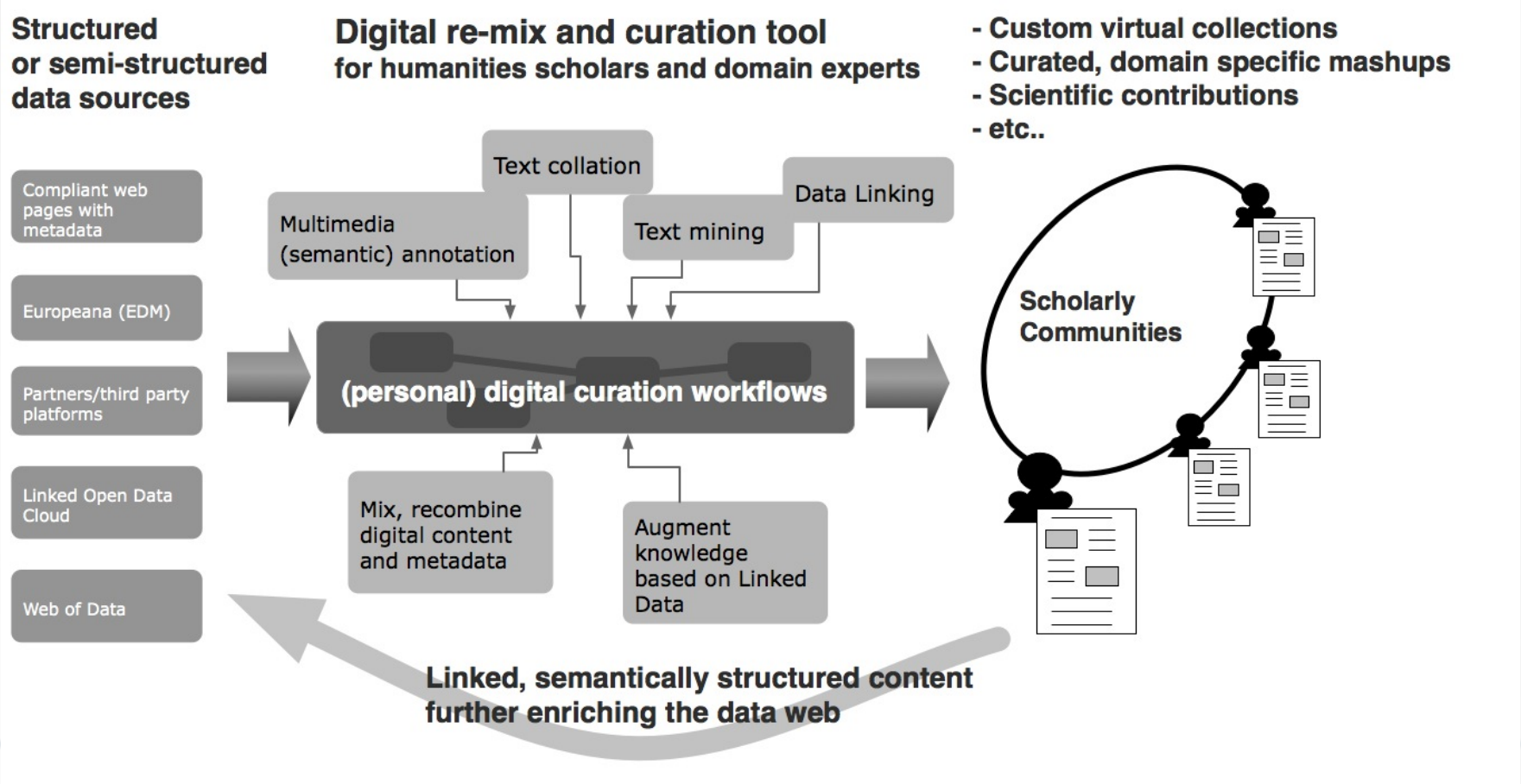
Digitised Manuscripts to Europeana (DM2E): What?

