The Europeana Data Model: tackling interoperability via modelling

Carlo Meghini, Antoine Isaac, Stefan Gradmann, Guus Schreiber, et al.

DL.org Autumn School

Athens, October 5, 2010

Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

Outline

- **Part I**
  - Background
  - Requirements
  - Status
- **Part II**
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

Europeana: the vision

'A digital library that is a single, direct and multilingual access point to the European cultural heritage.'

European Parliament, 27 September 2007

'A unique resource for Europe's distributed cultural heritage… ensuring a common access to Europe's libraries, archives and museums.'

Horst Forster, Director, Digital Content & Cognitive Systems Information Society Directorate, European Commission

Growing political engagement

- European Commission funding projects that promote interoperability of European information — Since the 1990s

- Google Print's library partnerships announced — December 2004

- Letter from 6 Heads of State to the President of the European Commission — April 2005

- Commission launches i2010, a 5 year digital-led strategy for growth and jobs — June 2005

- The Commission's Directorate for Information Society and Media launches Digital Libraries Initiative — September 2005

Gaining political momentum

- Commission recommendation to Member States   August 2006
  to create a European digital library

- Endorsement by the Council of Culture Ministers   November 2006
  representing all the Member States

- Outcome:
  - European digital library (EDL) Thematic Network   July 2007 to
  - 18-month project   March 2009
  - Funded by the digital libraries initiative under
    the eContent*plus* call
  - To create a prototype web portal: *Europeana*

The Commission's objectives for Europeana

- To create a multilingual public-domain access point to Europe's cultural and scientific heritage

- To use digitised cultural and scientific heritage resources as input for a wide range of information products and services

- To play a key role in the future growth of sectors such as learning and tourism

- To inspire new creative enterprise and innovation

- To promote understanding of our common European background and the sense of a European identity

Achieving political endorsement

- European Parliament votes to support a multilingual access point to Europe's common heritage — September 2007

- Commission issues Communication detailing each Members' progress on the digital libraries initiative — August 2008

- Europeana strategy briefing for policy advisors and digital strategists in all Ministries of Culture — October 2008

- Council of Culture Ministers meeting publishes Conclusions on the European digital library which express strong political support — 20 November 2008

europeana
think culture

Council conclusions, 20 November 2008

' Digitisation and online accessibility of cultural material are essential to highlight cultural heritage, to inspire the creation of new content and to encourage new online services to emerge. They help to democratise access to culture and knowledge and to develop the information society and the knowledge-based economy.'

europeana
think culture

Building Europeana

- Core Projects
  - Europeana version 1.0 (started March 2009)
  - Europeana:connect
- Many projects keep joining
  - Services (ASSETS, ARROW)
  - Content (EFG, Judaica, Athena, …)
- Releases:
  - Rhine (July 2010)
  - Danube (April 2011)
- Future: Europeana version 2.0
- …

The EDM context

- Why
  - to define what information is necessary in order to enable the functionality of Europeana
- What
  - Classes, arranged in a taxonomy
  - Properties, arranged in a taxonomy
  - Constraints: domain/range, cardinality of properties
- Who
  - The Europeana experts
- When
  - July 2010, Danube specs

Outline

- **Part I**
  - Background
  - Requirements
  - Status
- **Part II**
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

Requirements

- Data integration
- Support rich functionality (*e.g.*, semantic search)
- Optimize the use of resources in time

Requirements: Data integration

- Standard approach in a sound software development process:
  - Requirement
    - Collection
    - Specification
  - Design
    - Analysis of the functionality
    - Algorithms
    - Required data
  - Implementation
  - Testing
  - Validation

Requirements: Data integration

- Europeana is a data integration system
  - A living organism, consisting of
    - Central Repository
    - Local Sources
  - In continuous expansion:
    - More data coming from the local sources → **Consistency, data scalability**
    - More sources being added → **Extensibility in data model**
    - More users → **Workload scalability**
    - More functionality → **Extensibility in function**

europeana
think culture

Requirements: Data integration

- A data integration system is built by taking into account the data models of the sources
  - At requirement collection time: collect the model of each source
  - At design time:
    - How to integrate the existing data in order to achieve the required functionality
    - May lead to: revision of requirements or addition of extra functionality
- In the present case, the sources are:
  - Large and important = lots of data, users, expectations
  - In different domains = significantly different data models
  - Very many = lots of significantly different data models
  - An open set = **who knows what data may come tomorrow**

Requirements: Data integration

- Two possible venues for data modeling:
  - Cross-domain element set
    - a common set of properties capturing features shared by all objects, e.g. the Dublin Core Element Set
  - An ontology
    - a **complete** conceptualization, emphasizing the fundamental notions around Cultural Heritage Objects that allows Europeana to accommodate the data coming from providers **regardless of the original models**
- Cross-domain venue: Rhine, set up the basic infrastructure
  - Europeana Semantic Elements
- What about Danube?

Requirements: Support rich functionality

- Europeana must outdo the competition in the Cultural Heritage domain, notably web search engines
  - richness: collect **all the data there is**
  - intelligence: connect data to Knowledge Organization Systems
  - coverage: multilingualism
  - For Danube, we need to go the ontology venue in order to support rich functionality
  - richness: a special ontological entity to represent aggregates
  - intelligence: classes to represent knowledge and properties to connect knowledge to objects
  - coverage: multilingualism is core in Europeana (more on this later)

Requirements: Optimize resources

- Minimize and protect the investment required for accumulating knowledge:
    - Re-use existing models
        - ontology is a controversial area of philosophy
        - recently, the controversy has reached computer science
        - very recently, the controversy has reached Europeana too
    - Build on standards
        - Institutions are making their data and their Knowledge Organization Systems available in the Web, using URIs, RDF/S, SKOS, Linked Data, and more
        - Need to buy into the Web Architecture and standards
        - Europeana wants to follow institutions rather to push them

Requirements: wrap up

- In sum, the EDM must:
  - be a simple ontology for capturing all relevant aspects of Cultural Heritage Objects
    - integrate the providers' data
    - support rich functionality
  - offer a structure for collecting data from contributors
  - re-use existing ontology and models
  - buy into the Web architecture and models
- Not obvious at first, a result
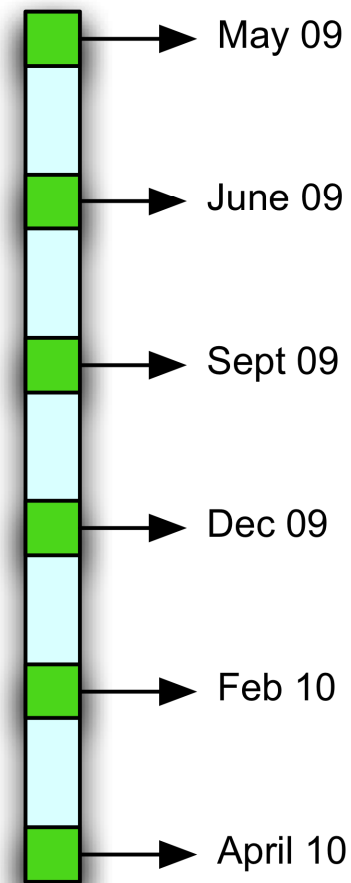
Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

