Application of LOD to Enrich the Collection of Digitized Medieval Manuscripts at the University of Valencia

José Manuel Barrueco

barrueco@uv.es

Cristina García Testal

testal@uv.es

University of Valencia (Spain)

Contents:

2: The UV manuscripts collection:

- **The UV ancient books collection:**
  - Manuscritps: +1,100 volumes going back to the XIII century
  - Incunabula: 334 volumes
  - Printed books (XVI – XIX centuries): +40,000 volumes
- **The UV has been involved in digitization projects since 2000.**
- **Partner in the Europeana Regia project (2010-2012):**
  - EU founded project to create a virtual library with the most important European royal collections of documents from the Middle Ages to the Renaissance.
  - Bibliothèque nationale de France, Bayerische Staatsbibliothek, Herzog August Bibliothek and the Koninklijke Bibliotheek van België.
  - http://www.europeanaregia.eu
  - The UV contributes with **92 codex** (Royal Library of the Aragonese Kings of Naples).
  - **They have been used as test bed for this work**

- Explore the oportunities of LOD to enrich the collection of digitized medieval manuscripts by providing additional information about authors:
    - Name (with variations)
    - Occupation (Historian, Poet…)
    - Biography
    - Picture
    - Main works

- Integrate LOD into a productive library application:
    - book viewer for digitized matherials.

- Analyze the problems faced by institutions whiling to consume LOD:
    - Availability of data sources
    - Licenses used
    - Technical issues …
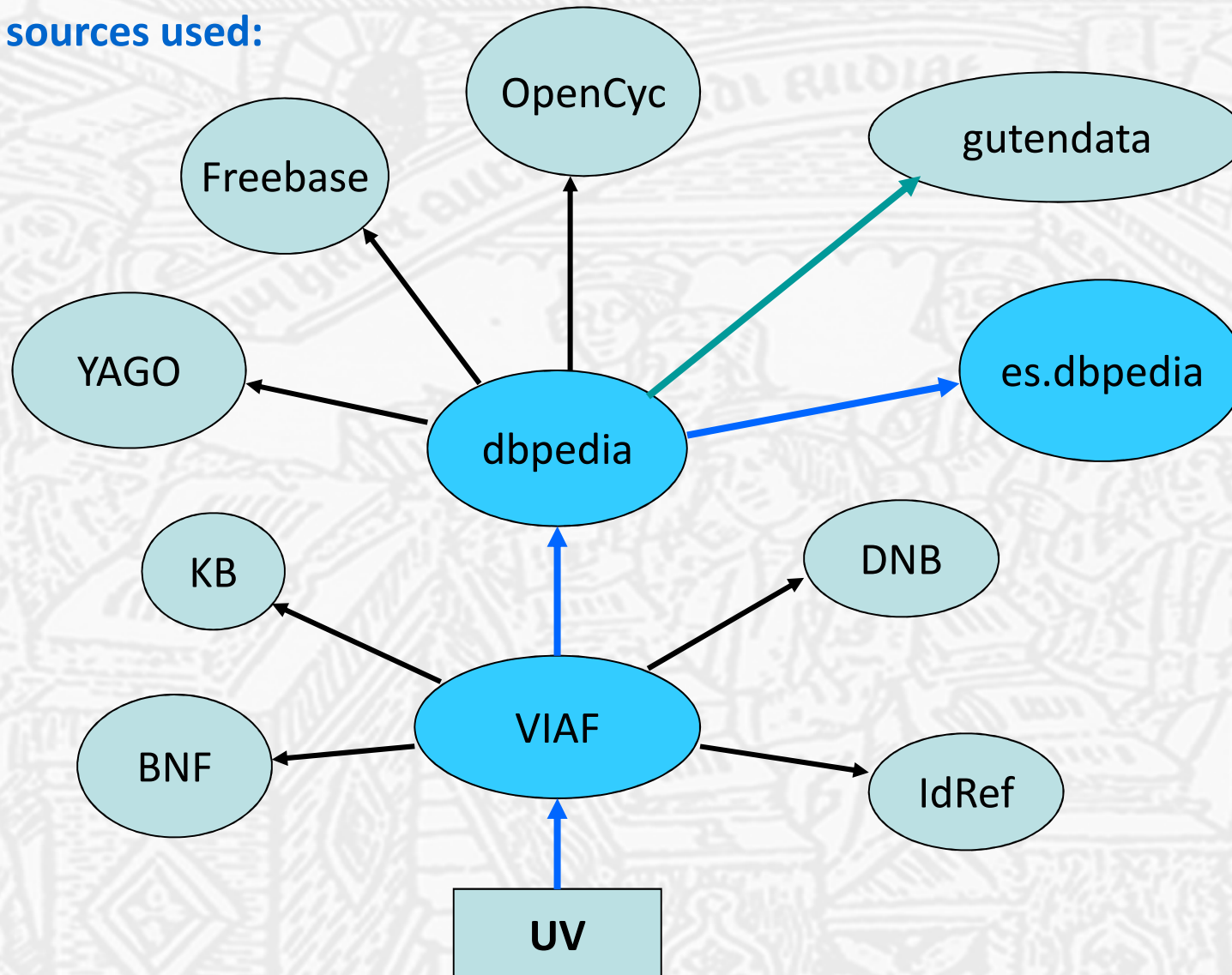
3: Implementation at the UV:

What?

- We want to provide for each author (at least):
  - Name (with variations)
  - Occupation (Historian, Poet...)
  - Biography
  - Picture
  - Main works

How?

- Integration of the data in the book viewer
- Not storing locally any RDF data
- Working with XML -> HTML conversions on the fly
- Storing the resulting data as HTML to present to the user
- Crawling the web of data
- Starting point: VIAF
  - Including VIAF URIs in the authority records of the institutional repository

SWIB13
Semantic Web in Libraries

- **Data sources used:**

OpenCyc

gutendata

Freebase

YAGO

dbpedia

es.dbpedia

KB

DNB

BNF

VIAF

IdRef

UV

3: Implementation at the UV:

- **Digitized books included in institutional repository:**
  - DSpace with locally developed book viewer:
    - METS metadata + JP2000 image files
    - XSLT (METS -> HTML)
    - IIP image server
- **Application developed (perl + xslt):**
  - **Input : VIAF URI for each author or contributor**
  - Dereference URI:
    - Take name variations in *foaf:name*
    - Dereference *owl:sameAs* link to dbpedia if exist
      1. Take *dbpedia-owl:abstract* [en|es|ca]
      2. Take *foaf:depiction* or *dbpedia-owl:thumbnail*
      3. Take *dbpedia-owl:occupation* (follow URI until spanish label)
      4. Take *dbpprop:notableWorks* (follow URI to works description)
      5. Dereference *owl:sameAs* link to es.dbpedia if exists
         - Repeat 1-4 completing missing data
  - **Output: HTML static page with the description of the author**

Application of LOD to enrich the collection of digitized manuscripts …

3: Implementation at the UV:

RODERIC
REPOSITORI DE CONTINGUT LLIURE

VNIVERSITAT ID VALÈNCIA

SOMNI

Valencià   Castellano

FAQ   TESIS   El meu RODERIC   Contacto

Inicio RODERIC ➤ Somni ➤ Manuscrits (Duc de Calàbria) ➤ Ver documento

Buscar en RODERIC

Ir   Búsqueda avanzada

◉ Manuscrits (Duc de Calàbria)
○ Buscar en RODERIC

Todo RODERIC
Comunidades y colecciones
Títulos
Autores
Temas
Tipo de documento

Esta colección
Títulos
Autores
Temas
Tipo de documento

Ver estadísticas

[Publii Vergilii Maronis Opera]

Ver estadísticas   RefWorks

Virgili Maró, Publi, 70-19 aC

Virgili Maró, Publi, 70-19 aC. Bucolica; Virgili Maró, Publi, 70-19 aC. Georgica; Virgili Maró, Publi, 70-19 aC. Aeneis; Donat, Elius. Publii Virgilii Maronis vita; Servi; Mestre d'Ippolita, fl. 1459-, il.

Este documento es un/a libro manuscrito, creado en: 1465.

Lugar: Itàlia Milà

Lengua: la

Descripció física: [328] f. : il. ; 433 x 280 mm enquadernació 450 x 310 mm

Soporte: Vitel·la

Tipo: Manuscrit

Resumen: Manuscrit copiat en 1465 i il·luminat amb motiu de les noces d'Ippolita Sforza amb Alfons II de Nàpols, transcriu les Bucòliques, Geòrgiques i l'Eneida. Aprofitant la copia hom afegí al començament la biografia de Virgili atribuïda a Elius Donat. S'hi inclouen els comentaris de Servi. Precedint l'Eneida hom troba uns versos amb el contingut de tots els llibres que segons la rúbrica escriví Ovidi, i que, però, son coneguts també com "Basilii de duodecim libris Aeneidos". Finalment, hom pot llegir també a l'inici de cadascun dels llibres de la Eneida l'argument escrit per Ovidi **Fuente:** Universitat de València. Biblioteca Històrica

