

Status Quo and (current) limitations of Library Linked Data

Asunción Gómez-Pérez¹, Philipp Cimiano² and
Daniel Vila-Suero¹

¹Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

<http://www.oeg-upm.net>
asun,dvila@fi.upm.es

²AG Semantic Computing, CITEC, University of Bielefeld
cimianog@cit-ec.uni-bielefeld.de

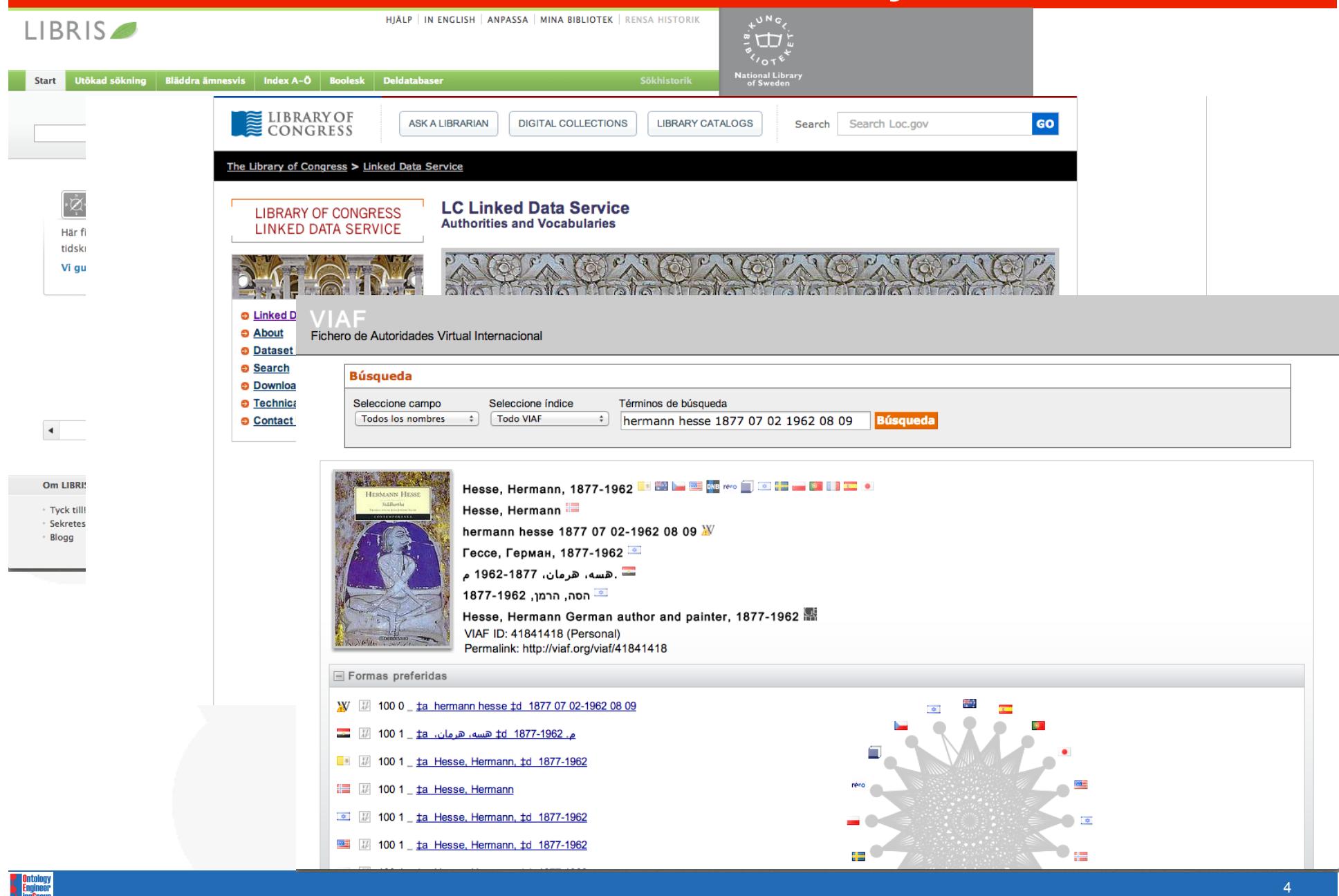
Acknowledgments: BNE team, DNB team, Uldis Bojars (NLL), Jodi Schneider (NUIG) and others

