



*The*  
**European  
Library**

*Connecting knowledge*

# **Links, languages and semantics: linked data approaches in The European Library and Europeana.**

**Valentine Charles, Nuno Freire & Antoine Isaac**

14<sup>th</sup> August 2014, IFLA2014 satellite meeting, Paris



# The European Library and Europeana: two aggregation services



# The European Library

- Collects metadata for digital and non-digital materials from 48 National Libraries and more than 50 University / Research libraries
  - <http://www.theeuropeanlibrary.org/>
- A very large scholarly dataset
  - 200 million bibliographic records
  - 600 collections and their descriptions
  - 19 million digitised objects
  - Large corpus of full-text pages
- Data distributed as Linked Data
  - 61 datasets available as RDF dumps at <http://www.theeuropeanlibrary.org/tel4/access/data/opendata>



# Europeana

- Aggregates metadata from the cultural heritage sector in Europe
  - libraries, museums, archives and audio-visual archives
- Provides a portal for users to access data and objects
  - <http://www.europeana.eu/>
  - Metadata under Creative Commons Zero - public domain
  - Previews and links to source
- Data distributed via
  - API <http://labs.europeana.eu/api/>
  - Linked Data (currently being updated) <http://data.europeana.eu/>



---

# Facing the same challenges

## → Data aggregation issues

- Too many interpretations of cataloguing practices and bibliographic data models
- Legacy data affected by diverse data migration processes

## → Multilinguality issues

- Provide access to multilingual resources
- Allow the search for items in various languages
- Make sure users can understand the descriptions of these items



# Create new data frameworks

## → Europeana Data Model (EDM)

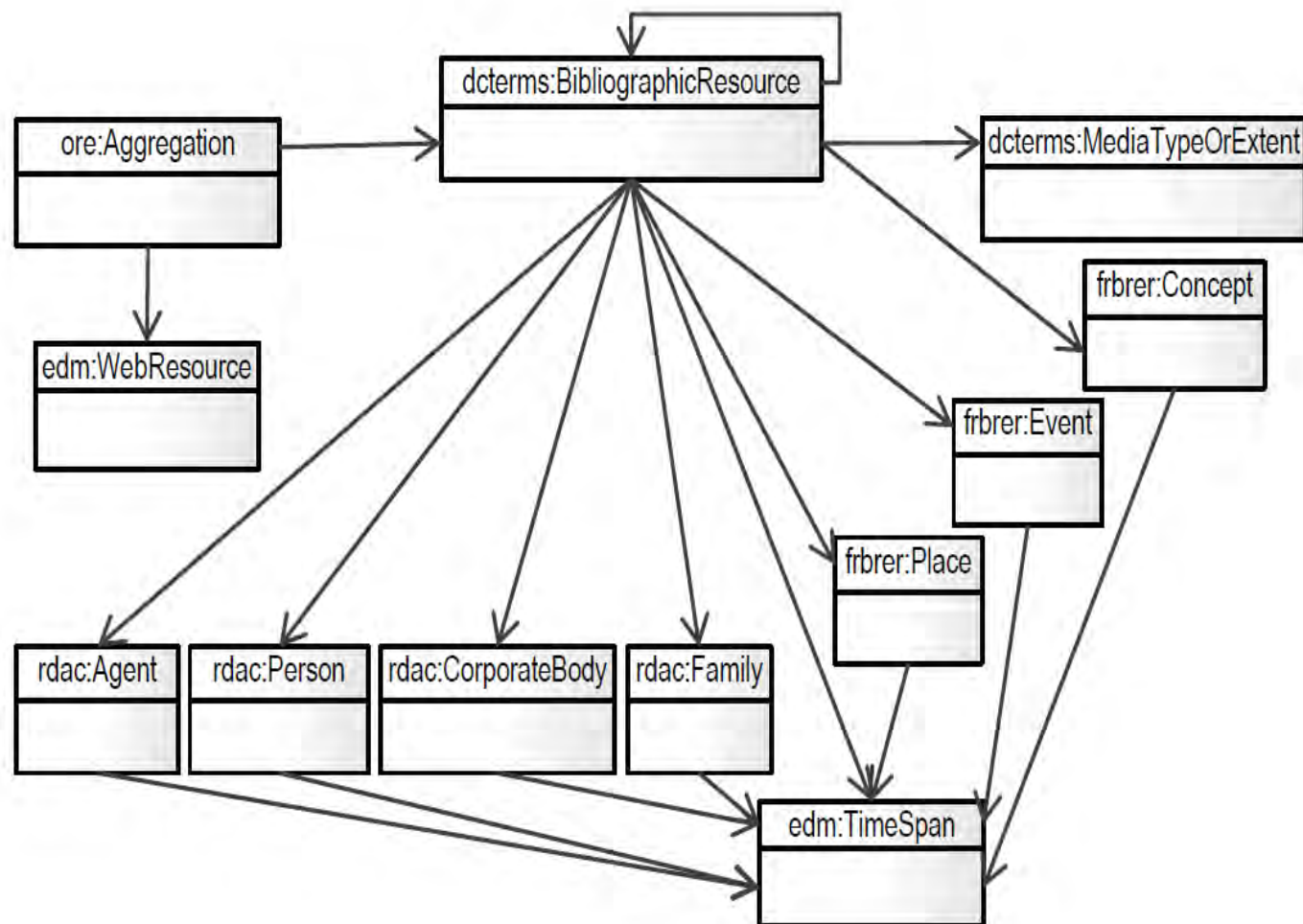
- Re-uses several existing Semantic Web-based models: Dublin Core, OAI-ORE, SKOS, CIDOC-CRM...
- More granular metadata
  - links e.g. between objects and context entities (persons, places)
  - multilingual & semantic linked data for contextual resources (e.g. Concepts)