- 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 and 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.

cc BY SA

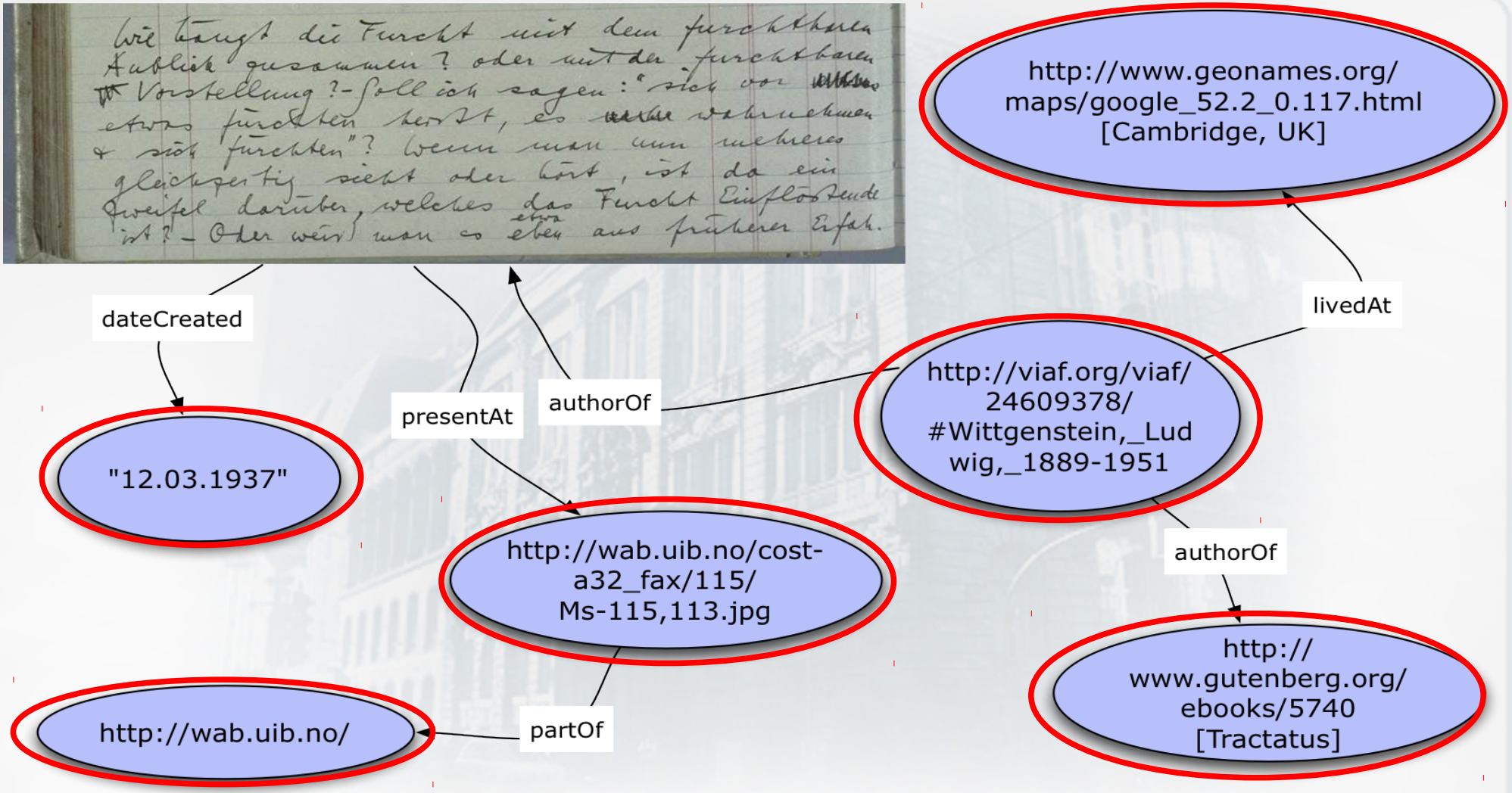
WP3: Digital Humanities Related Engineering

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



BY SA

Contextualisation



CC BY SA

Digital Humanities Functionality (WP3)

- Can we 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 what is the use of the resulting increasingly complex **social semantic scholarly graph** that extends the EDM data with RDF statements such as
  - Version_A – isSuccessorOf – Version_B
  - Scribe_Y – copiedFrom – Scribe_Z
  - Statement_1 – contradicts - Statement_2
- → what do you **obtain** from on inferencing on this graph …
- → and which are the **limitations** of such an approach?

CC BY SA

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, Available online 13 July 2011, http://www.sciencedirect.com/science/article/pii/S0167739X11001178

LibreOffice The Document Foundation

Thank you!

BY SA