SWIB12 – Cologne, 28-11-2012

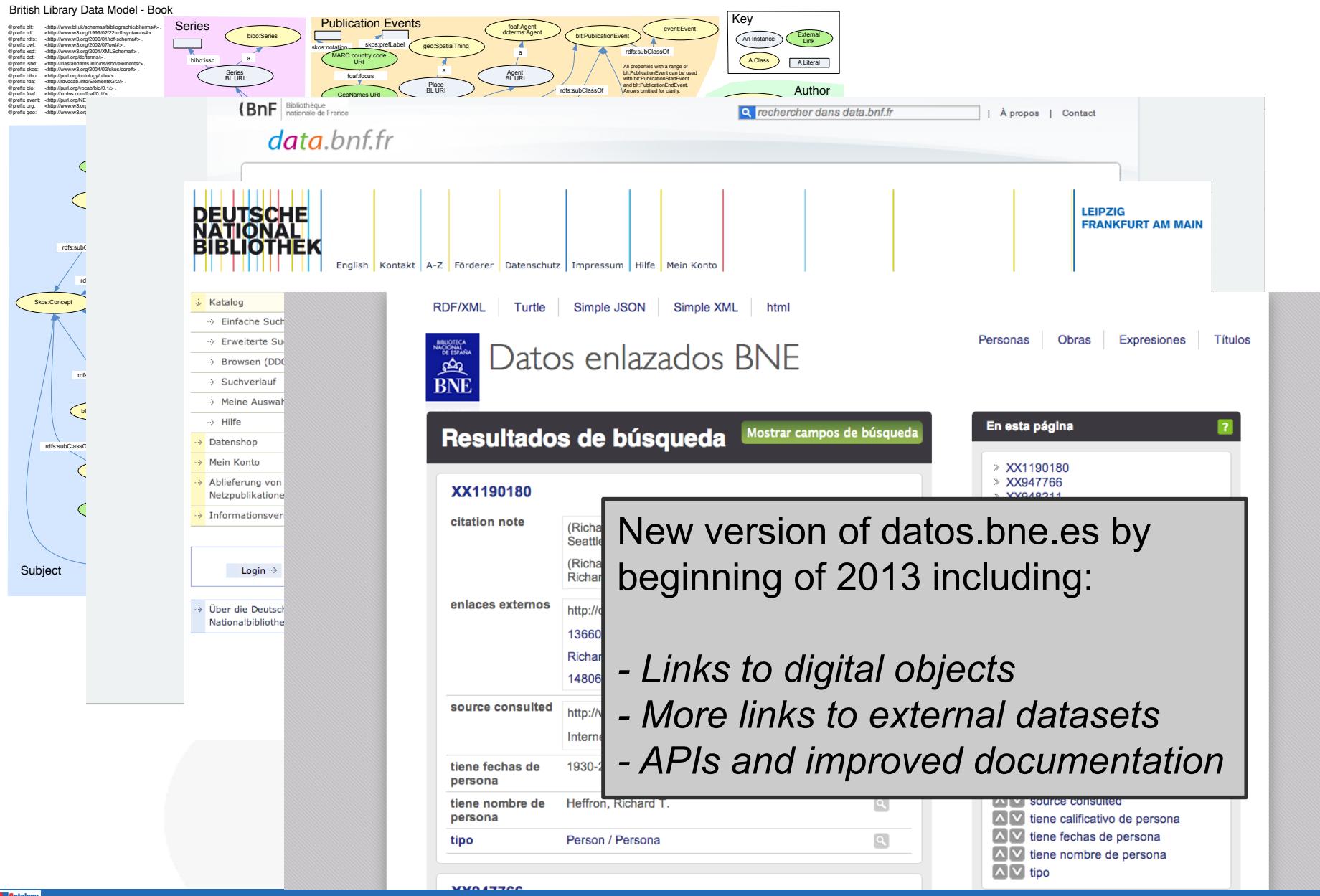
- The Library Linked Data Cloud
- A (library) Linked Data Life-cycle
- A collection of current limitations
- Conclusions