EDM development

- version 1: initial surrogate model with rich set of contextualization properties
- version 2: OAI-ORE aggregations and SKOS concepts
  - release 2: 1st Europeana plenary
- version 3
- version 4: IRW ontology
  - release 1: December 2009
  - release 2: February 2009
- version 5: integration of ESE, evaluation through domain meetings
  - release 1: April 2010
  - release 2: June 2010

May 09

June 09

Sept 09

Dec 09

Feb 10

April 10

Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

The general picture
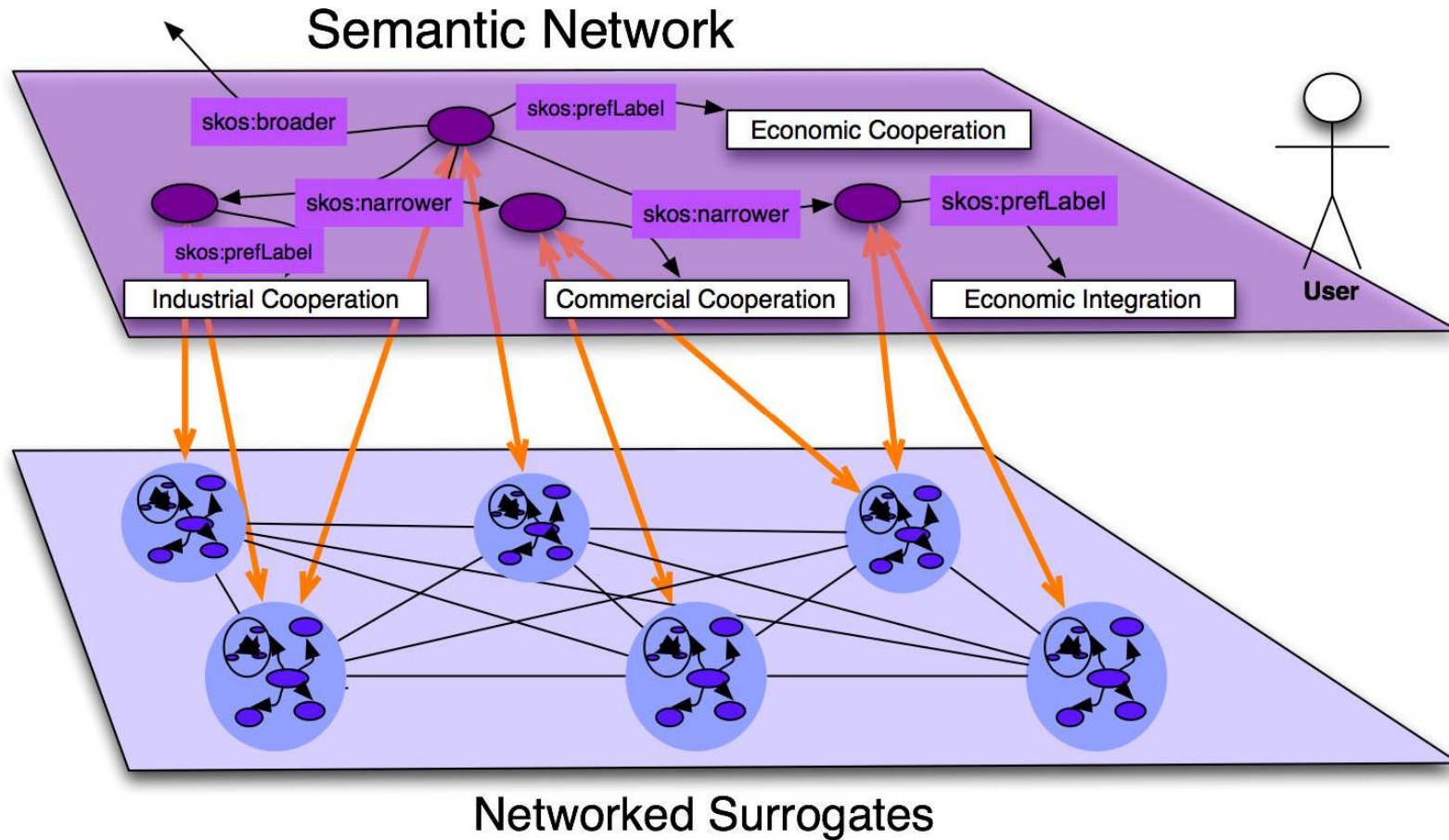
Semantic Network

skos:broader — skos:prefLabel → Economic Cooperation

skos:narrower → skos:narrower → skos:prefLabel

skos:prefLabel

Industrial Cooperation — Commercial Cooperation — Economic Integration

User

Networked Surrogates

europeana
think culture

Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

The class taxonomy
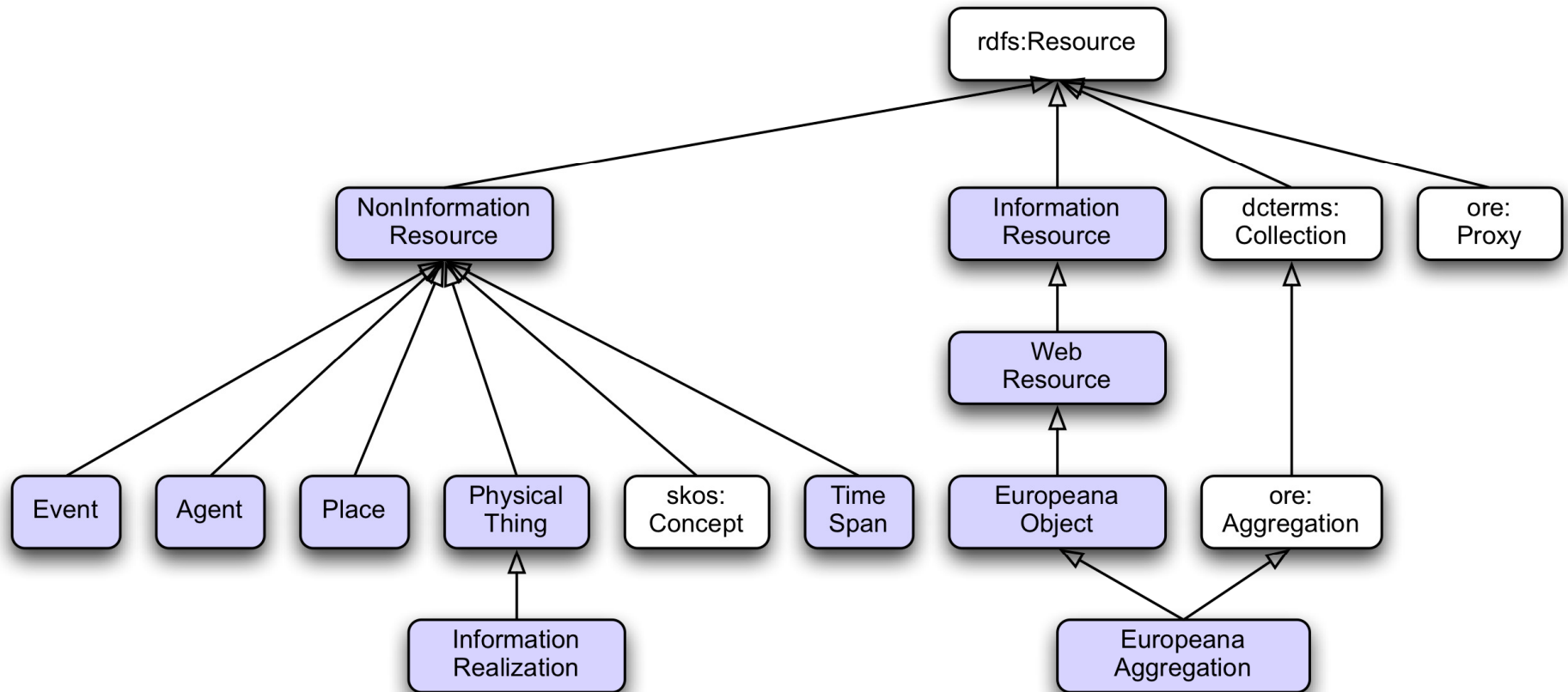
Europeana Aggregation

- The set of resources related to a single Cultural Heritage Object that collectively represent that object in Europeana.
  - all descriptions about the object that Europeana collects from (possibly different) content providers
    - including thumbnails and other abstractions
  - the description of the object that Europeana builds
- Every Cultural Heritage Object known to Europeana is represented by an instance of EuropeanaAggregation
- Every instance of EuropeanaAggregation represents a Cultural Heritage Object.

Europeana Object

- Any digital object on which Europeana has rights
  - Aggregations
  - Europeana content
    - Annotations (this class is the range of ens:hasAnnotation)
    - Deliverable of one the Europeana projects
  - Any content provider's object on which Europeana has acquired some right
    - A thumbnail of the painting Mona Lisa owned by the Louvre and offered to Europeana as an illustration of the painting, along with some rights (e.g., display)
    - A digitization of a photograph of the first page of issue number 56 of the title "Le Temps"
    - The text of the first page of issue number 56 of the title "Le Temps" s

Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

Property taxonomy (without ESE properties)