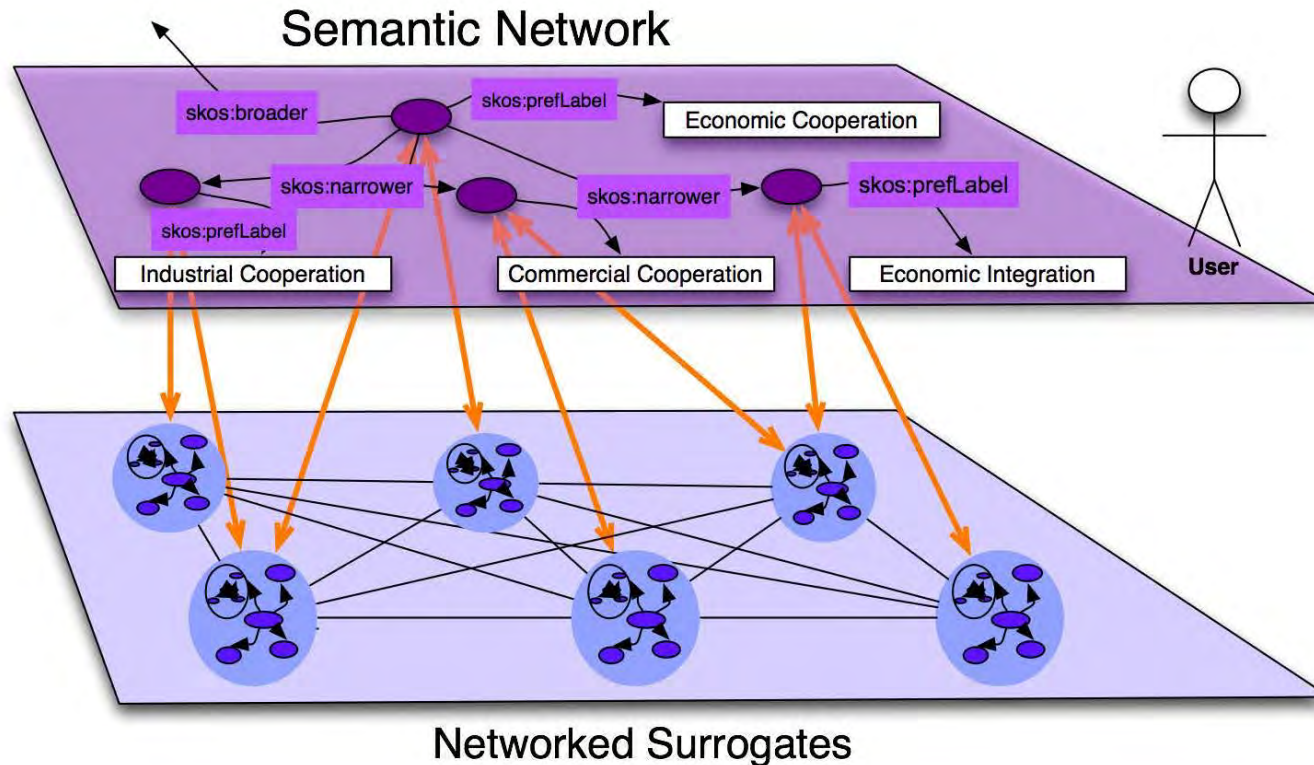
# Create new data frameworks

→ The European Library created its model to publish Linked Data



- Re-uses RDA, FRBR, EDM
- Enables description of bibliographic resources, their digital representations and contextual information

# Rely on knowledge organisation systems



- Create a “semantic layer” on top of cultural heritage objects
- Include multilingual “value vocabularies”
  - From Europeana’s providers or from third-party data sources





---

# Encouraging providers to contribute their own vocabularies

- Benefit from data links made at data providers' level
- Ingestion of vocabularies is made possible if the vocabularies used the data structures EDM expects
  - For instance SKOS for concepts



# In Europeana



© Free access - no re-use

View item at  
[Muzej za umjetnost i obrt, Zagreb](#)



Share



Cite on Wikipedia



Translate details

Select language ▼

Powered by Microsoft® Translator

## PARK LUXEMBOURG: akvarel

**Creator:**

[Josip Račić](#) (Josip Račić); [Račić, Josip](#) (Račić, Josip)

**Contributor:**

[Muzej za umjetnost i obrt](#)

**Geographic coverage:**

[Paris](#) ; ; [France](#) ; [Paris](#) ; [France](#)

**Date:**

1907 - 1907

