

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)

1. The UV manuscripts collection
2. What are we trying to do
3. Implementation of LOD at the UV
 - Data sources used
 - Application development
 - How it looks like
4. Results
5. Questions for LOD consumers:
 - Data sources available
 - Quality of the data
 - Licenses used
6. Conclusions

2: The UV manuscripts collection:

- **The UV ancient books collection:**
 - Manuscripts: +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**

1: What are we trying to do:

- Explore the opportunities 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 materials.
- Analyze the problems faced by institutions willing to consume LOD:
 - Availability of data sources
 - Licenses used
 - Technical issues ...

3: Implementation at the UV:

What?

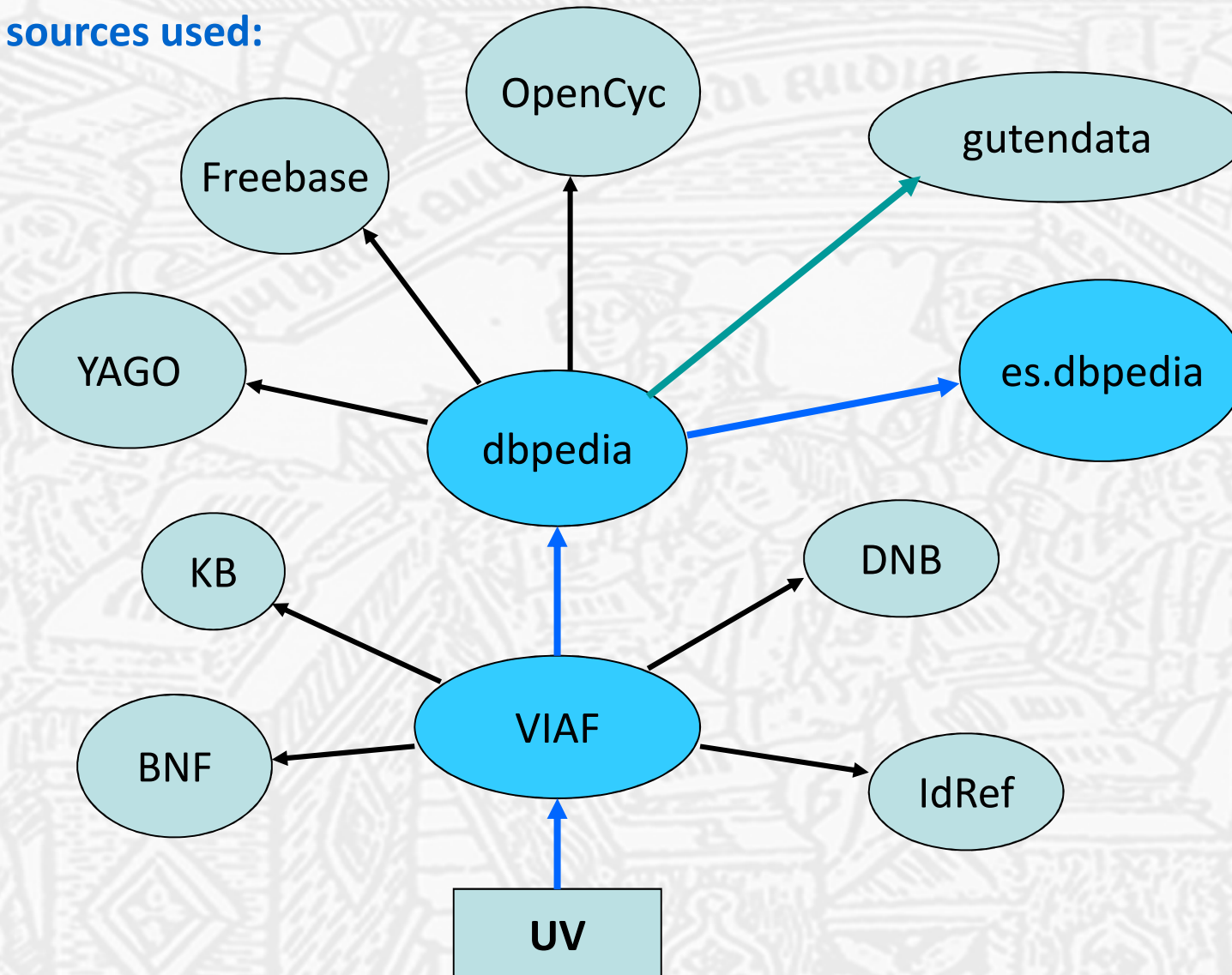
- We want to provide for each author (at least):
 - Name (with variations)
 - Occupation (Historian, Poet...)
 - Biography
 - Picture
 - Main works