DL.org Autumn School
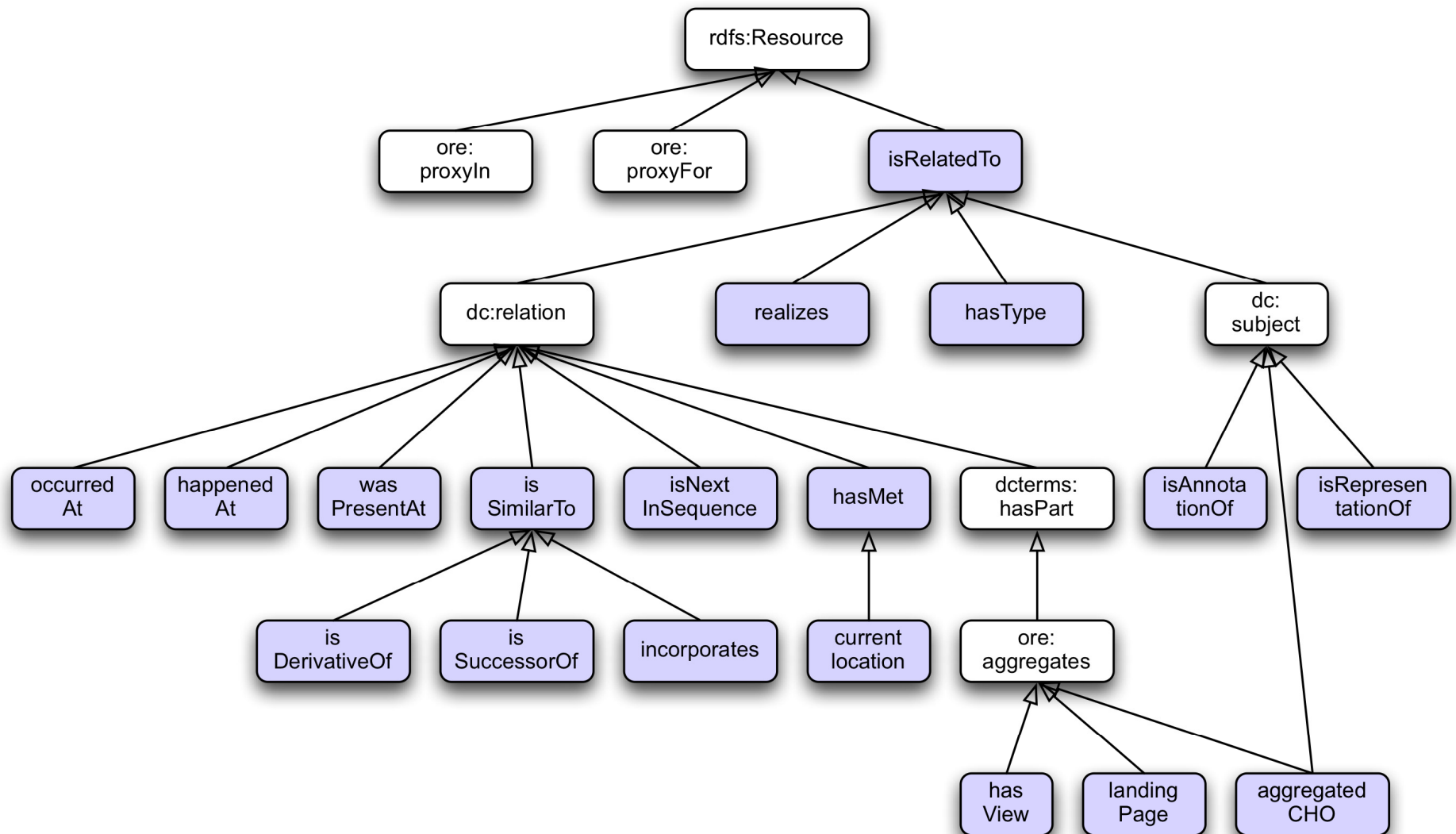
The Example - 1

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

The Example - 2

© Musée du Louvre/A. Dequier - M. Bard

Denon
1 e étage
Salle de la Joconde
Salle 6

détail de la bouche
détail des mains
détail des yeux
détail du buste

Leonardo di ser Piero DA VINCI, dit Léonard de Vinci

Vinci, 1452 - Amboise, 1519

Portrait de Lisa Gherardini, épouse de Francesco del Giocondo, dite *Monna Lisa*, la *Gioconda* ou la *Joconde*
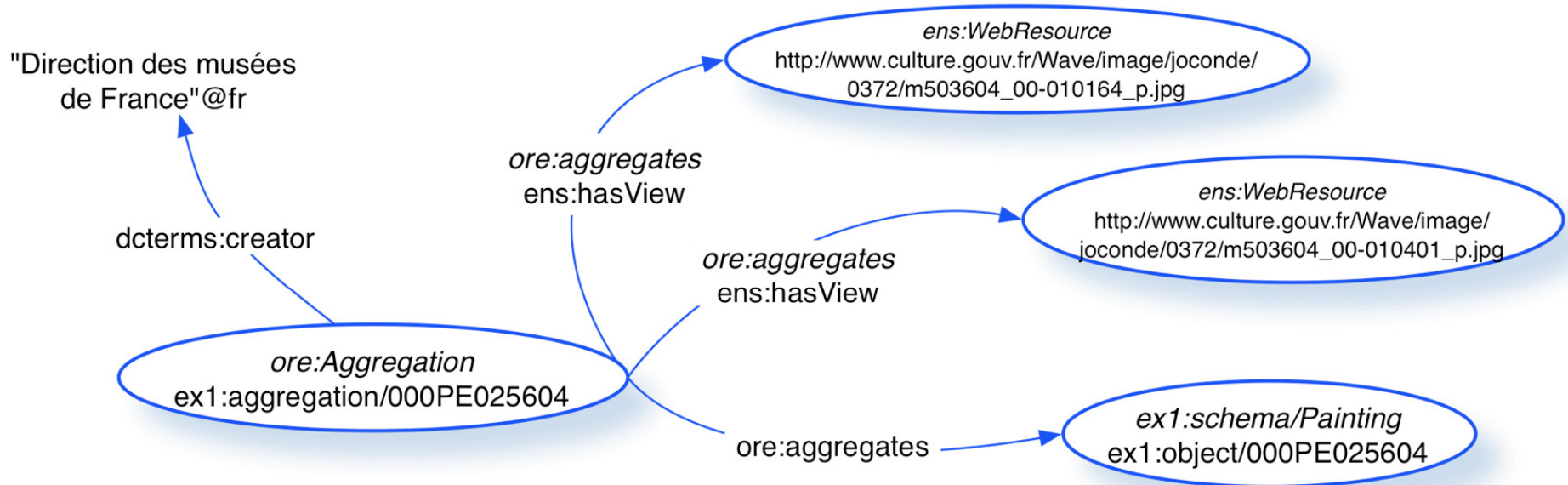
Vers 1503 - 1506
Peint à Florence
Bois (peuplier)
H. : 0,77 m. ; L. : 0,53 m.

