

Linked Open Data & Europeana

Nicola Aloia, Cesare Concordia

cesare.concordia@isti.cnr.it



Europeana

- Started in 2007
 - Cluster of projects funded by EU
- 32,698,244 information objects (as of August 2014)
 - Paints, maps, drawings, photographs, music, books, newspapers, journals, diaries...
- 31 languages
- More than 2,300 data contributors
- 4m unique visitors came to the Europeana website in the 2013



Europeana





Search Subjects ▾ brasil

Search

Help

[Return to search results](#)

[< Previous](#) [Next >](#)



© Free access - no re-use

View item at
[Ateneu Barcelonès](#)

Share

Cite on Wikipedia

Translate details

Select language ▾

Powered by Microsoft® Translator

Subsidios para a organisação da carta physica do Brazil estudo geographico

Description:

44 p. : il., 1 map. ; 27 cm

Creator:

[Marcondes Homem de Mello, Francisco Ignacio, Barão Homem de Mello, \(1837-1918\)](#)

Date:

[1876]; [2011]

Type:

text

Subject:

[Geografia física-Brasil](#)

Relation:

<http://biblioteca.ateneubcn.org:8080/ateneu/ca/consulta/registro.cmd?id=5823366>

Language:

Source:

Publicat originalment a: Rio de Janeiro : Imperial Instituto Artístico, [1876]; Memòria Digital de Catalunya

Provider:

[Ateneu Barcelonès](#)



europæana
think culture

http://labs.europeana.eu



[Home](#)

[APIs](#)

[Data](#)

[Apps](#)

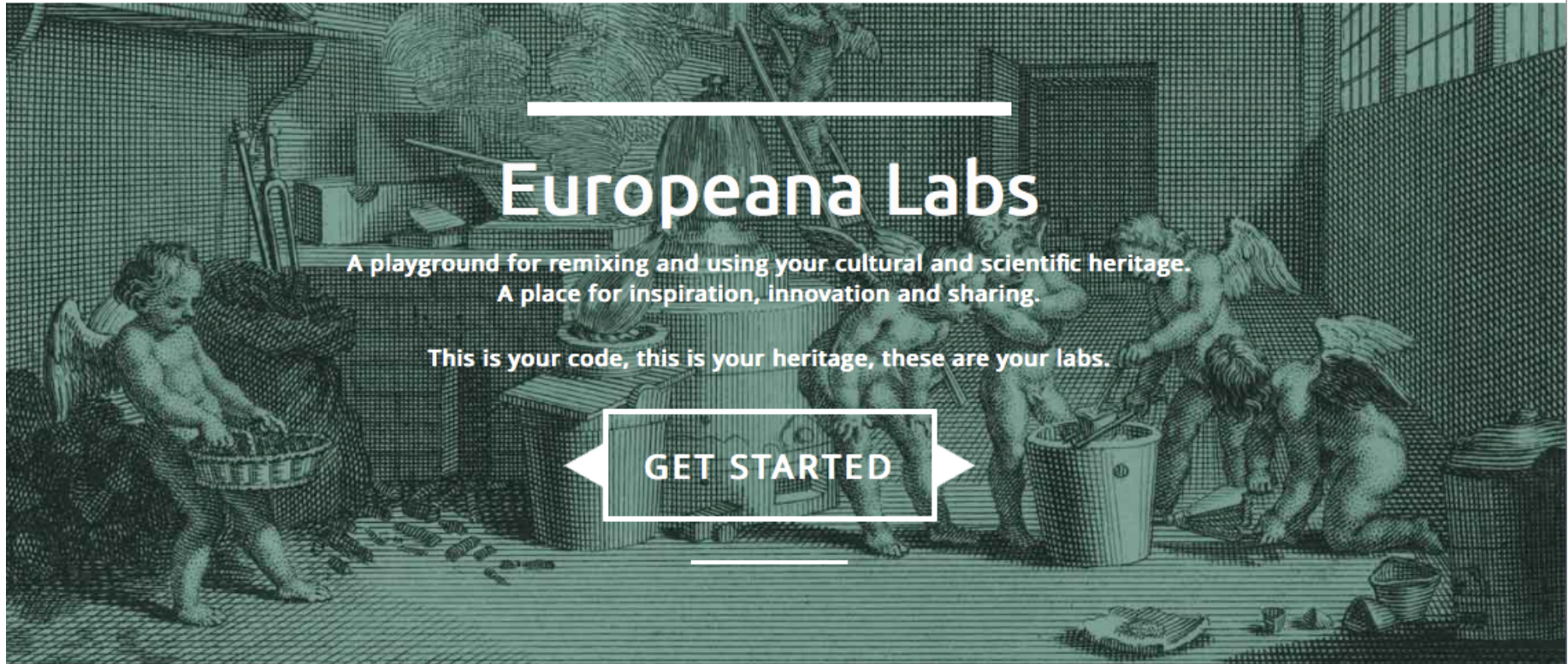
[Blog](#)

[Events](#)

[Incubation](#)

[Support](#)

Search



Europeana Labs

A playground for remixing and using your cultural and scientific heritage.
A place for inspiration, innovation and sharing.

This is your code, this is your heritage, these are your labs.

GET STARTED

europeana
think culture

Europeana API

Beta



Home [APIs](#) Data Apps Blog Events Incubation Support

Search



Get your API key

Introduction

REST API

Introduction

Getting Started

[Get an API-key](#)

Console

Authentication

Search

Query Syntax

Record

Record in JSON-LD

MyEuropeana



Get your API key.


Sign up for your free API key and get developing!
The rest is up to you.

Email address:



Sign up

Europeana Apps

Beta



Home APIs Data **Apps** Blog Events Incubation Support

Apps

Discover and be inspired by the wide range of applications using the Europeana APIs and our Open Data. Is your app missing here? If so get in touch.


Filter by:

- ALL (152)
- OPEN SOURCE TOOLS (53)
- HACKATHON PROTOTYPE (39)
- API IMPLEMENTATION (36)
- R&D RESULTS (28)
- SEARCH (17)

+SHOW MORE

Contribute to the Gallery


Help improve this content



Serendip-o-matic


As their page states: "Serendip-o-matic connects your sources to digital materials located in libraries, museums, and archives around the world. By first examining your research interests, and then identifying related content in locations such as the Digital Public Library of America (DPLA), Europeana, and Flickr Commons, our serendipity engine helps you discover photographs, documents, maps and other primary sources."