**Type:**

[watercolours \(paintings\)](#); [watercolor \(painting\)](#); [figura](#); [figure](#); [Jugendstil](#); [Secessionstil](#); [IMAGE](#); [figures \(representations\)](#); [Figuren \(Darstellungen\)](#); [Art Nouveau](#)

**Format:**

image/jpg; width 28,7 cm; height 22,2 cm;

**Identifier:**

MUO\_027043

**Is part of:**

SLIKARSTVO

## What ▾

**Concept Term:** <http://partage.vocnet.org/part00365>

**Concept Label:** [Jugendstil] (hr) ; [Jugendstil] (sl) ; [Jugendstil] (ca) ; [Jugendstil] (it) ; [Jugendstil] (no) ; [Szecesszió (Jugendstil)] (hu) ; [secese (Jugendstil)] (cs) ; [Jugendstil] (de) ; [Jugend (Jugendstil)] (fi) ; [Jugendstil] (pt) ; [Jugendstil] (pl) ; [Jugendstil] (fr) ; [Jugend (Jugendstil)] (sv) ; [Jugendstil] (en) ; [Jugendstil] (es) ; [Jugendstil] (nl)

**Concept Broader Label:** <http://partage.vocnet.org/part00353>; <http://partage.vocnet.org/part00349>

**Concept Term:** <http://partage.vocnet.org/part00501>

**Concept Label:** [akvareli] (hr) ; [akvareli] (sl) ; [acquarelli] (it) ; [aquarel-les] (ca) ; [akvarell] (no) ; [akvarell] (hu) ; [akvarel] (cs) ; [Aquarell (visuelles Werk)] (de) ; [akvarellimaalaus] (fi) ; [aguarela] (pt) ; [akwarele] (pl) ; [akvareller] (sv) ; [aquarelle] (fr) ; [watercolours (paintings)] (en) ; [acuarelas] (es) ; [aquarellen] (nl)