Acquis par François Ier en 1518
Département des Peintures

INV. 779

europeana
think culture

Providing an aggregation of digital resources for a cultural object

*RDF graph with specific conventions for resource types and sub-properties*

Modeling Mona Lisa

- There's a resource that stands for Mona Lisa as an object in Museum

> *ex1:schema/Painting*
> ex1:object/000PE025604

- Ideally identified by a URI assigned by Direction des musées de France
- classified using a DMF ontology
- But DMF has a specific description for that object
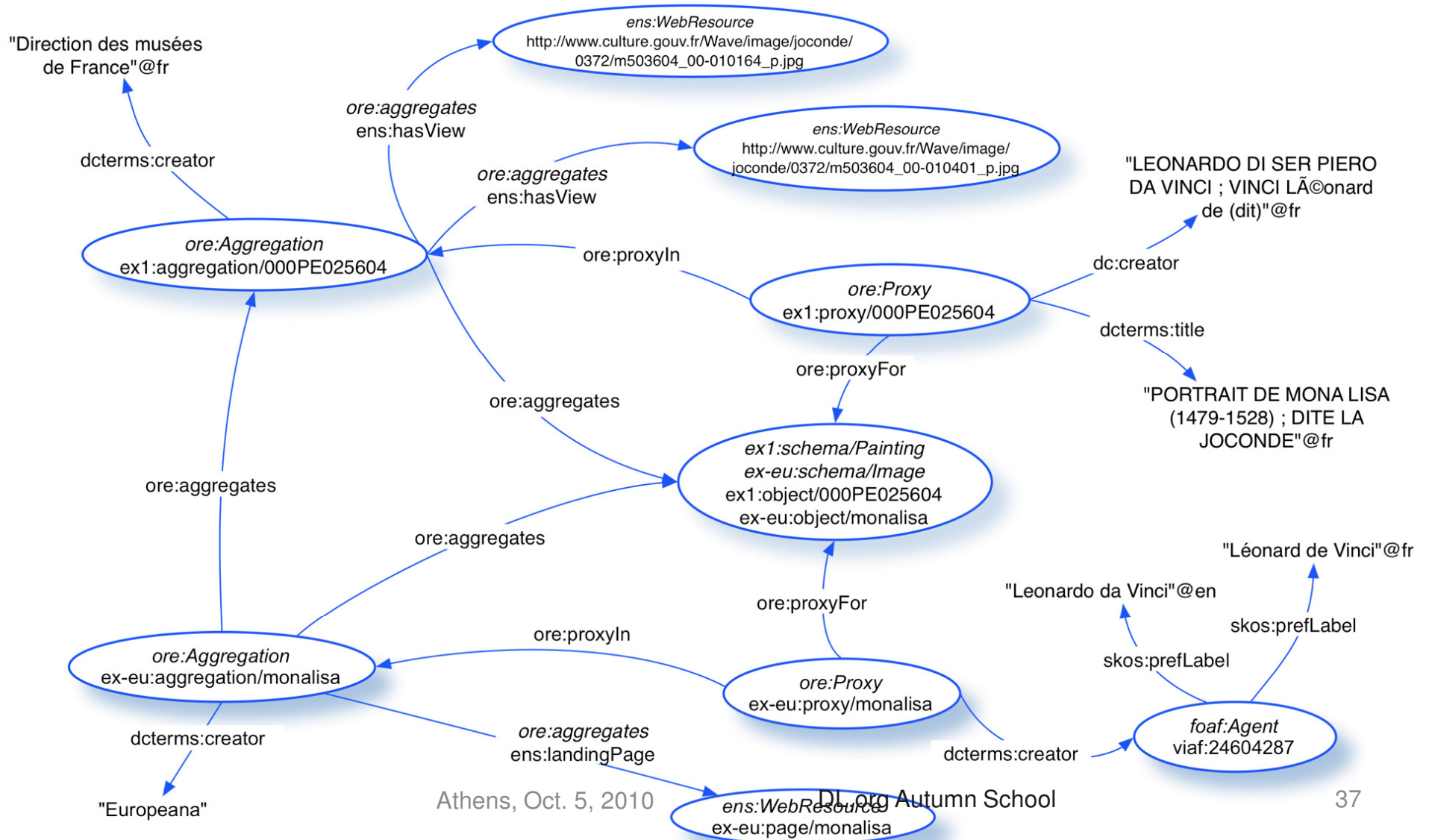 - Other institutions might have a different one!