API IMPLEMENTATION | DISCOVERY



Kringla

API IMPLEMENTATION



Europeana
think culture

Linked Data

- Semantic web *”a web of data that can be processed directly and indirectly by machines”* Tim Berners-Lee
- Linked Data is a set of principles and technologies providing a publishing paradigm for sharing and reusing data on the Web

Linked Data design issues (T.B. Lee)

1. Use URIs to denote things.
2. Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents.
3. Provide useful information about the thing when its URI is dereferenced, leveraging standards such as RDF, SPARQL.
4. Include links to other related things (using their URIs) when publishing data on the Web.



Linked Data simple rules (T.B. Lee, TED 2009)

1. *All kinds of conceptual things, they have names now that start with HTTP.*
2. *I get important information back. I will get back some data in a standard format which is kind of useful data that somebody might like to know about that thing, about that event.*
3. *I get back that information it's not just got somebody's height and weight and when they were born, it's got relationships. And when it has relationships, whenever it expresses a relationship then the other thing that it's related to is given one of those names that starts with HTTP.*



Linked Data & Europeana

- Europeana goal is to provide integrated access to digital objects of the cultural heritage organizations of all the nations of the European Union
- Linked data can enable the use of digital representations of cultural artifacts for generating knowledge



Linked Data & Europeana

- Issues preventing the adoption of the LD paradigm from the beginning of Europeana:
 - lack of metadata expressed in RDF
 - missing links to other sources
 - missing data provider agreements

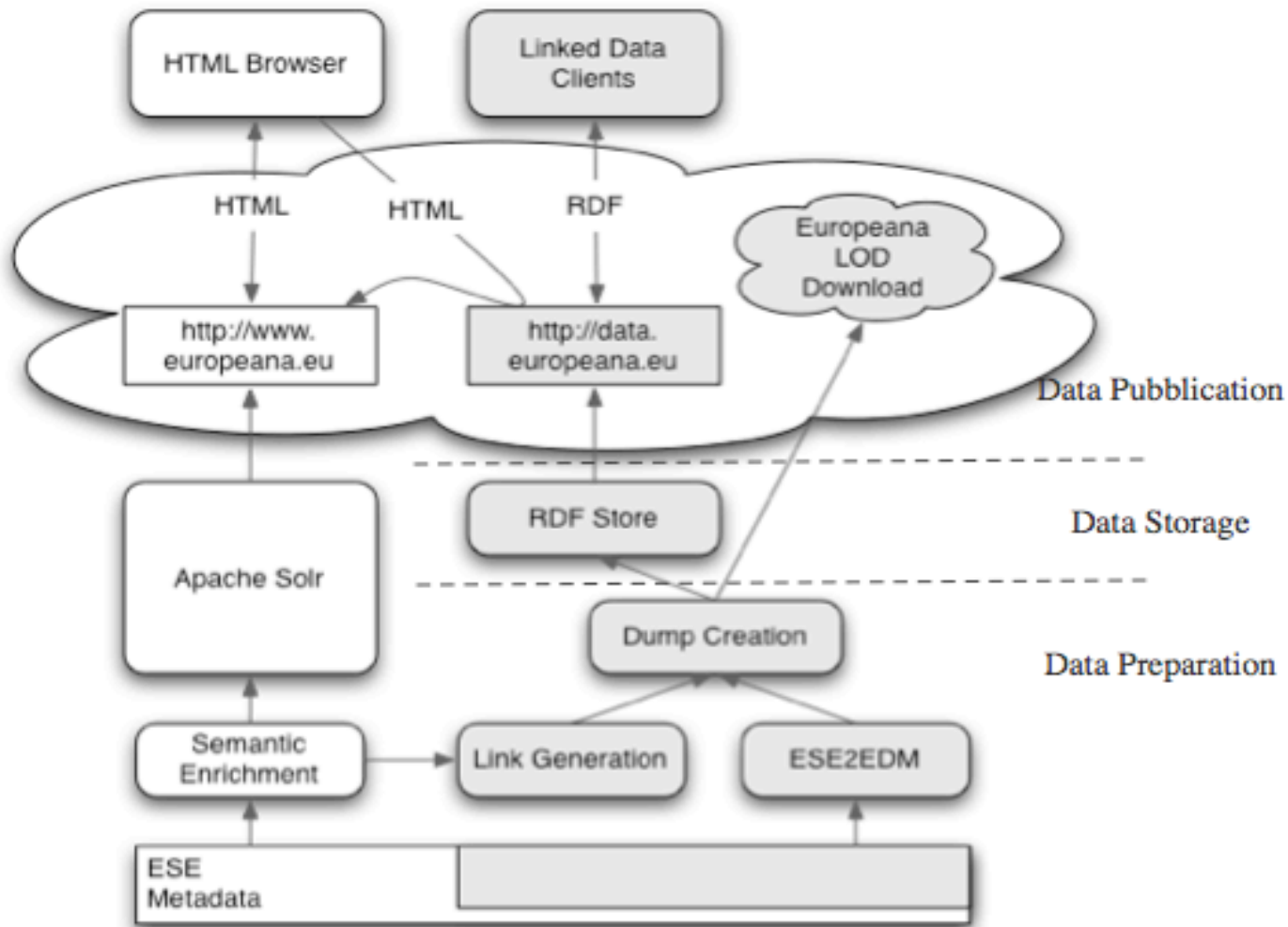


The Europeana Linked Data Pilot

- The Europeana Linked Data Pilot server:
 - Decoupled from the Europeana portal
 - Datasets contains EDM records
 - converting ESE record to EDM structured records
 - Datasets contain collections of institutions which signed the Europeana Data Exchange Agreement (<http://pro.europeana.eu/licensing>)



Europeana LD server overall architecture



Europeana LD Server: overall approach

- Convert Europeana Semantic Element (ESE) metadata into RDF/XML EDM metadata records
 - XML stylesheets, using XSLT 1.0
- Enrich selected metadata fields using controlled vocabularies
 - Annocultur tool (developed at Europeana foundation)
- Link to existing LOD services maintained by Europeana partners (National Library of Hungary, Swedish culture aggregator...)
- Publish the datasets
 - File dump download, RDF triple store



ESE & EDM

- ESE based on unqualified DC + specific fields (e.g. dataProvider)
 - Main issues: *flat* model, values as string, in the same metadata record values belonging to different entities
- EDM designed to open the Europeana information space
 - Key features: distinguish ‘real world objects’ from their digital representations, allow several description for one item, support for complex item representation, re-use and links to existing reference vocabulary reference
 - EDM solves ESE shortcomings