Signatura:
BH Ms. 0891
Olim 780

Identificadores:
http://roderic.uv.es/uv_ms_0891
http://hdl.handle.net/10550/23142

Colección: Europeana Regia. Biblioteca dels reis aragonesos de Nàpols

Procedencia: Biblioteca de Alfonso, duque de Calabria; Biblioteca del Duque de Calabria; "Librería de San Miguel de los Reyes"

Adquisición: Desamortització, 1837

Descripción general: Ms. il·luminat
Títol donat pel catalogador
Col·lació: Vitel·la, f. ii (paper) + 328 + f. ii (paper) ; foliació moderna a llapis ; reclams horitzontals i verticals cada huit fulls
Composició: Escrit en vers el text, i el comentari a línia tirada
Escriptura: Humanística antiqua el text, i humanística cursiva el comentari. Espais en blanc deixats pel copista per a les paraules que deurien escriure's en grec
Decoració: Al frontispici (f. 1r) orla miniada en or i colors formada per fulls de palma i corona que emmarquen coloms; una d'aquestes duu el lema "A bon

3: Implementation at the UV:

3: Implementation at the UV:

4: Results:

- **92 manuscripts used as test bed:**
  - **97 authors and coauthors with VIAF URIs:**
    - **73 main authors and 24 scribes, illuminators, miniaturists…**

	All authors	Main Authors
Name forms (from VIAF)	**97 (100%)**	**73 (100%)**
Biography (from dbpedia)	**37 (38.14%):** 33: in English and Spanish 4: only in English	**37 (50.68%):** 33: in English and Spanish 4: only in English
Picture (from dbpedia)	**31 (31.95%)**	**31 (42.46%)**
Occupation (from dbpedia & es.dbpedia)	**8 (8.24%):** 7 spanish label	**8 (10.95%):** 7 spanish label
Main works (from dbpedia & es.dbpedia)	**5 (5.15%):** 2 URIs to works description	**5 (6.84%):** 2 URIs to works description
Main works (from gutendata)	*10 (10.30%)*	

5: Questions for LOD consumers:

- **How to know the data sources available?**

  - Need for registries of linked data sets

  - The datahub (http://datahub.io)

- **What datasets are available for use?**

  - September 2011: Bizer et al. [1] identified 295 linked open datasets

  - October 2013: 8,920 datasets, **891 of them are LOD** (10%)

  - That sounds good! but .....

    - bioportal ontologies: 244 datasets

    - rkb-explorer: 55 datasets

    - **~ 594 LOD data sets**

[1] Bizer, C. ; Jentzsch, A ; Cyganiak, R. State of the LOD Cloud. Version 0.3, 09/19/2011. http://lod-cloud.net/state/