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

3: Implementation at the UV:

- Data sources used:



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**

3: Implementation at the UV:

The screenshot shows the Roderic website interface. At the top, there is a navigation bar with the Roderic logo and the text 'REPOSITORI DE CONTINGUT LLIURE'. Below this, there are language options (Valencià, Castellano) and utility links (FAQ, TESIS, El meu Roderic, Contacto). The main content area displays the title '[Publii Vergilii Maronis Opera]' and a search bar. A sidebar on the left contains navigation menus for 'Todo Roderic' and 'Esta colección'. The main record for 'Virgili Maró, Publi, 70-19 aC' includes a thumbnail of a manuscript page, a description in Catalan, and detailed metadata in Spanish. The metadata includes physical description, support, type, summary, source, signature, identifiers, collection, provenance, acquisition, general description, and decoration.

Valencià Castellano

FAQ TESIS El meu Roderic Contacto

Inicio Roderic > Somni > Manuscrits (Duc de Calàbria) > Ver documento

Buscar en Roderic

Búsqueda avanzada

Manuscrits (Duc de Calàbria)

Buscar en Roderic

Todo Roderic

Comunidades y colecciones

Títulos

Autores

Ternas

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 escriu 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>

Colecció: Europea Regia. Biblioteca dels reis aragonesos de Nàpols

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

Adquisició: Desamortització, 1837

Descripció 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

Esriptura: Humanística antiga el text, i humanística cursiva el comentari.

Espais en blanc deixats pel copista per a les paraules que deuriu escriure's en grec

Decoració: Al frontispici (f. 1r) orla miniada en or i colors formada per fulls de palma i corona que emmarquen columnes d'arxuesques d'una i l'altre costat del lema "A hon

Application of LOD to enrich the collection of digitized manuscripts ...

3: Implementation at the UV:

The screenshot shows the Roderic digital library interface. At the top, there is a header with the Roderic logo and the text 'REPOSITORI DE CONTINGUT LLIURE'. Below the header, there is a navigation bar with 'Valencià Castellano' and 'Somni' buttons. The main content area displays the title '[Publii Vergilii Maronis Opera]' and a search bar. A browser window is overlaid on the page, showing the manuscript viewer. The browser window title is 'Espill v12: Visualitzador per a continguts digitalitzats a Somni - Google Chrome'. The browser address bar shows 'webliboteca.uv.es/cgi/view.pl?source=uv_ms_0891'. The browser window content shows the manuscript viewer interface with the title '[Publii Vergilii Maronis Opera] (1465)'. Below the title, there is a list of authors and their dates: 'Virgili Maró, Publi, 70-19 aC [més info]', 'Virgili Maró, Publi, 70-19 aC [més info]', 'Virgili Maró, Publi, 70-19 aC. [més info]', 'Donat, Elius. [més info]', 'Servi [més info]', and 'Mestre d'Ippolita, fl. 1459, il. [més info]'. The browser window also shows a list of contents and a thumbnail of the manuscript page. The manuscript page is titled 'VITA VIRGILII MARONIS' and features a large illuminated initial 'V' and a decorative border. The page number 'f. 1 r' is visible at the bottom right of the manuscript image.

Virgili Maró, Publi, 70-19 aC
<http://viaf.org/viaf/8194433/>

Esriptura: Humanística antiga el text, i humanística cursiva el comentari.
Espais en blanc deixats pel copista per a les paraules que deurién escriure's en grec
Decoració: Al frontispici (f. 1r) orta miniada en or i colors formada per fulls de palma i corona que emmarquen coloms: una d'aquestes duri el lema "A hon

3: Implementation at the UV:

The screenshot displays the RODERIC (Repositori de Contingut Lliure) digital library interface. The main header features the RODERIC logo and the text 'REPOSITORI DE CONTINGUT LLIURE'. A navigation bar includes language options (Valencià, Castellano) and links for FAQ, TESIS, El meu RODERIC, and Contacto. The breadcrumb trail reads: 'Inici RODERIC > Somni > Manuscrits (Duc de Calàbria) > Ver documento'. A search bar is present with the text 'Buscar en RODERIC' and a 'Búsqueda avanzada' button. The left sidebar contains navigation links for 'Comunidades y colecciones', 'Títulos', 'Autores', 'Temas', and 'Tipo de documento', as well as 'Esta colección' and 'Ver estadísticas'. The main content area shows a manuscript viewer for '[Publii Vergilii Maronis Opera] (1465)'. A browser window titled 'Espill v12: Visualitzador per a continguts digitalitzats a Somni - Google Chrome' is overlaid on the page, displaying the manuscript's content. A popup window titled 'Informació sobre Virgil' is open, providing biographical information about Publio Virgilio Marón. The popup includes a photograph of a marble bust of Virgil and two columns of text: one in Spanish and one in English. The Spanish text describes Virgil as a Roman poet, author of the Aeneid, Bucolics, and Georgics, and his influence on Dante's Divine Comedy. The English text provides a similar overview, mentioning his birth date (October 15, 70 BC) and death (September 21, 19 BC).