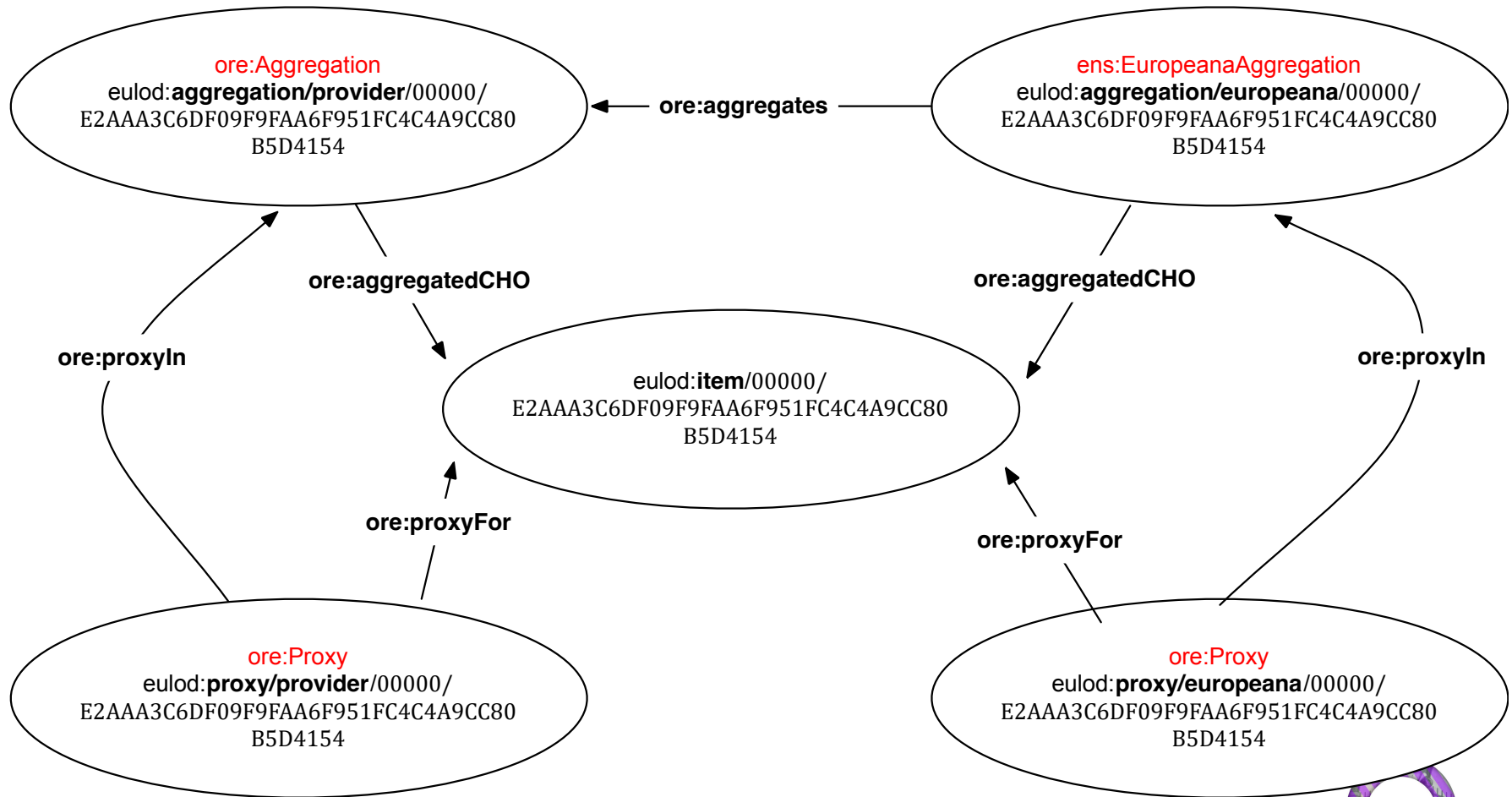
ESE: record example

dc:identifier	02-Lemba-Lakkous-Lady.tif
dc:title	Stone figurine
dc:subject	Cyprus--Antiquities
dc:description xml:lang="en"	Stone vessel from Choirokoitia with embossed decoration.
dc:publisher	Cyprus Archaeological Museum
dc:type	Image
dc:format	image/tiff
dcterms:temporal	mid 3rd millenium B.C.
dc:rights	Cyprus Archaeological Museum
dc:title	Λίθινο ειδώλιο
dc:description	Λίθινο ειδώλιο από τη θέση Λέμπα ...
Europeana:provider	Cyprus Library
Europeana:type	IMAGE
Europeana:rights	http://creativecommons.org/licenses/by/3.0/nl/
Europeana:dataProvider	Cyprus Archaeological Museum
Europeana:isShownBy	http://www.mcw.gov.cy/mcw/da/da.nsf/All/8308A7AB9CA6CB5EC22574CC003FF5E1/\$file/02-Lemba-Lakkous-Lady.jpg?OpenElement
Europeana:uri	http://www.europeana.eu/resolve/record/00000/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154
Europeana:country	cyprus
Europeana:language	gr

ESE2EDM

- The ESE2EDM mapping procedure does:
 - create the EDM entities (items, aggregations, proxies, etc)
 - assign dereferencable URI id to entities
 - add metadata fields to entities

EDM skeleton example



Provider Metadata

xmlns:eulod: "http://data.europeana.eu/"