Library Linked Data cloud

Library Linked Data Cloud



Library Linked Data Cloud



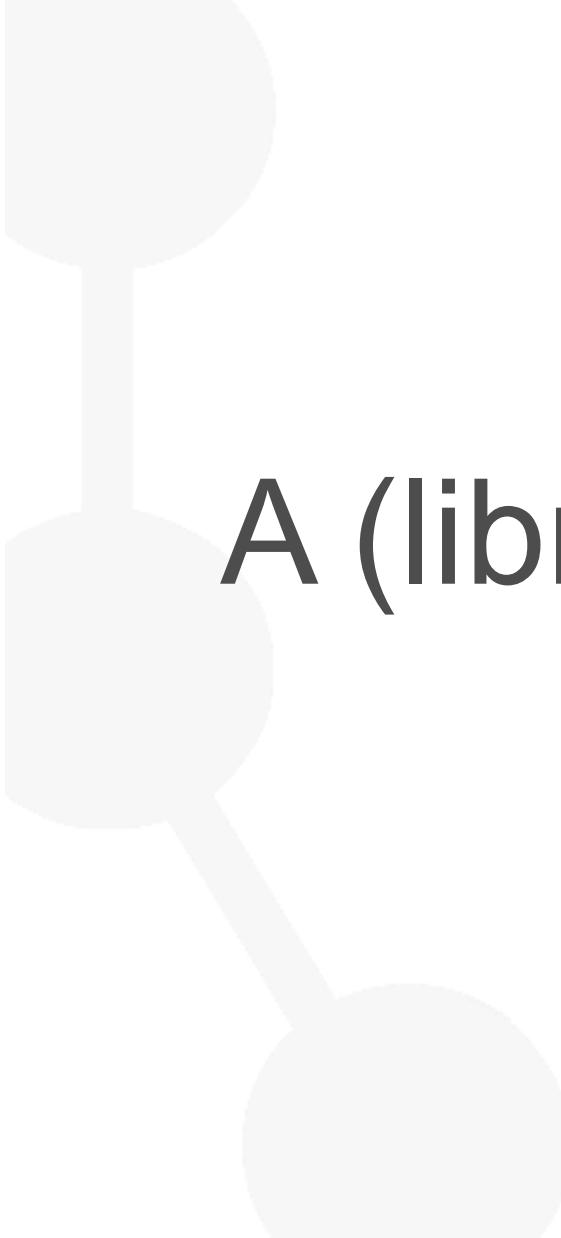
Library Linked Data Cloud

The screenshot shows the Europeana professional homepage. At the top, there's a banner with the Europeana logo and several horse silhouettes. Below the banner is a navigation bar with links: Home, About us, News & Events, Projects, Europeana Foundation, Europeana Network, Provide data, and Re-use data. The main content area features a large image of a person's face. On the left, there's a sidebar with a menu: Re-use data, API services, Linked Open Data (with sub-links: In brief, Tech data, Datasets, Credits, Case studies, Cultural Co..., Hackathons), and T... S... (partially visible). The central part of the page displays several cards: 'CultureSampo' (Finnish Culture on the Semantic Web 2.0, note: the portal is only partially available in English, sorry), 'Map Browse and Search', 'Relation Search', 'Search and Organize', 'Collections', and 'Finnish History 1860-1919'. At the bottom right, there are language links: Language: suomi svenska English.