**Concept Broader Label:** <http://partage.vocnet.org/part00044>

**Concept Term:** <http://partage.vocnet.org/part01201>

**Concept Label:** [Figur (Darstellung)] (de) ; [figures (representations)] (en)

**Concept Broader Label:** <http://partage.vocnet.org/part01377>

**Concept Term:** [http://dbpedia.org/resource/Art\\_Nouveau](http://dbpedia.org/resource/Art_Nouveau)

**Concept Label:** [Jugendstil] (de) ; [新艺术运动] (zh) ; [Art Nouveau] (it) ; [Art nouveau] (pt) ; [Secesja (sztuka)] (pl) ; [Art nouveau] (fr) ; [Jugend] (sv) ; [Art Nouveau] (en) ; [Модерн] (ru) ; [Modernismo (arte)] (es) ; [Jugendstil] (nl) ; [アール・ヌーヴォー] (ja)



---

# Automatic enrichment in Europeana

- Objects are connected to third-party sources
- Places from GeoNames
  - Agents from DBpedia
  - Concepts from GEMET and DBpedia



# In Europeana



© Free access - no re-use

View item at  
[Royal Library of Belgium](#)

Share

Cite on Wikipedia

Translate details

Select language ▼

Powered by Microsoft® Translator

## Guerre de 1914 carte spéciale

### Description:

70 kilomètres = [7,8 cm];

1 carton: "Colonies allemandes"

### Geographic coverage:

[Paris](#) ; ;

### Date:

[1914](#); Part of: [Early 20th century](#); From: 01-01-1914 — To: 31-12-1914

### Type:

[map](#); [Monograph](#); [Still image](#)

### Format:

Printed; col., ill., texte; 1 carte; 51 x 54 cm

### Subject:

[First World War](#); [Great War](#); [WW1](#); [WWI](#); [World War 1](#); [World War I](#); [World War One](#); [World War, 1914-1918--Campaigns--Western Front](#); [World War, 1914-1918--Historiography](#)

### Identifier:

1\_1300771

[Auto-generated tags](#) ▼



Auto-generated tags ▾

What ▾

**Concept Term:** <http://www.eionet.europa.eu/gemet/concept/14934>

**Concept Label:** [proiecție geografică] (ro) ; [coğrafi projeksiyon] (tr) ; [geografisk projeksjon] (no) ; [földgömb ábrázolása síklapon] (hu) ; [ģeogrāfiskā projekcija] (lv) ; [geografine projekcija] (lt) ; [Geographische Projektion] (de) ; [geographical projection, 地理投影] (def) ; [maantieteellinen projektio] (fi) ; [projection géographique] (fr) ; [geografisk projektion] (sv) ; [Географска проекция] (bg) ; [kartografska projekcija] (sl) ; [geografické zobrazenie] (sk) ; [geografisk prognose] (da) ; [proiektzio geografiko] (eu) ; [proiezione geografica] (it) ; [projezzjoni ġeografika] (mt) ; [مسقط جغرافی] (ar) ; [projekce geografická] (cs) ; [γεωγραφική προβολή] (el) ; [odzworowanie geograficzne] (pl) ; [projecções geográficas] (pt) ; [geographical projection] (en) ; [географическая проекция] (ru) ; [kaardiprojektsioon] (et) ; [proyección geográfica] (es) ; [geografische projectie] (nl)