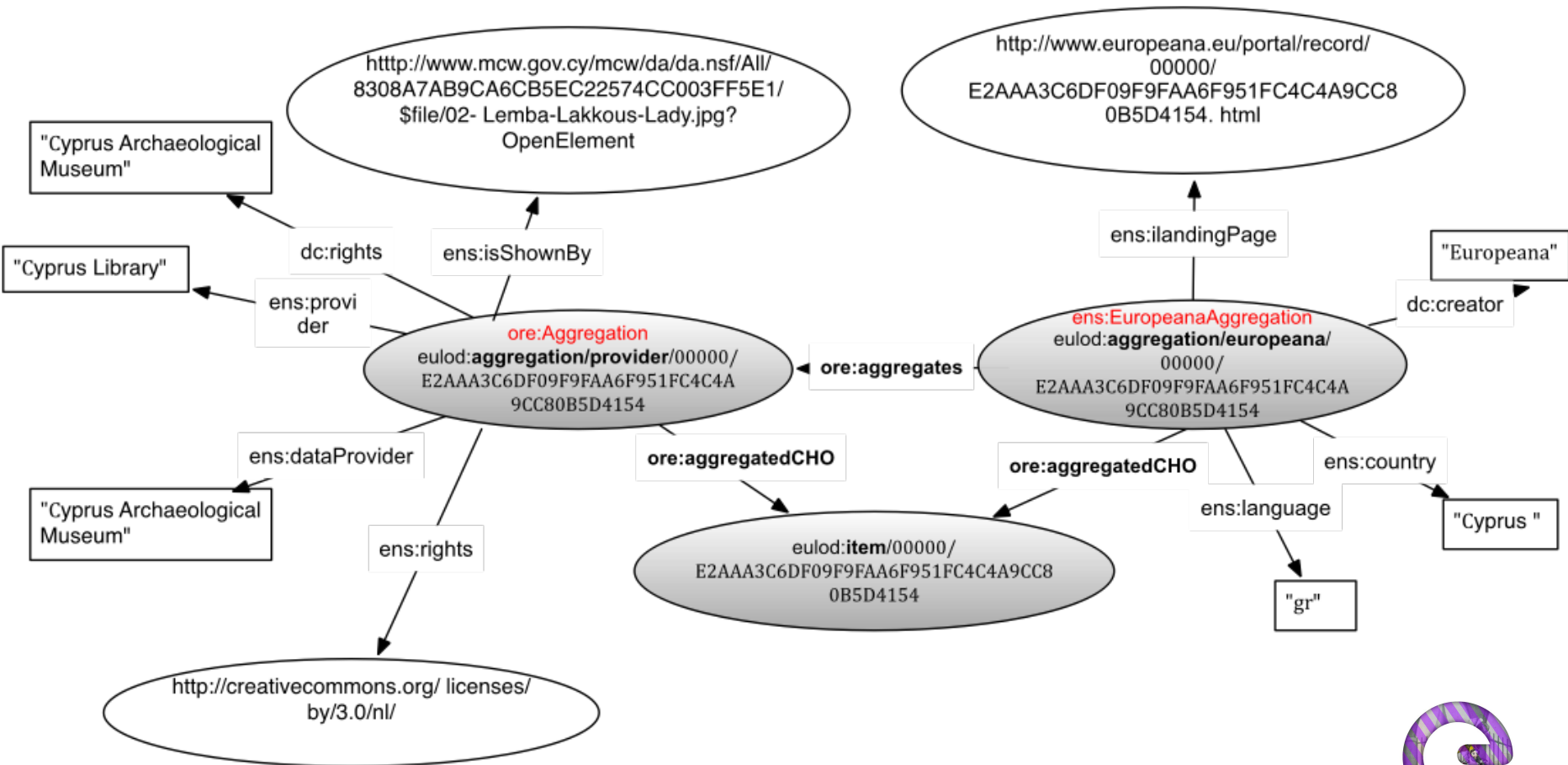
xmlns:ens = "http://www.europeana.eu/schemas/edm/"

xmlns:ore = "http://www.openarchives.org/ore/terms/"

Europeana Metadata



EDM aggregations example



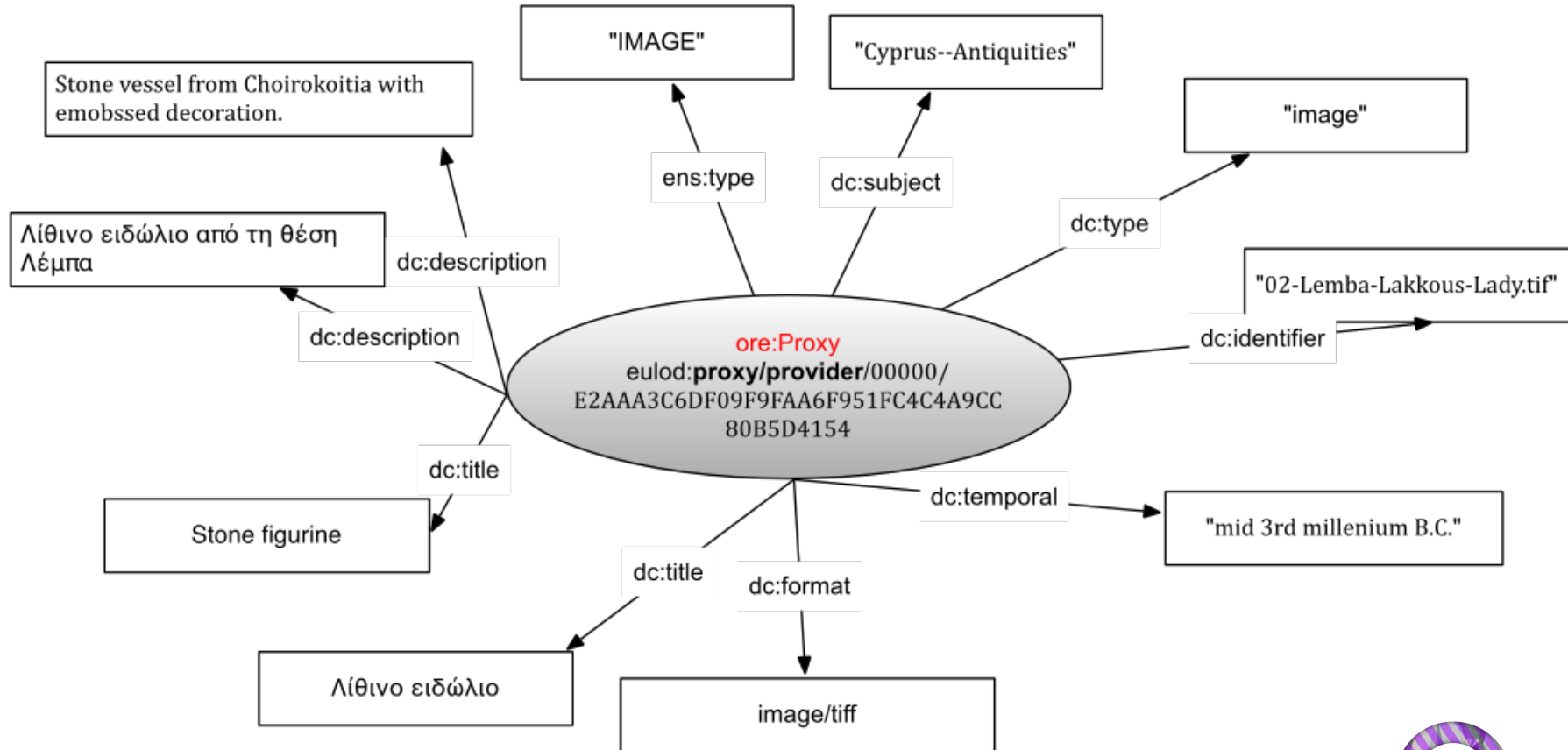
xmlns:eulod: "http://data.europeana.eu/"

xmlns:ens = "http://www.europeana.eu/schemas/edm/"

xmlns:ore = "http://www.openarchives.org/ore/terms/"



EDM proxy example



xmlns:eulod: "http://data.europeana.eu/"

xmlns:ens = "http://www.europeana.eu/schemas/edm/"

xmlns:ore = "http://www.openarchives.org/ore/terms/"