So we create Proxies
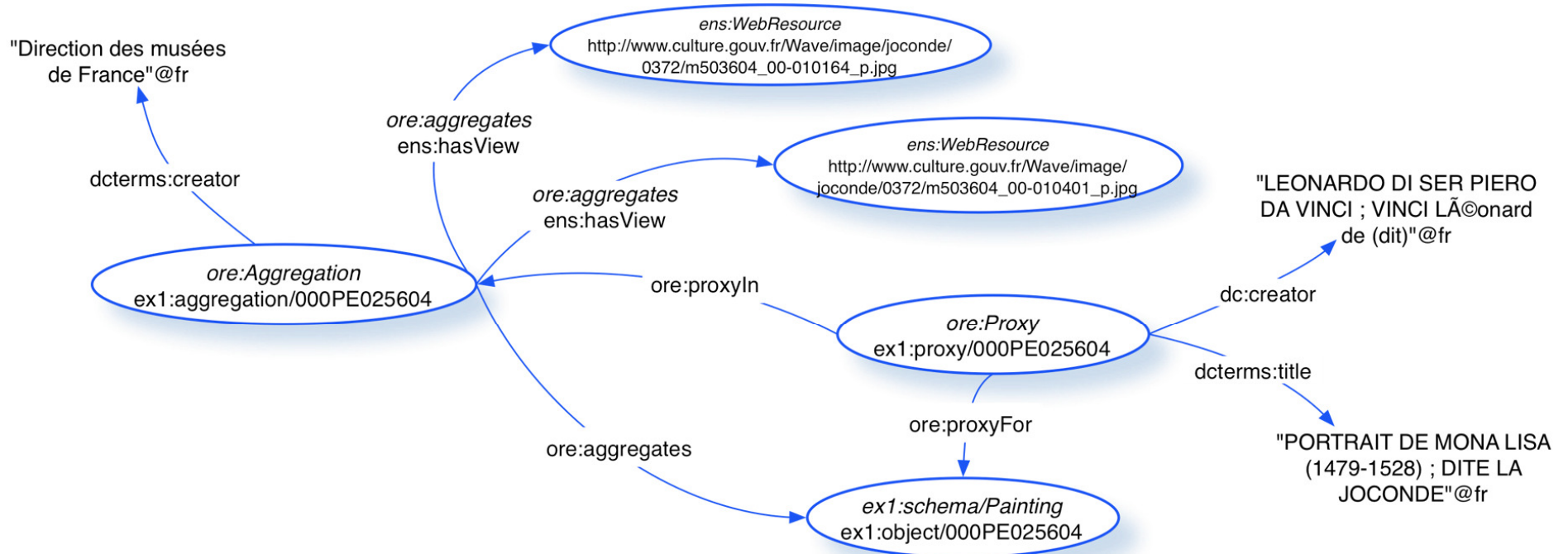
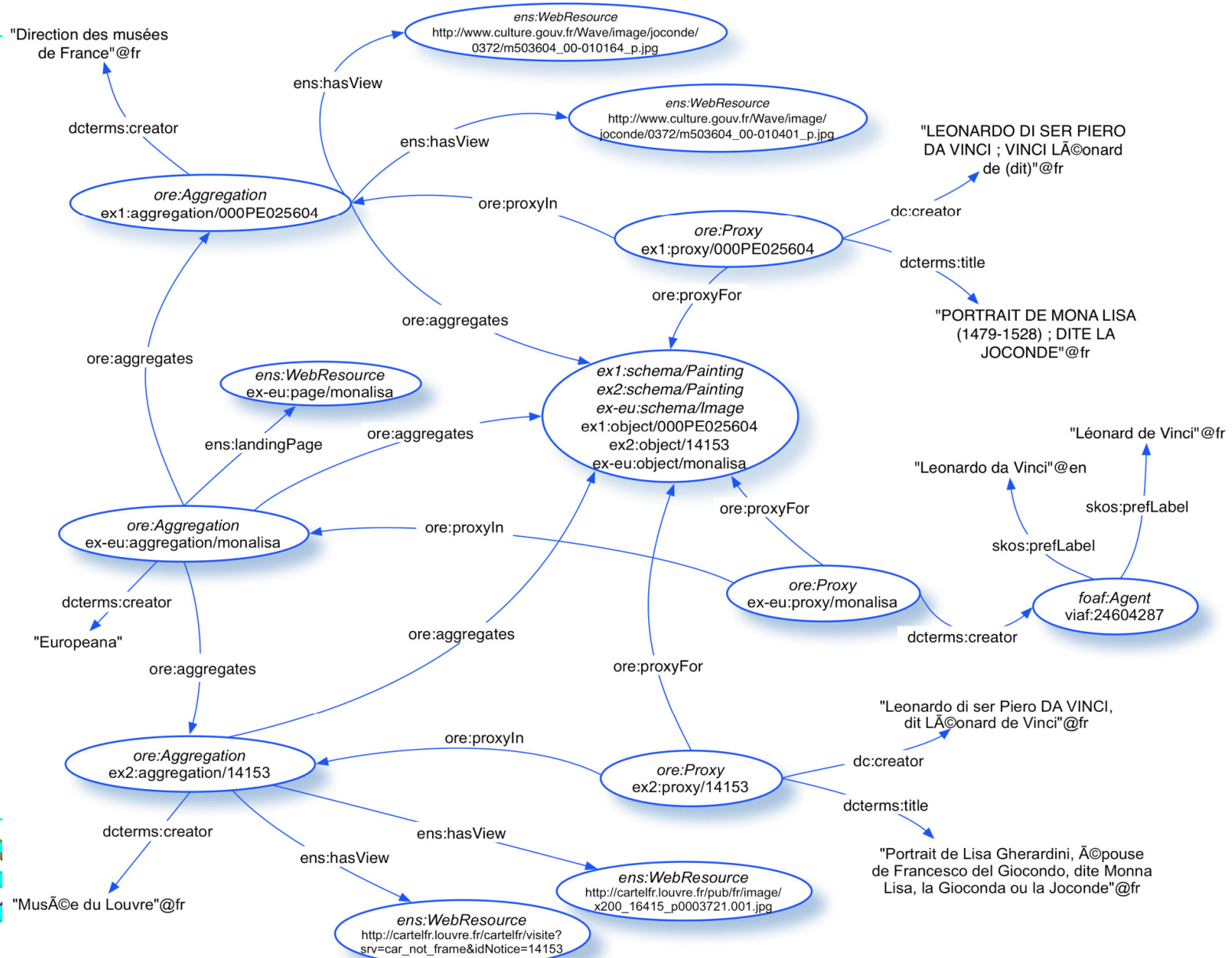# Becomes really handy when there are several records for a same object
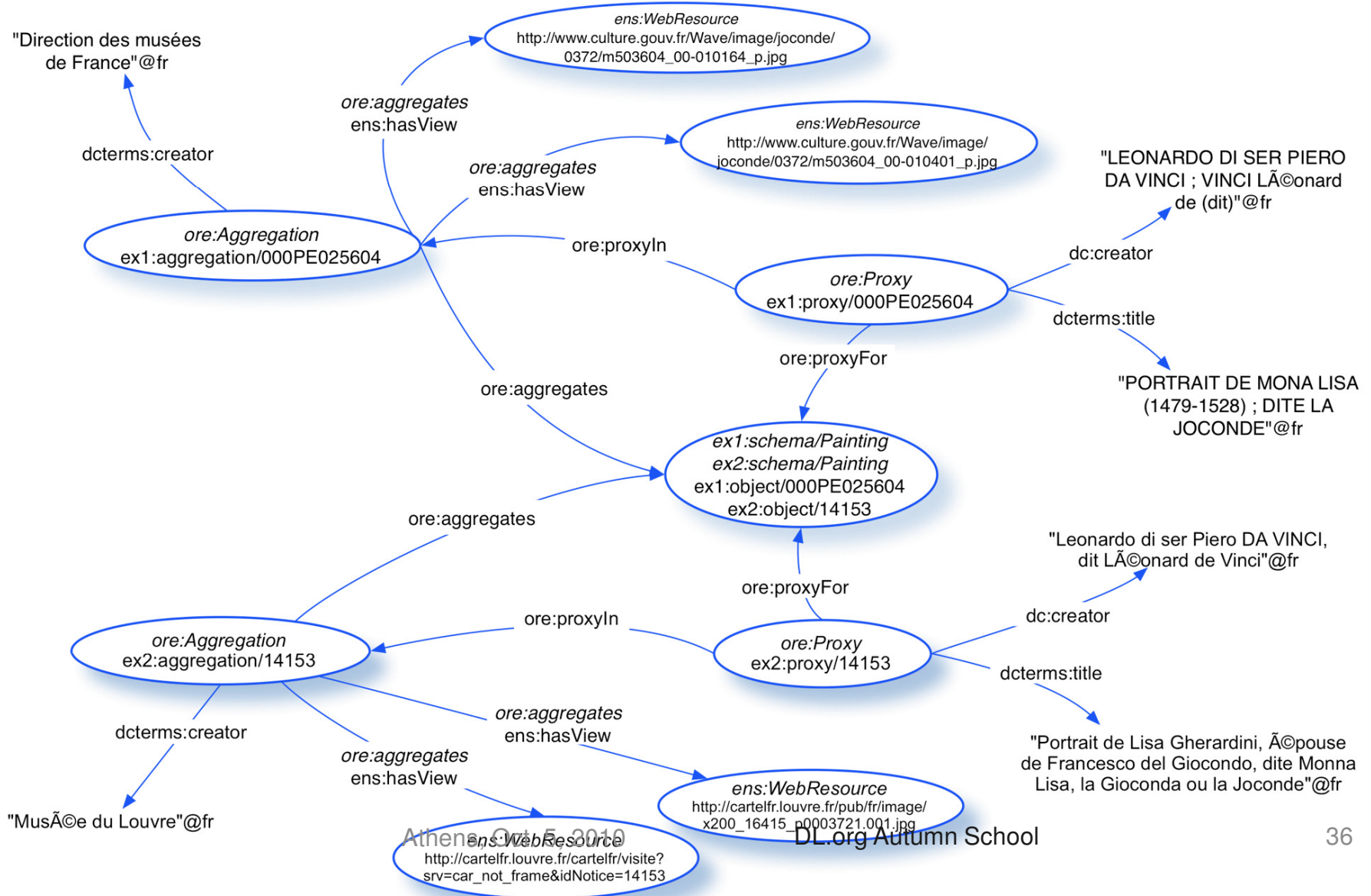
And there are always several information providers: think of Europeana!