**Concept Broader Label:** <http://www.eionet.europa.eu/gemet/concept/2286>

Geographical coverage ▾

**Place Term:** <http://sws.geonames.org/2988507/>

**Place Label:** [Lungsod ng Paris, Paris] (tl) ; [París] (ca) ; [Paris] (tr) ; [Paris] (no) ; [Paris] (nn) ; [Parys] (fy) ; [Париж] (tg) ; [Paris] (gd) ; [Pariz] (br) ; [Páras] (ga) ; [ປາຣີສ] (th) ; [París] (oc) ; [Pariisi] (fi) ; [Παρίσι] (ta) ; [Paris] (sv) ; [Lutece, Paname, Pantruche, Paris, Ville-Lumière] (fr) ; [Париж] (bg) ; [பேர்ப்பு] (ka) ; [Paris] (sw) ; [Парыж] (be) ; [Pariz] (sl) ; [Paris] (kw) ; [Paríž] (sk) ; [Париж] (os) ; [Paris] (da) ; [Париж] (sr) ; [París] (ku) ; [Paris, Parisi] (sq) ; [파리 시] (ko) ; [Paris] (cy) ; [פאריס] (he) ; [ಪ್ಯಾರಿಸ್] (kn) ; [Paříž] (cs) ; [Paries] (li) ; [París - Paris] (gl) ; [Lutetia Parisorum] (la) ; [Paarys] (gv) ; [Paryž] (pl) ; [Париж] (ru) ; [Paräis] (lb) ; [Pariz] (hr) ; [Pari] (ln) ; [巴黎] (zh) ; [Paris] (ro) ; [Pa-ri, Paris] (vi) ; [Párizs] (hu) ; [Paríze] (lv) ; [Paryžius] (lt) ; [Paris] (de) ; [Paris] (id) ; [Paris] (ia) ; [75091, Paris] (def) ; [Paris] (mg) ; [Φωληή] (hy) ; [Париж] (mk) ; [París] (is) ; [Paris] (eu) ; [Париж] (uk) ; [Parigi] (it) ; [पेरिस] (mr) ; [پاریس] (ug) ; [Paris] (ms) ; [پاریس] (fa) ; [پیرس] (ur) ; [پاریس] (ar) ; [Paris] (io) ; [Paris] (ty) ; [Paris] (na) ; [Παρίσι] (el) ; [ፖሪስ] (am) ; [París] (an) ; [Paris] (pt) ; [Parizo] (eo) ; [Paris] (en) ; [Pariis] (et) ; [París] (es) ; [Parijs] (nl) ; [پاریس] (ja) ; [Parys] (af)

**Geo Space:** 48.85341 ; 2.3488





---

# In The European Library

- The European Library has access to a large set of subjects headings and authority data
  - Most if these authorities are local and language-dependent
  - Data processing and conversion results in loss of information (synonyms, hierarchical links)
- The European Library has to apply information extraction techniques to align open data pivot sources



---

# Consolidation of author data using VIAF

- First matching of data about contributors to VIAF entries is based on the comparison of authority record identifiers
  - Depends on the involvement of the library in VIAF and the presence of authority record identifiers
- Second matching relies on additional data such as birth and death dates, names variants...
- Further enrichment performed against the International Standard Name Identifier (ISNI)



# Multilingual Access to Subjects (MACS)

- MACS project has produced manual and semi automatic alignments between:
  - Library of Congress Subject Heading (LCSH)
  - RAMEAU
  - Schlagwortnormdatei (SWD)
- 120,000 links created
- MACS is integrated in The European Library as links included in all bibliographic data.



```
<?xml version="1.0" encoding="UTF-8"?>
<add xmlns="http://www.theeuropeanlibrary.org">
  <doc>
    <field name="id">MACS0000005</field>
    <field name="SWD">Schauspieler</field>
    <field name="SWD_number">4052154-0</field>
    <field name="LCSH">Actors</field>
    <field name="LCSH_number">sh 85000744</field>
    <field name="RAMEAU">Acteurs</field>
    <field name="RAMEAU_number">FRBNF133184427</field>
  </doc>
</add>
```

An example of a MACS record before and after additions by The European Library :