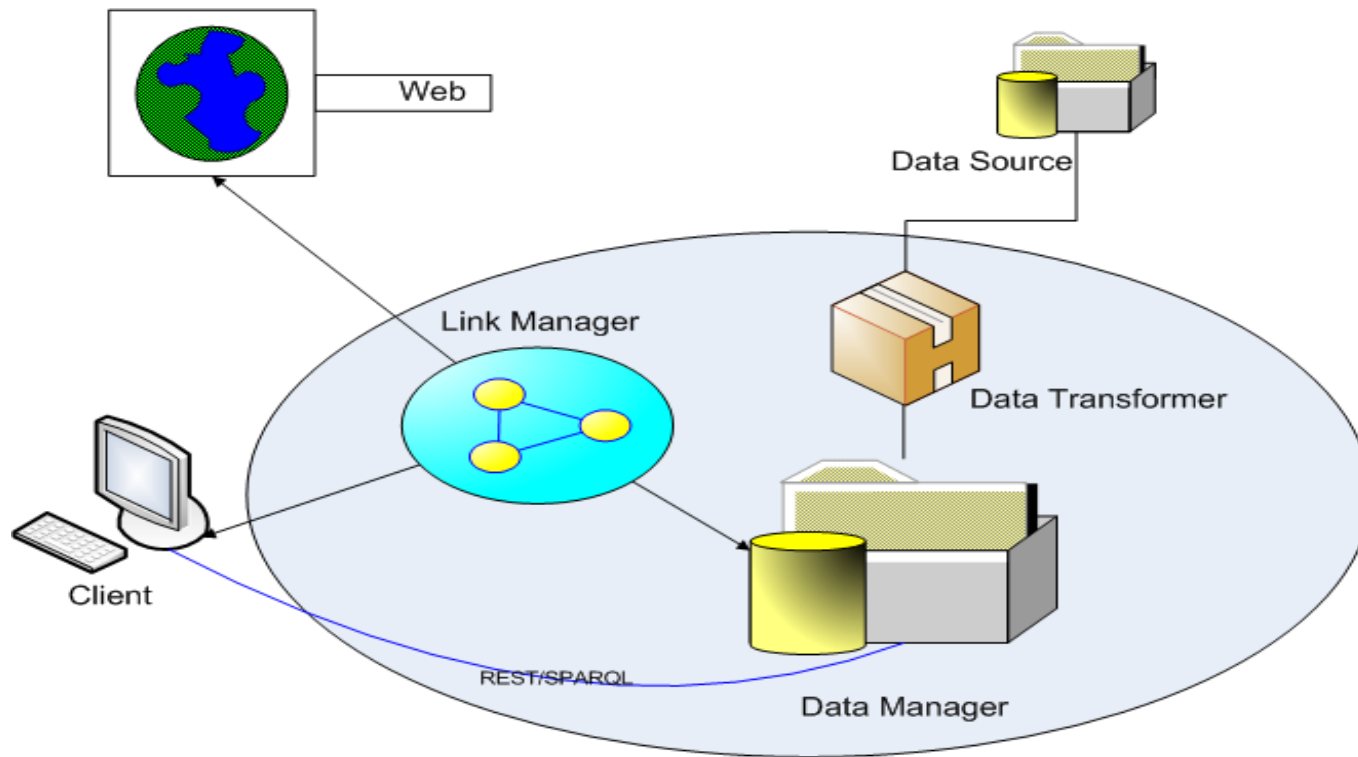
EDM metadata enrichment

- Metadata enrichment is performed by matching values of selected metadata fields with labels of resources from controlled vocabularies
 - Geonames gazetter, GEMET, DBPedia
 - E.g.: `ens:country="Cyprus"` becomes
`ens:country=http://www.geonames.org/146669/`
and additional fields are added for labels, geo coordinates etc.
- Metadata fields produced in semantic enrichment are attached to Europeana's proxy

ESE2EDM: lessons learned

- ESE records are lists of property value pairs while EDM data are networked resource
 - identify the target EDM resource for a given ESE property
- Complex network of resources not easy to ‘consume’
 - feedback from data consumers
- Enhance navigability between resources
 - Advanced RDF store configuration, new properties
- Disseminating meta-level information (provenance, licensing...):
 - Linked Data framework lacks a standardized suite to express such information
- URI design

LD server implementing architecture



Europeana LD Server: data publishing

- Implemented by a Web Server and by a library of Java servlets
- The Web Server receives a request and redirect it to
 - the download area, if a dump file is requested,
 - the servlets library, if a resource is requested.



Linked Data: URI dereferencing strategies

- Two implementation approaches:
 - The 303 URIs: if the server recognizes that the URI identify a real object or an abstract concept , it sends to the client a HTTP response code “303 See Other” and a link to a web document describing the resource, the client then asks for this document
 - The hash URIs: the fragment identifier of a URI (the part of a URI that follows the # symbol) is used to identify real-world objects and abstract concepts, without creating ambiguity.

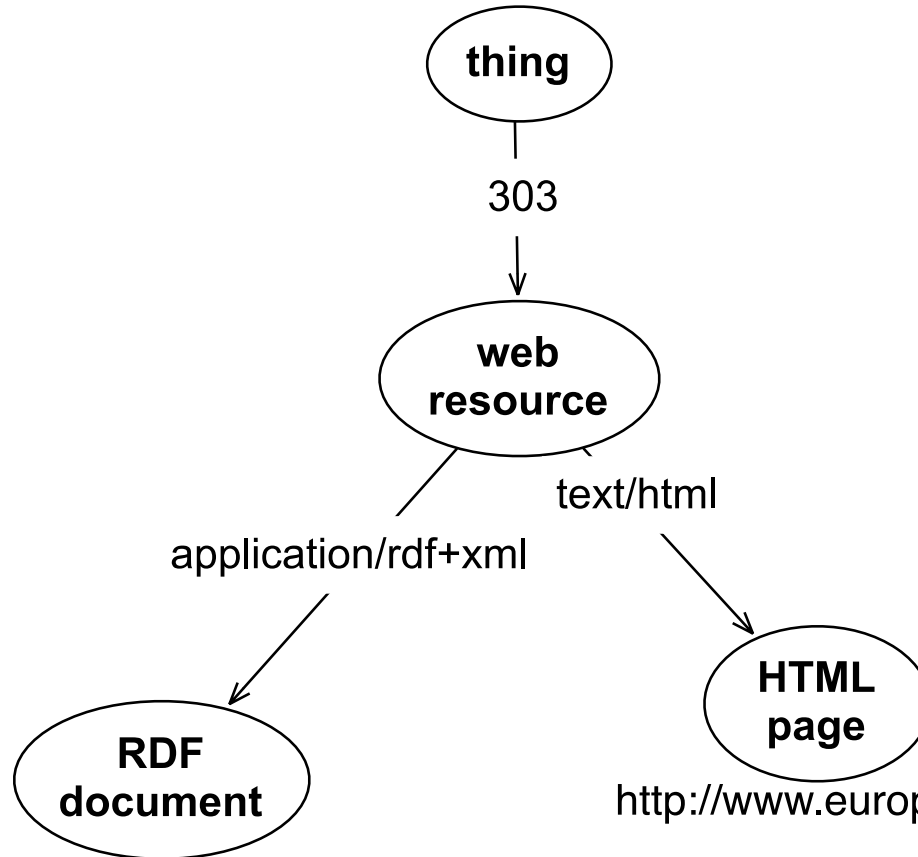


Europeana LD Server: URI dereferencing

- The servlets implement the 303 URIs dereference strategy.
- The implementation algorithm is based on the HTTP server-driven content negotiation mechanism, which enables HTTP clients and servers to negotiate a possible response to a specific request.
 - HTTP “Accept” header