- **What is the scope of the available datasets?**
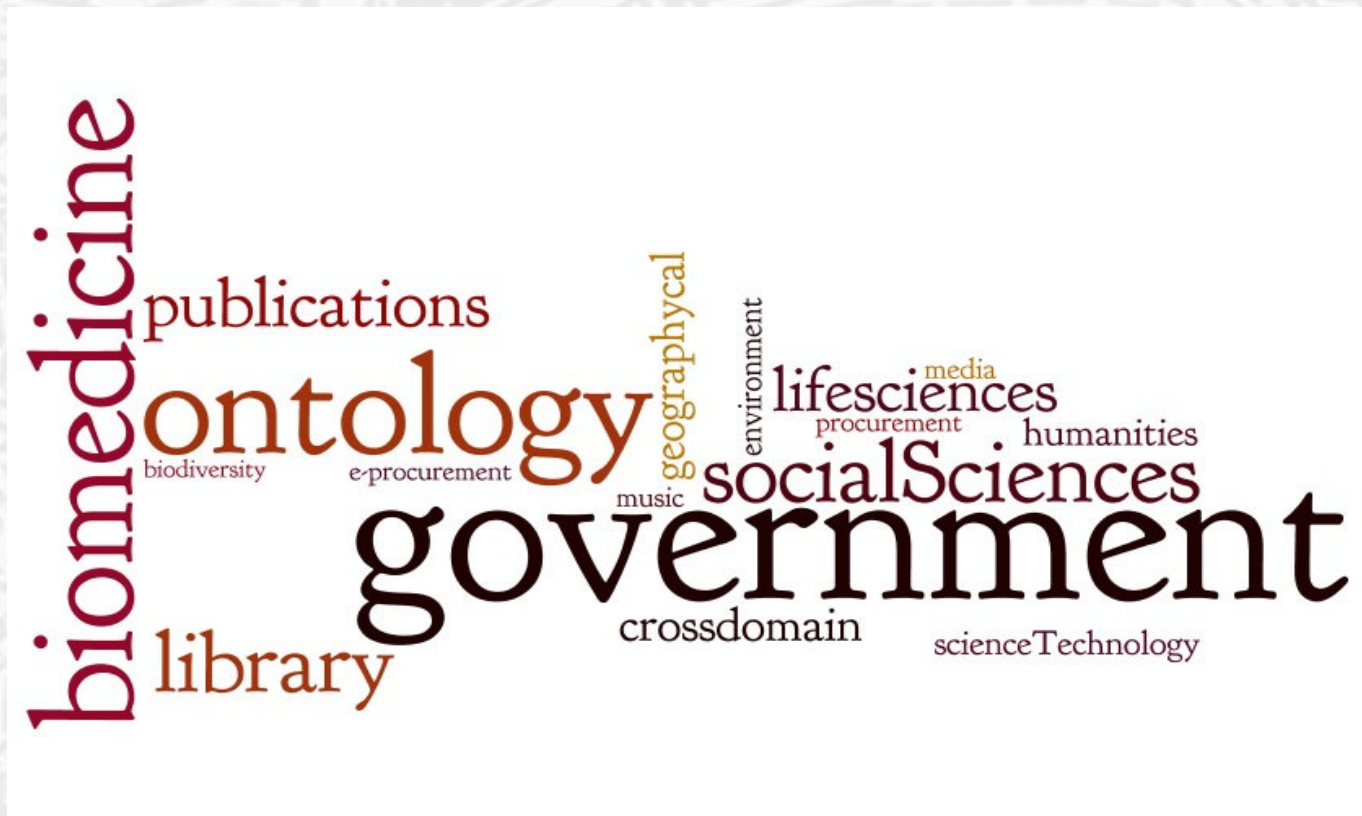
5: Questions for LOD consumers:

- **Are they compliant with best practices for data provisioning?**

RDF links pointing at other data sources:

Out-Links	Datasets (Sep 2011)	Datasets (Sep 2013)
No links	30 (10.17 %)	212 (23.79%)
up to 1,000	90 (30.51 %)	243 (27,27%)
1,000 to 10,000	58 (19.66 %)	190 (21.32%)
10,000 to 100,000	45 (15.25 %)	135 (15.15%)
100,000 to 1,000,000	43 (14.58 %)	69 (7.74%)
more than 1,000,000	29 (9.83 %)	42 (4.71%)
	295	891

Provide dataset-level metadata:

	Datasets (Sep 2011)	Datasets (Sep 2013)
voiD descriptions	95 / 295 (32.20 %)	222 / 891 (24.91 %)
Sitemaps	53 / 295 (17.97 %)	87 / 891 (9.76 %)
voiD description or sitemaps	109 / 295 (36.95 %)	243 / 891 (27.27 %)
voiD description and sitemaps		66 / 891 (7.40 %)
nothing	186 / 295 (63.05 %)	648 / 891 (72.72 %)
	295	891

Application of LOD to enrich the collection of digitized manuscripts …

5: Questions for LOD consumers:

• **Licenses used to distribute the data:**

- **In which way can we use the data sets?**
- **Are they really open?**

- 197 (22.33%) datasets without license information
- 694 (77.89%) with some type of license

License type	Datasets (Sep 2013)
Undefined license model (open)	287 (41.48%)
Creative Commons	277 (39.91%)
Open Data Commons	92 (13.25%)
Undefined license model (not open)	49 (5.49%)
UK Open Government Licence	23 (3.31%)
ukcrown-withrights	6 (0.86%)
GNU Free Documentation License	6 (0.86%)
General Public License	2 (0.28%)
apache	1 (0.14%)
	694

Application of LOD to enrich the collection of digitized manuscripts …

- **As consumers of LOD we would ask for:**

  - comprehensive registries of data sources

  - comprehensive metadata at data set level

  - licenses following any of the available models (CC, DC, …)

  - more *owl:sameAs* links to interconect data islands

  - … more data sets

- **As librarians implementing an application of LOD:**

  - we have been able to easily develop an application to integrate LOD in a collection of our institutional repository

  - Enrich the collection of manuscripts providing biographical information for almost half of the authors

  - Our future plan:

    - to extent the coverage to other matherials starting by early printed books

![SWIB13 — Semantic Web in Libraries]

Thanks for your attention!

more information:
http://somni.uv.es

José Manuel Barrueco
barrueco@uv.es
Cristina García Testal
testal@uv.es