- ARK identifiers
- LOD URIs

```
<?xml version="1.0" encoding="UTF-8"?>
<add xmlns="http://www.theeuropeanlibrary.org">
  <doc>
    <field name="id">MACS0000005</field>
    <field name="SWD">Schauspieler</field>
    <field name="SWD_number">4052154-0</field>
    <field name="SWD_uri">http://d-nb.info/gnd/4052154-0</field>
    <field name="LCSH">Actors</field>
    <field name="LCSH_number">sh 85000744</field>
    <field name="LCSH_uri">http://id.loc.gov/authorities/subjects/sh85000744</field>
    <field name="RAMEAU">Acteurs</field>
    <field name="RAMEAU_number">FRBNF133184427</field>
    <field name="RAMEAU_ark">ark:/12148/cb133184428</field>
    <field name="RAMEAU_uri">http://data.bnf.fr/ark:/12148/cb133184428</field>
  </doc>
</add>
```

link ( <a href="#">related_resource</a> )	
url	http://catalogimages.wiley.com/images/db/jimages/9781118105375.jpg
linkStatus	NOT_CHECKED
bibliographic_level	MONOGRAPH
language ( <a href="#">language_of_content</a> )	ENG
language ( <a href="#">language_of_cataloguing</a> )	ENG
edition	
number	2
statement	2nd ed.
topic ( <a href="#">lcc</a> )	
topicName	HG4026
identifiers	
identifier	http://id.loc.gov/authorities/classification/HG4026
topic ( <a href="#">ddc</a> )	
topicName	658.15
edition	23
topic ( <a href="#">lcsch</a> )	
topicName	Corporations
identifiers	
identifier	1000000185745
scope	TEL
generalSubdivision	Finance.
topic ( <a href="#">lcsch</a> )	
identifiers	
identifier	http://id.loc.gov/authorities/subjects/sh85032938
topic ( <a href="#">rameau</a> )	
identifiers	
identifier	http://data.bnf.fr/ark:/12148/cb11980284x
topic ( <a href="#">gnd</a> )	
identifiers	
identifier	http://d-nb.info/gnd/4206416-8

An internal representation of a bibliographic record at TEL showing multilingual subject enrichments added by using MACS

Additional subject heading

Source Subject Heading

Enrichments added through MACS



Metadata format: EDM

```
<rdf:RDF>
  <edm:ProvidedCHO
rdf:about="http://data.theeuropeanlibrary.org/BibliographicResource/3000094132561">
  <dcterms:spatial>Hoboken, N.J.</dcterms:spatial>
  <dcterms:spatial rdf:resource="http://id.loc.gov/vocabulary/countries/us"/>
  <dcterms:issued>2012</dcterms:issued>
  <dcterms:issued>c2012.</dcterms:issued>
  <dcterms:extent>ill. ;</dcterms:extent>
  <dcterms:extent>xix, 505 p. :</dcterms:extent>
  <dcterms:extent>26 cm.</dcterms:extent>
  <dcterms:isPartOf>CFA Institute investment series</dcterms:isPartOf>
  <dcterms:isPartOf>http://data.theeuropeanlibrary.org/Collection/a1195_test</dcterms:isPartOf>
  <dc:description xml:lang="en">Illustrations</dc:description>
  <dc:type xml:lang="en">Monograph</dc:type>
  <dc:description>2nd ed.</dc:description>
  <dc:contributor>Michelle R. Clayman</dc:contributor>
  <dc:contributor>Martin S. Fridson</dc:contributor>
  <dc:contributor>George H. Troughton</dc:contributor>
  <dc:publisher>John Wiley & Sons</dc:publisher>
  <dc:subject>Library of Congress Classification: HG4026</dc:subject>
  <dc:subject>Dewey Decimal Classification: 658.15</dc:subject>
  <dc:subject>Corporations, Finance.</dc:subject>
  <dc:subject rdf:resource="http://id.loc.gov/authorities/subjects/sh85032938"/>
  <dc:subject rdf:resource="http://data.bnf.fr/ark:/12148/cb11980284x"/>
  <dc:subject rdf:resource="http://d-nb.info/gnd/4206416-8"/>
  <dc:subject rdf:resource="http://dewey.info/class/658/about"/>
  <dc:identifier>http://data.copac.ac.uk/id/64151214</dc:identifier>
  <dc:language>en</dc:language>
  <dc:relation rdf:resource="http://catalogimages.wiley.com/images/db/jimages/9781118105375.jpg"/>
  <dc:format xml:lang="en">Printed</dc:format>
  <dc:description>[edited by] Michelle R. Clayman, Martin S.
Fridson, George H. Troughton.</dc:description>
  <dc:description>Formerly CIP.</dc:description>
  <dc:description>Includes bibliographical references (p.
475-479) and index.</dc:description>
  <dc:title>Corporate finance a practical
approach</dc:title>
  <edm:type>TEXT</edm:type>
</edm:ProvidedCHO>
<core:Aggregation
rdf:about="http://data.theeuropeanlibrary.org/BibliographicResource/3000094132561#aggregation">
```