Europeana LD Server: URI dereference

http://data.europeana.eu/item/item_id



http://www.europeana.eu/portal/record/item_id

"DESCRIBE <http://data.europeana.eu/item/item_id>"

Dereferencing:

<http://data.europeana.eu/item/2022109/648CDD89BF914AE7EB9429E0A8F0246A21F98F6F>



cesare.isti.cnr.it

HEAD
/item/2022109/648CDD89BF914AE7EB9429E0A8F0246A21F98F6F
Accept: application/rdf+xml
User-Agent: cesare.isti.cnr.it



data.europeana.eu

303
Location:
<http://data.europeana.eu/data/item/2022109/648CDD89BF914AE7EB9429E0A8F0246A21F98F6F>

HEAD
/data/item/2022109/648CDD89BF914AE7EB9429E0A8F0246A21F98F6F
Accept: application/rdf+xml
User-Agent: cesare.isti.cnr.it

200
Content-type: application/rdf+xml;charset=UTF-8
Vary: Accept



europeana
think culture


Europeana LOD server

- The Europeana Linked Open Data server currently publishes more than 22m
 - Records belonging to providers, who want to make their data available on the web
- The LOD server is separated from the Europeana production server
 - <http://data.europeana.eu>





data.europeana.eu

Beta



Home [APIs](#) Data Apps Blog Events Incubation Support

Europeana Linked Open Data

Introduction

Linked Open Data is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources.

REST API

Linked Open Data

The metadata for all the objects in the Europeana portal is **open**, in that it is all licensed under the [CC0 Public Domain Dedication](#) under the terms of the [Data Exchange Agreement](#) (DEA), and can be freely downloaded via the API.

[Introduction](#)

In October 2012, a large subset of this data was transformed into **linked data** and made available from data.europeana.eu. The data is represented in the Europeana Data Model (EDM). For more information, see our [datasets](#) page.

[Data Downloads](#)

[SPARQL-endpoint](#)

[Data Structure](#)

[FAQ](#)

[Credits](#)

data.europeana.eu started as an experimental pilot in February 2012 with a small number of data providers who committed at an early stage to Europeana's initiative of promoting more open data. The current version of the pilot contains metadata on 20 million texts, images, videos and sounds gathered by Europeana. They cover a great variety of heritage objects, such as this [slovenian version of "O Sole Mio"](#) from the National Library of Slovenia, or Neil Robson's [memories of the herring business](#) from the Tyne and Wear Archives & Museums.

Search Widget

Terms of Use

[Linked Open Data - What is it?](#) from Europeana on [Vimeo](#) (also in [French](#), [German](#), [Italian](#) and [Spanish](#).)



europeana
think culture

Europeana SPARQL endpoint (experimental)

Forest SPARQL Search... Full Search Login



Welcome on the SPARQL end-point of data.europeana.eu!

data.europeana.eu currently contains open metadata on 20 million texts, images, videos and sounds gathered by Europeana. The data follows the terms of the [Creative Commons CC0 public domain dedication](#). The data is described using the [Resource Description Framework \(RDF\)](#) format, and structured using the [Europeana Data Model \(EDM\)](#). We provide more detail about the published EDM data at the [technical details page](#).

Please take the time to check out the [list of collections currently included in the pilot](#).

The terms of use and external data sources appearing at data.europeana.eu are provided on the [Europeana Data sources](#) page.

Sample queries are available on the [sparql](#) page.

Repository overview

Engine: OWLIM SE	Version: 5.3
Cultural Heritage Objects: 20,163,672	Data version: 2.0
From: http://data.europeana.eu/download/2.0/	Last update: 14-Sep-2012
Inference ruleset: owl-horst-optimized	Number of expl. statements: 998,471,854
Number of entities: 265,799,020	Number of statements: 3,798,446,742

© 2009-2013 Ontotext AD. All rights reserved.

Build Timestamp: 2014-07-29 14:06:32

Build Revision: 5714

Querying the Europeana LoD dataset with SPARQL:creash course



Simple Protocol and RDF Query Language (SPARQL)

- *A query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format. (Wikipedia)*
- W3C Recommendation since 2008, last version 1.1 (March 2013)
- SPARQL lets us translate LD's interlinked, graph data into other more readable formats, for instance a normalized, tabular data



SPARQL Query

- A SPARQL query comprises, in order:
 - Prefix declarations, for abbreviating URIs
 - Dataset definition, stating what RDF graph(s) are being queried
 - A result clause, identifying what information to return from the query
 - The query pattern, specifying what to query for in the underlying dataset
 - Query modifiers, slicing, ordering, and otherwise rearranging query results