Back to proxies: don't look at that!

Events

- EDM supports:
    - simple object-centric approaches, typical in libraries
    - more sophisticated event-aware approaches, typical in museums
- In fact, museum objects often come with complex descriptions

Amphora of Tuthmosis III

Identifier: Λ2409

Classification: Amphora

Event: Type: Excavation

> **Agent:** Stylianos Alexiou
>
> **Date:** 1951, October
>
> **Place:** Katsampas, Tomb of the "blue coffin", Heraklion

Event: Type: Deposition

> **Place:** Katsampas, Tomb of the "blue coffin", Heraklion
>
> **Period:** LMIII A1 (14th century BC)

Event: Type: Production

> **Place:** Egypt
>
> **Period:** 18th Dynasty, reign of Tuthmosis III (15th century BC)

Current Location: Archaeological Museum of Heraklion Crete

**Current Owner:** Archaeological Museum of Heraklion Crete

Description: Intact, veined, Egyptian alabaster jar. It has a piriform body, short neck, flat everted rim, foot of biconcave profile, defined by a ring with hollow underside, imitating a slightly asymmetrical base. Two vertical strap handles separate the shoulder from the top of the belly. On one side of the belly is a rectangular frame enclosing a hieroglyphic inscription with the name of Tuthmosis in two cartouches. The inscription reads:

Amphora of Tuthmosis III

europeana
think culture
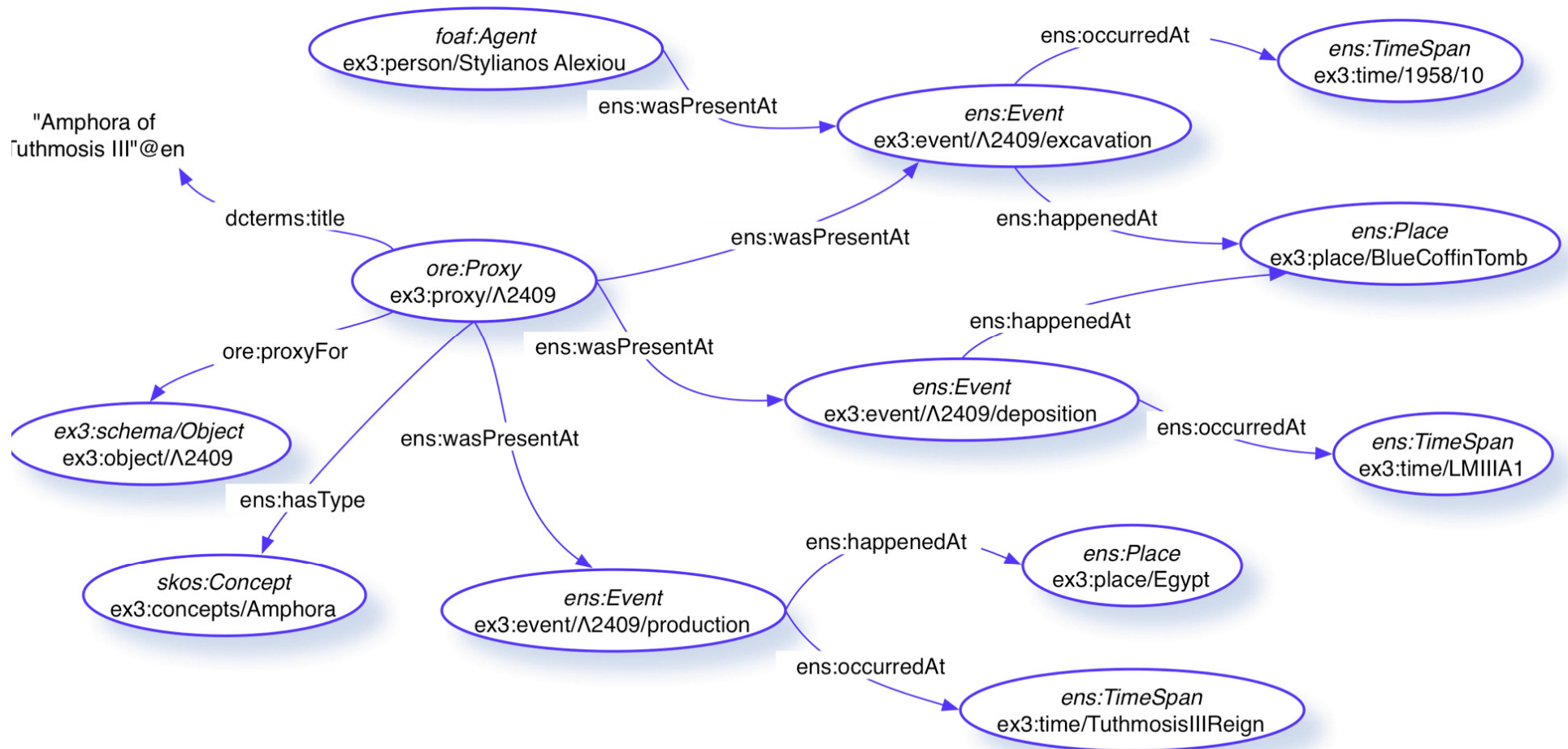
Enabling event-aware descriptions:
Was Present At, Happened At, Occurred At
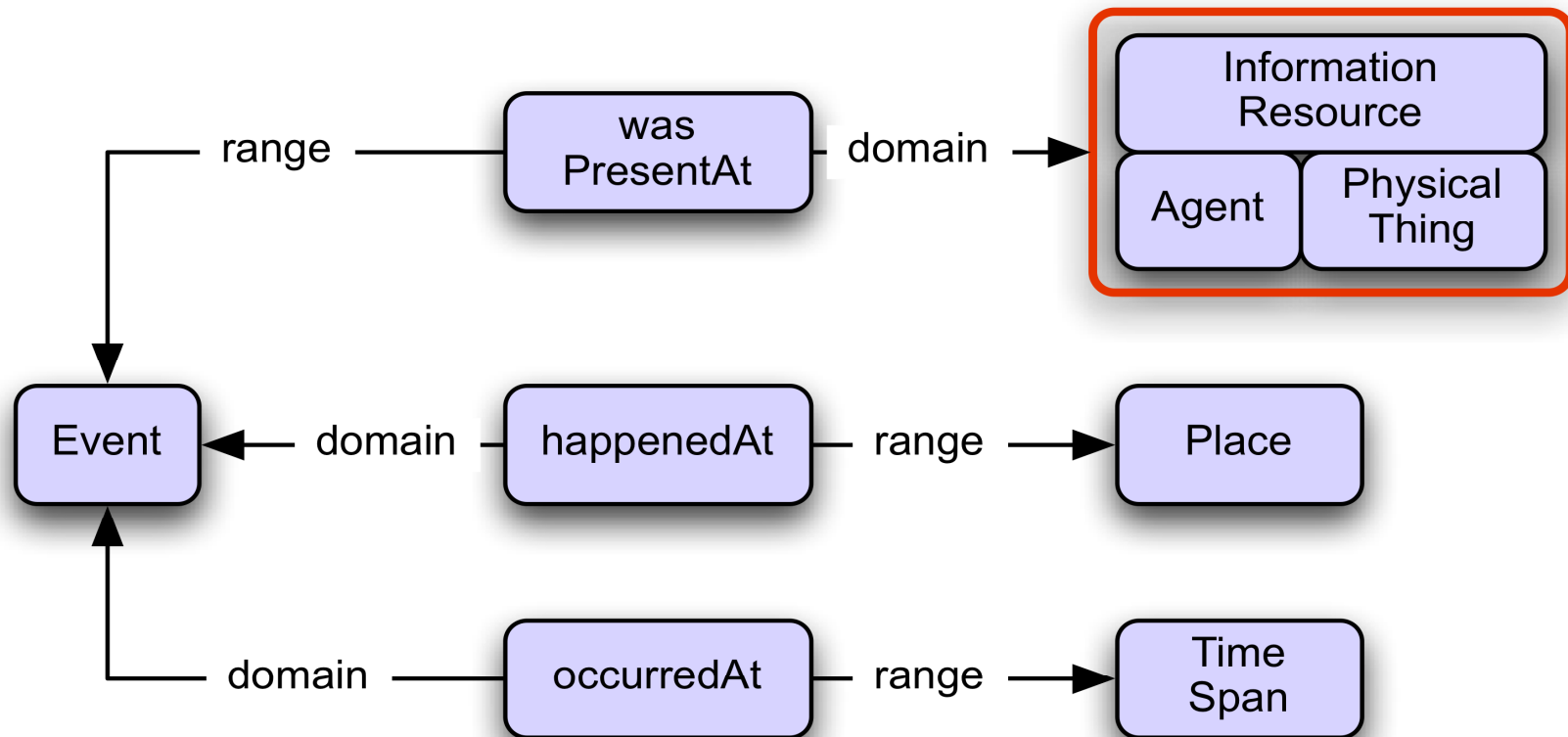
Interoperability at the value level