The subject enriched  
record in EDM for delivery  
to Europeana

Enrichments  
added through  
MACS





# Building library LOD services

The vision: accelerate the availability of library LOD through The European Library

→ Publishing LOD becomes easier for libraries

- It is built on TEL's data aggregation infrastructure
- Less technical expertise and IT requirements

→ Library data is more extensively linked through centralised library data

- Linking bibliographic data within an aggregated dataset is easier than across distributed datasets
- Each library benefits from the linking done for other libraries
- Each external dataset linked to, benefits all libraries' data



# Building LOD services at The European Library

## Current Status

→ The European Library Open Dataset was first published in June 2014

- Over 60 million Bibliographic Resources from 15 countries
- Likely to double during 2014, and reach 200 million during 2015
- Data distribution agreements with partner libraries are underway

→ <http://www.theeuropeanlibrary.org/tel4/access>

The screenshot displays the 'Access Our Data' page of The European Library. The header includes navigation links: About, Membership, For Current Partners, Log In, and English (en). A search bar is present with the text 'Search all European Library content...' and a 'GO' button. Below the header, the page is divided into several sections:

- About The European Library datasets:** This section explains that the library offers a unique set of data from Europe's libraries and provides access to raw data via the Open Search API.
- How to access the datasets:** This section guides users on how to sign up for an API key and access the data.
- Available APIs and Datasets:** This section lists four available services:
  - The European Library API v2.0:** Offers access to bibliographic and digital collections from Europe's national and research libraries.
  - The European Library Open Dataset:** Offers access to a vast set of bibliographic data made available by its partner libraries.
  - Metadata Harvesting - OAI-PMH Repository:** Provides access to its datasets metadata harvesting applications via an OAI-PMH v2.0 compliant repository.
  - RLUK Linked Open Data:** Provides access to the RLUK Database, an established Union Catalogue provided by Research Libraries UK.

On the right side of the page, there are links for 'Get access / login', 'About APIs', 'The European Library Data Forum', and 'FREQUENTLY ASKED QUESTIONS'. At the bottom, there is a 'CURRENT NEWS' section with recent updates.

# Linking library data with the wider cultural heritage domain – WWI example

- Integration of library data with other cultural heritage data from specific communities
- Europeana 1914-1918 (<http://www.europeana1914-1918.eu/>) developed its own vocabulary based on a subset of LCSH
  - Terms translated in 10 languages and linked to id.loc.gov
  - Published in SKOS via the OpenSkos vocabulary service
- Future work
  - Re-use of the vocabulary by WW1 communities
  - Alignment with other vocs to make our collections more interoperable
  - Full integration of SKOS data in the portal



# Challenges ahead

Continue populating the “semantic layer”....

- Creating synergies between domains to create LOD at low cost
- Focus on cross-domain multilingual vocabulary alignment and publish the results as Linked Data
  - Should we develop MACS as LOD and consider it as an equivalent of VIAF for concepts resources?
  - More pivot vocabularies such as AGROVOC, STW Thesaurus for Economics integrated in The European Library and in Europeana



# Thank you !

[valentine.charles@europeana.eu](mailto:valentine.charles@europeana.eu)

[nuno.freire@theeuropeanlibrary.org](mailto:nuno.freire@theeuropeanlibrary.org)

[antoine.isaac@europeana.eu](mailto:antoine.isaac@europeana.eu)



think culture



The  
European  
Library

*Connecting knowledge*

14<sup>th</sup> August 2014, IFLA2014 satellite meeting, Paris