Application of LOD to enrich the collection of digitized manuscripts ...

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
<i>Main works</i> (from gutendata)	10 (10.30%)	

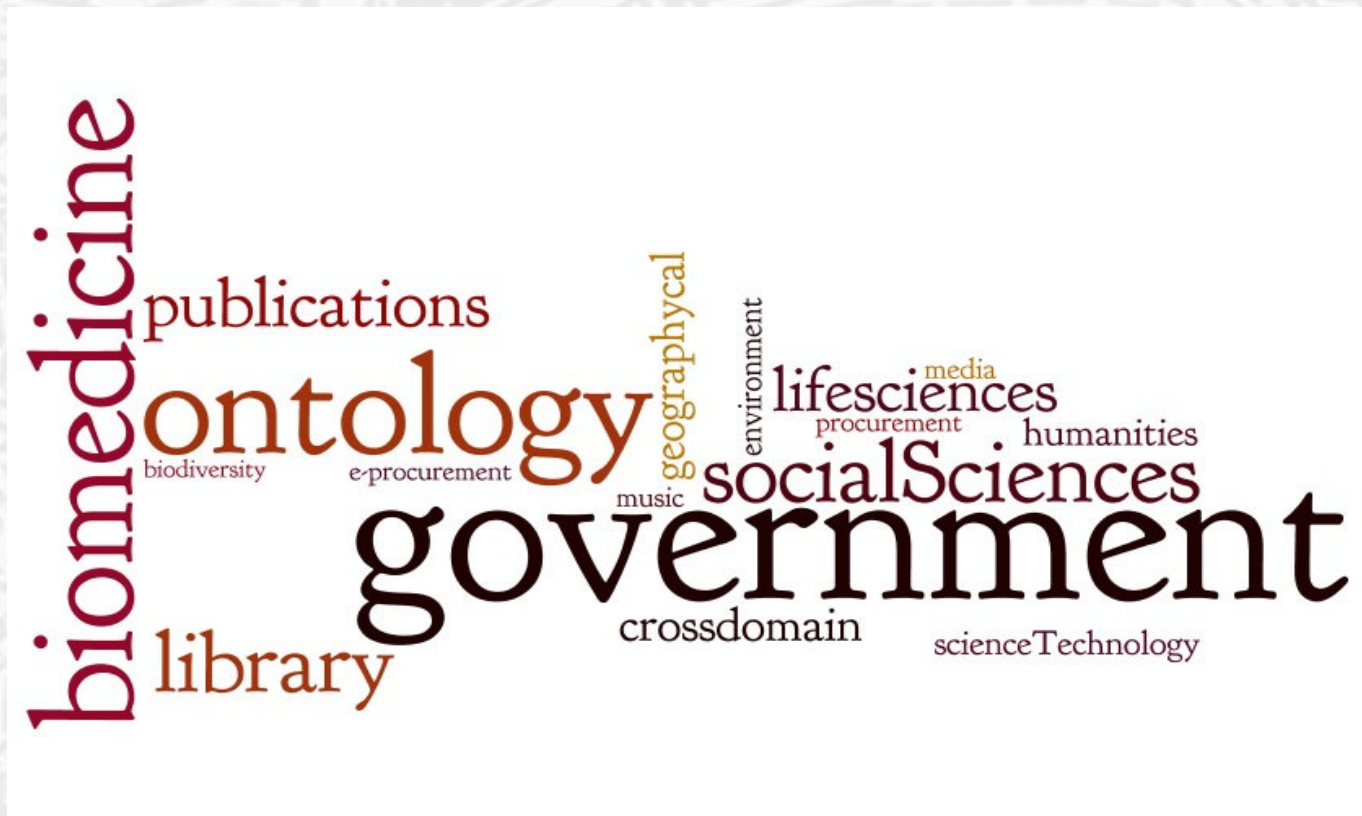
5: Questions for LOD consumers:

- **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/>

5: Questions for LOD consumers:

- 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

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

6: Conclusions:

- **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 interconnect 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

Thanks for your attention!

more information:

<http://somni.uv.es>

José Manuel Barrueco

barrueco@uv.es

Cristina García Testal

testal@uv.es