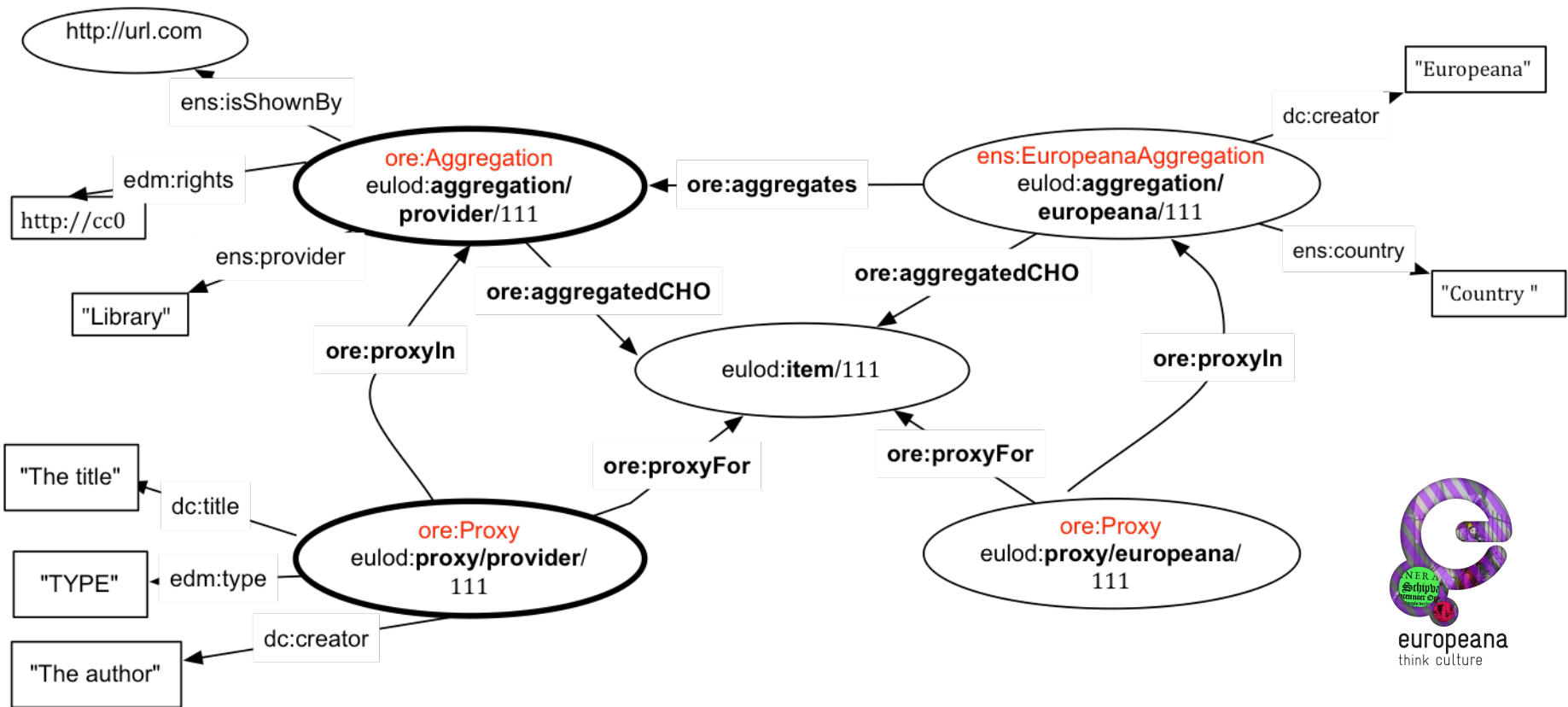
SPARQL Queries

- The results of SPARQL queries can be returned and/or rendered in a variety of formats:
 - XML. SPARQL specifies an XML vocabulary for returning tables of results.
 - JSON. A JSON "port" of the XML vocabulary, particularly useful for Web applications.
 - CSV/TSV. Simple textual representations ideal for importing into spreadsheets
 - RDF. Certain SPARQL result clauses trigger RDF responses, which in turn can be serialized in a number of ways (RDF/XML, N-Triples, Turtle, etc.)
 - HTML. When using an interactive form to work with SPARQL queries. Often implemented by applying an XSL transform to XML results.

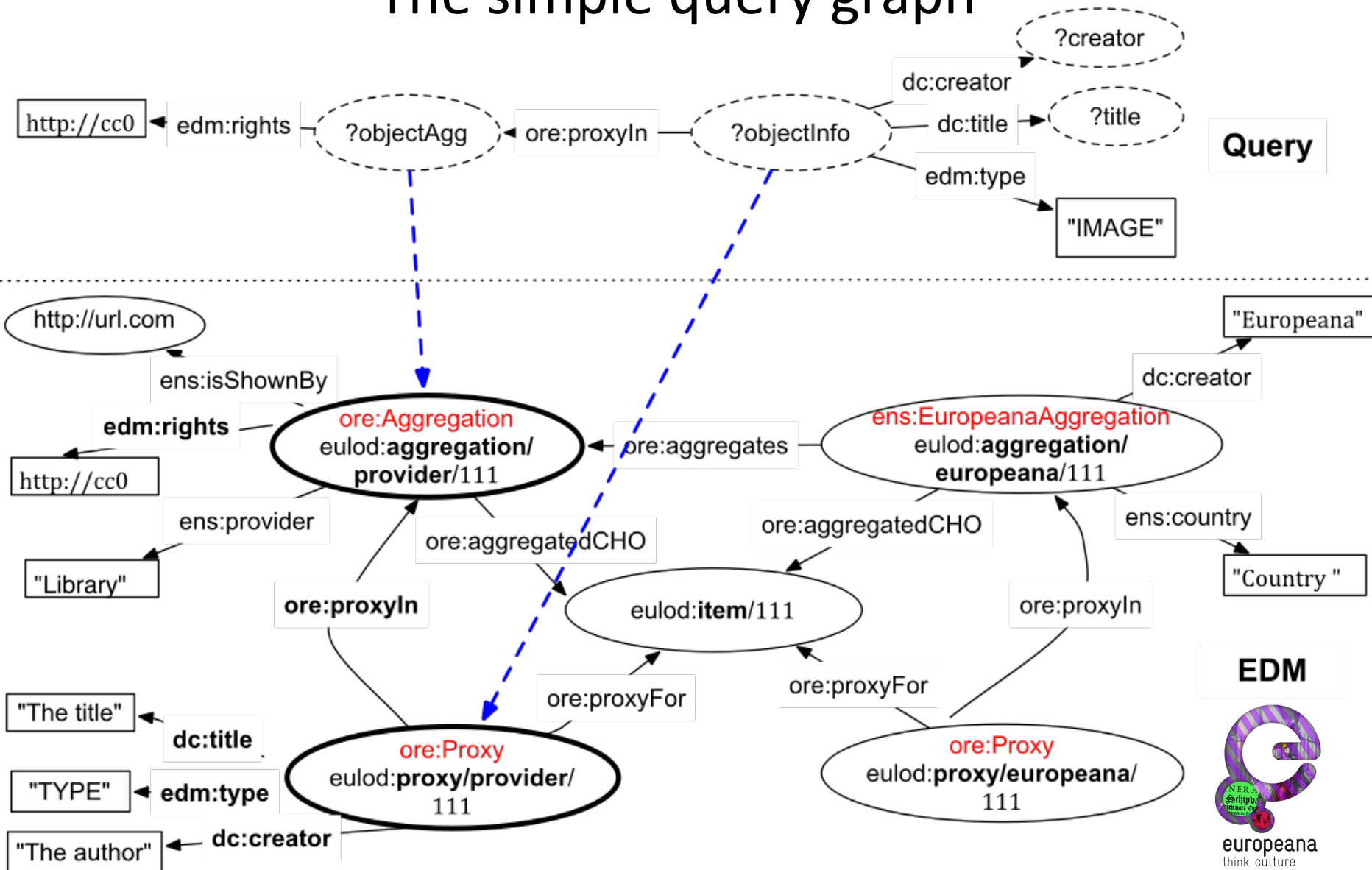


A simple query

- We want all images with CC0 license having a title and a creator.