And many others (see <http://thedatahub.org/group/lld>)

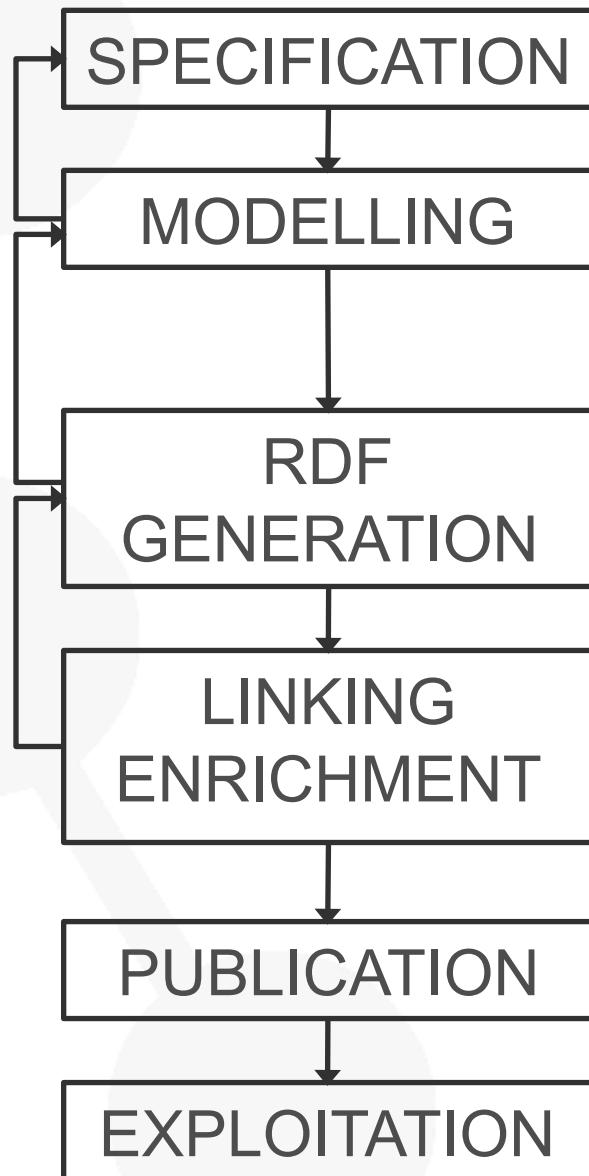
The screenshot shows the WorldCat website. At the top, there's a blue header with a search bar and links: Búsqueda avanzada | Encontrar una biblioteca, Comenzar, ¿Olvidó su contraseña?, and Establecer una cuenta gratuita. Below the header, there's a message: Esperando s7.addthis.com... and Last Viewed: video : Kätevät: hillihango. The main content area has several sections: 'Usar herramientas de WorldCat' (Learn more in a library, Widgets y servicios web, WorldCat Search API, Aplicaciones y plug-ins del navegador, Enlazar a WorldCat, Agregar una chapa WorldCat), 'Construir una bibliografía' (Video tutorial, Cómo usar citas de WorldCat (YouTube), Duración:1:47), 'Bibliotecas más populares', and 'Etiquetas más populares'. At the bottom, there's a footer with the Ontology Engineering Group logo and a page number: 6.

- **Availability** of Library Linked Data is already a reality
- Many **serious efforts**: id.loc.gov, VIAF, DNB, data.bnf.fr, Bibliographic Framework Initiative etc.
- Still **several challenges** and **limitations** preventing LLD full potential