- EDM offers high-level classes and properties for integrating (via specialization) classes and properties of existing models.

- What about values?

- Europeana collects metadata with values:
  - in many different languages
  - drawn from many different vocabularies
  - drawn from no vocabulary at all

- What to do? Enrich!

Initial metadata values

"PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE"@fr

"LEONARDO DI SER PIERO DA VINCI ; VINCI LÃ©onard de (dit)"@fr

"1503, entre"@fr

"1506 et"@fr

"bois"@fr

"François Ier"@fr

"femme"@fr

ens:hasMet
dcterms:created

ens:hasMet
dcterms:created

ens:hasType
dc:medium

ens:hasMet

ens:hasMet
dc:creator

dcterms:title

ore:Proxy
ex1:proxy/000PE025604

ore:proxyFor

ens:isAbout
dc:subject

ex1:schema/Painting
ex1:object/000PE025604

europeana
think culture

Enriched object

"Leonardo da Vinci"@en

"Léonard de Vinci"@fr

skos:prefLabel   skos:prefLabel

"PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE"@fr

foaf:Agent
ex1:person/leonardo

ens:TimeSpan
ex1:time/1503-1506

dcterms:title

dcterms:creator

dcterms:created

"bois"@fr   "wood"@en

ore:Proxy
ex1:proxy/000PE025604

skos:prefLabel   skos:prefLabel

dcterms:medium

ore:proxyFor

dcterms:subject

skos:prefLabel

skos:Concept
ex1:concept/bois

ex1:schema/Painting
ex1:object/000PE025604

skos:Concept
ex1:concept/femme

"femme"@fr

skos:prefLabel

"woman"@en

skos:broader

skos:Concept
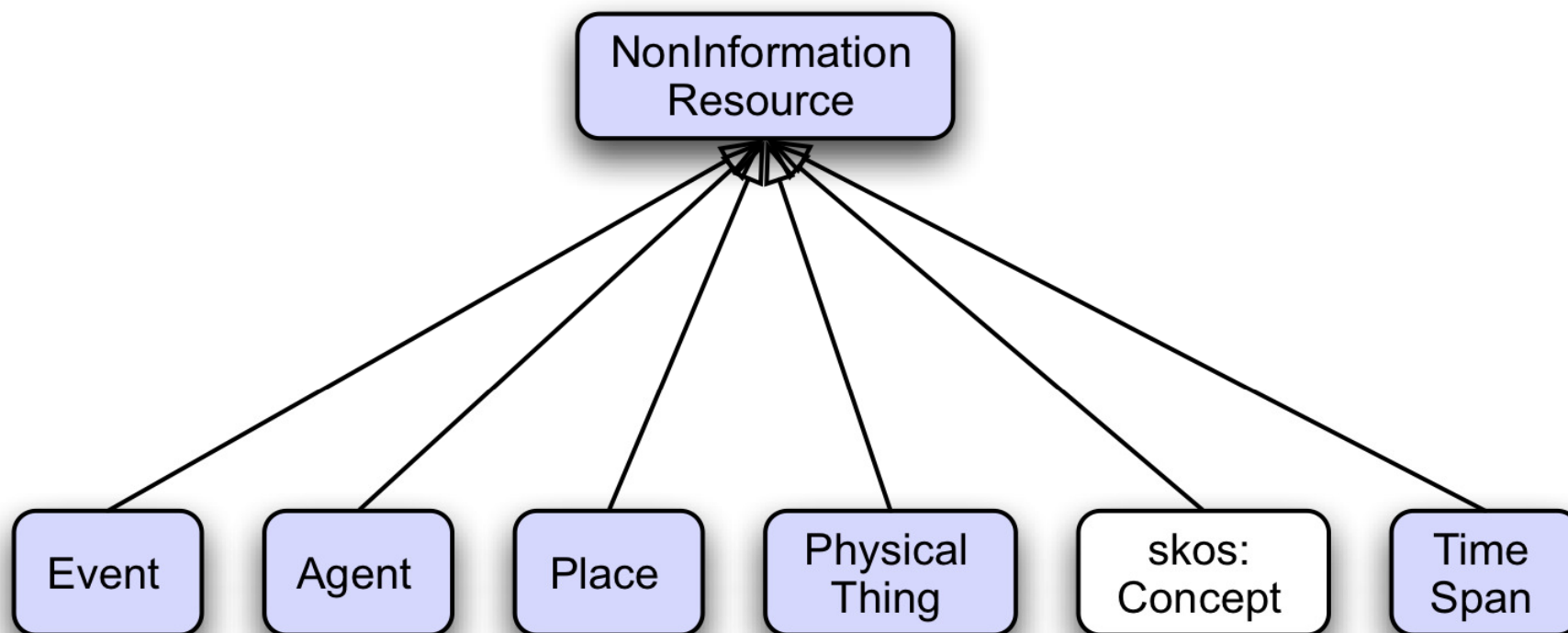ex1:concept/humain

skos:prefLabel

"humain"@fr

"human being"@en

skos:prefLabel

Reminder: classes for context entities

Who, what, when, where

More information

- EDM Specification
- EDM Primer

Outline

- Part I
  - Background
  - Requirements
  - Status
- Part II
  - The general picture
  - Classes
  - Properties
  - Examples
  - Future work

Future work

- Harmonize property values!
- Link to resources
  - Linked Data
- Evaluation:
  - Mapping real data to EDM
  - Functional check
  - Implementation

europeana
think culture

Thank you!

- Questions