The simple query graph



The data model: Prefixes

```
Prefixes
a      http://www.w3.org/1999/02/22-rdf-syntax-ns#type
cc:    http://creativecommons.org/ns#
data:  http://data.europeana.eu/
dc:    http://purl.org/dc/elements/1.1/
dct:   http://purl.org/dc/terms/
edm:   http://www.europeana.eu/schemas/edm/
eu:    http://www.europeana.eu/
foaf:  http://xmlns.com/foaf/0.1/
html:  http://www.w3.org/1999/xhtml/vocab#
ore:   http://www.openarchives.org/ore/terms/
```

- We'll be using these prefixes for our query:

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX edm: <http://www.europeana.eu/schemas/edm/>

PREFIX ore: <http://www.openarchives.org/ore/terms>



The query

PREFIX dc: <<http://purl.org/dc/elements/1.1/>>

PREFIX edm: <<http://www.europeana.eu/schemas/edm/>>

PREFIX ore: <<http://www.openarchives.org/ore/terms/>>

SELECT ?creator ?title

WHERE {

 ?objectInfo dc:title ?title .

 ?objectInfo dc:creator ?creator .

 ?objectInfo edm:type "IMAGE" .

 ?objectInfo ore:proxyIn ?objectAgg .

 ?objectAgg edm:rights <<http://creativecommons.org/publicdomain/zero/1.0/>>

}



SPARQL query example

- How many images in the dataset for every specific license provided by each data contributor

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX ore: <http://www.openarchives.org/ore/terms/>
SELECT ?edmrighs ?provider (COUNT(*) as ?count)
WHERE {
    ?objectAgg edm:provider ?provider .
    ?objectAgg edm:rights ?edmrighs .
    ?objectInfo ore:proxyIn ?objectAgg .
    ?objectInfo edm:type "IMAGE" .
}
GROUP BY ?edmrighs ?provider
ORDER BY DESC(?count)
```

Europeana LoD server: data access

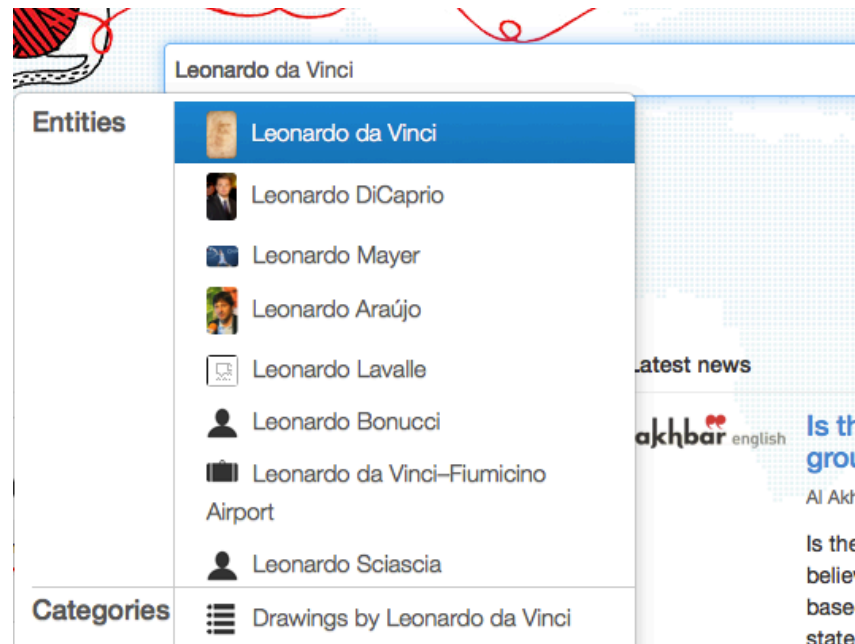
Publishing method	File Transfer	REST/SPARQL	HTTP/GET
Data published			
Complete dataset	Download dataset dump	N.A.	N.A.
Collection of resources	Download collection(s) dump	SPARQL 'Select' query	N.A.
Single resource	N.A.	SPARQL 'Describe' query	URI dereference

Current activities

- Distribute the whole Europeana dataset
 - Agreements with content providers
- Challenges:
 - Licensing: 64% (as of June 2013) of metadata records does not have clear info about content license
 - Improve metadata record quality
 - Optimizing data for reuse
 - Improve the LOD server performances

Future activities: Europeana Entity search

- Exploit Lod server in order to enable users to write queries by using *entities*
 - Entity Linking (EL)



Future activity: Europeana Entity search

- The Europeana dataset is processed in order to individuate entities, relevant for the collection. Entities will be collected with different techniques:
 - a. Exploiting the semantic annotations of Europeana
 - b. Using Named Entities Recognition and Disambiguation techniques
- For every entity the system will collect specific information from several controlled sources (Wikipedia, DBpedia, Freebase, WikiArt...),
 - Defined set of fields
- The resulting dataset is processed to create a consistent and networked 'collection' of entities



Future activity: Europeana Entity search

The screenshot shows search results for 'PABLO PICASSO' and 'Anette Robinson : Guernica de Picasso'. The Picasso entry includes a portrait, a description, creator (Arnold Newman), date (1954-12-24), type (Photo), format (image/jpeg), and identifier (spdk19960922-471710-3). The Guernica entry includes a thumbnail, contributor (Olivier Barrot), publication date (2003-05-14), type (Magazine), format (0h1m57s ; video/mpeg), identifier (2296766001), and a link to the INA website. Blue dashed lines connect the Picasso portrait to the Guernica title and the Guernica thumbnail to the Picasso name, indicating a search path.

A row of logos for semistructured resources: Wikipedia, Freebase, WikiArt (Visual Art Encyclopedia), and DBpedia. A box labeled 'Other semistructured resources' is positioned to the right of these logos.

The screenshot shows a detailed Wikipedia article for Pablo Picasso. It features a portrait of Picasso, a grid of various artworks, and a large image of 'Guernica'. The article text includes his birth (October 25, 1881, Málaga, Spain) and death (April 8, 1973, Mougins, France), his full name, and his role as a painter, sculptor, and designer. It also lists his children and the location of the 'Guernica' painting (Museum Reina Sofia, Madrid, Spain).



Thank you

- <http://data.europeana.eu>
- Dataset dump download area:
 - <http://labs.europeana.eu/api/linked-open-data/data-downloads/>
- Acknowledgment: the ESE2EDM mapping approach has been designed by Antoine Isaac and Bernhard Haslhofer