A (library) Linked Data life-cycle

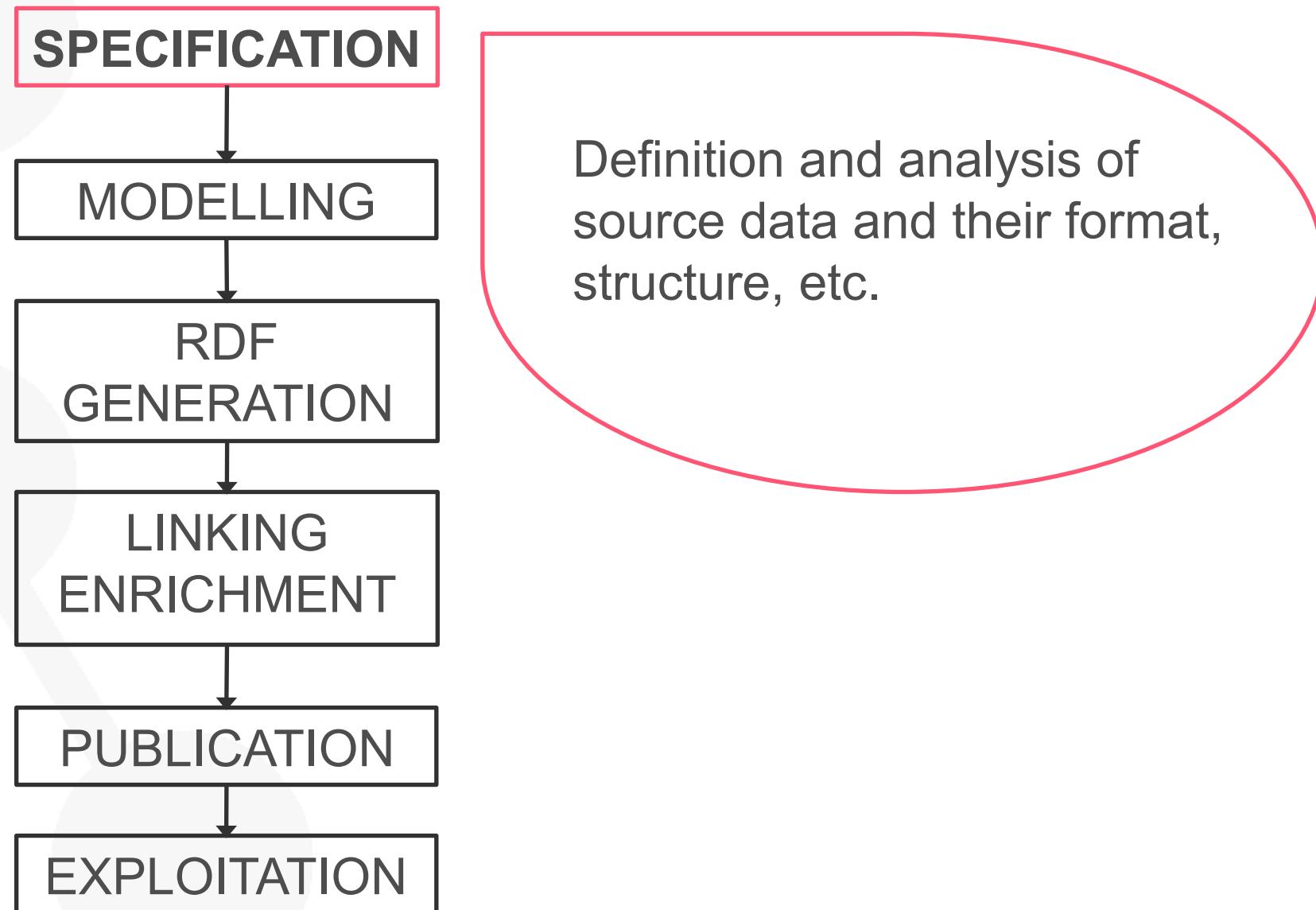
A (Library) Linked Data Life-cycle



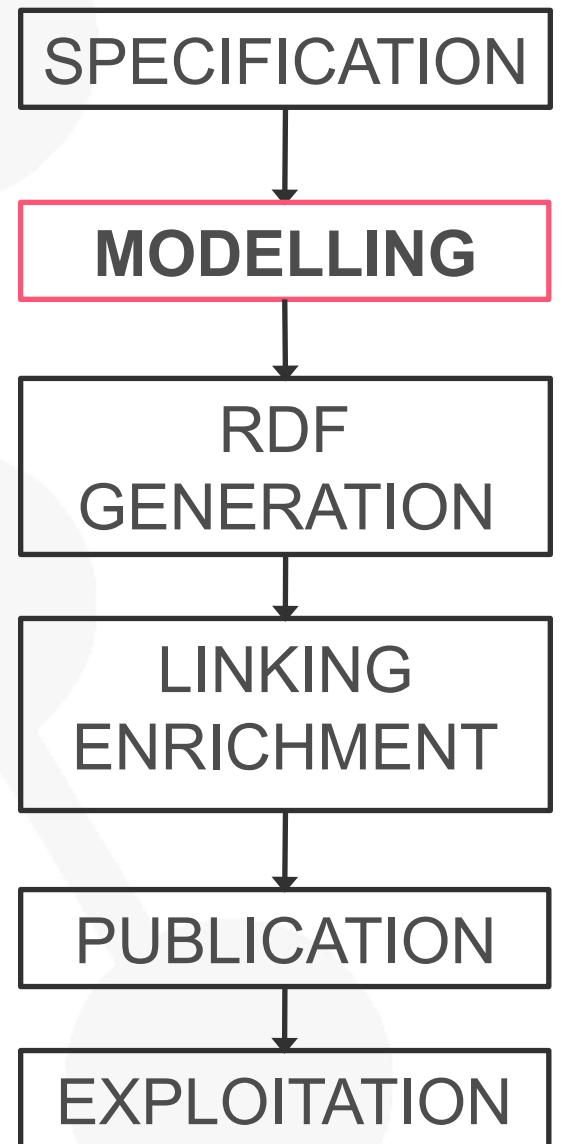
- A series of steps or phases
- Based on *Villazón-Terrazas et al.**
- **Others:** LOD2, Datalift, etc. (see http://www.w3.org/2011/gld/wiki/GLD_Life_cycle)

* *Methodological Guidelines for Publishing Government Linked Data,*
Boris Villazón-Terrazas, Luis. M. Vilches-Blázquez,
Oscar Corcho and Asunción Gómez-Pérez
Linking Goverment Data Book

A (Library) Linked Data Life-cycle

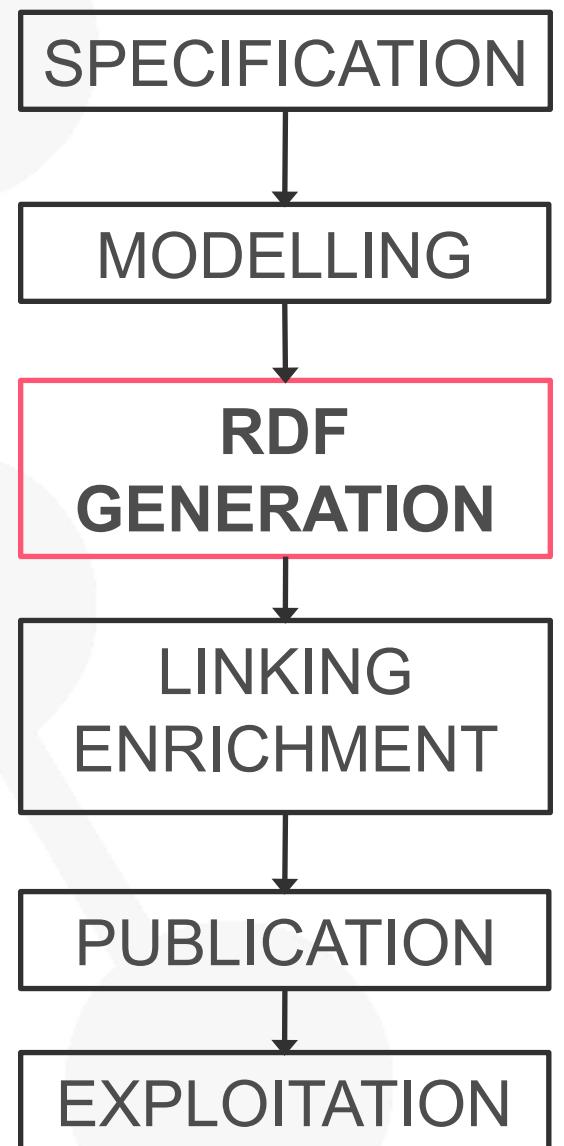


A (Library) Linked Data Life-cycle



- Selection and reuse of vocabularies to represent LD.
- Creation of new local terms and mapping to existing vocabularies

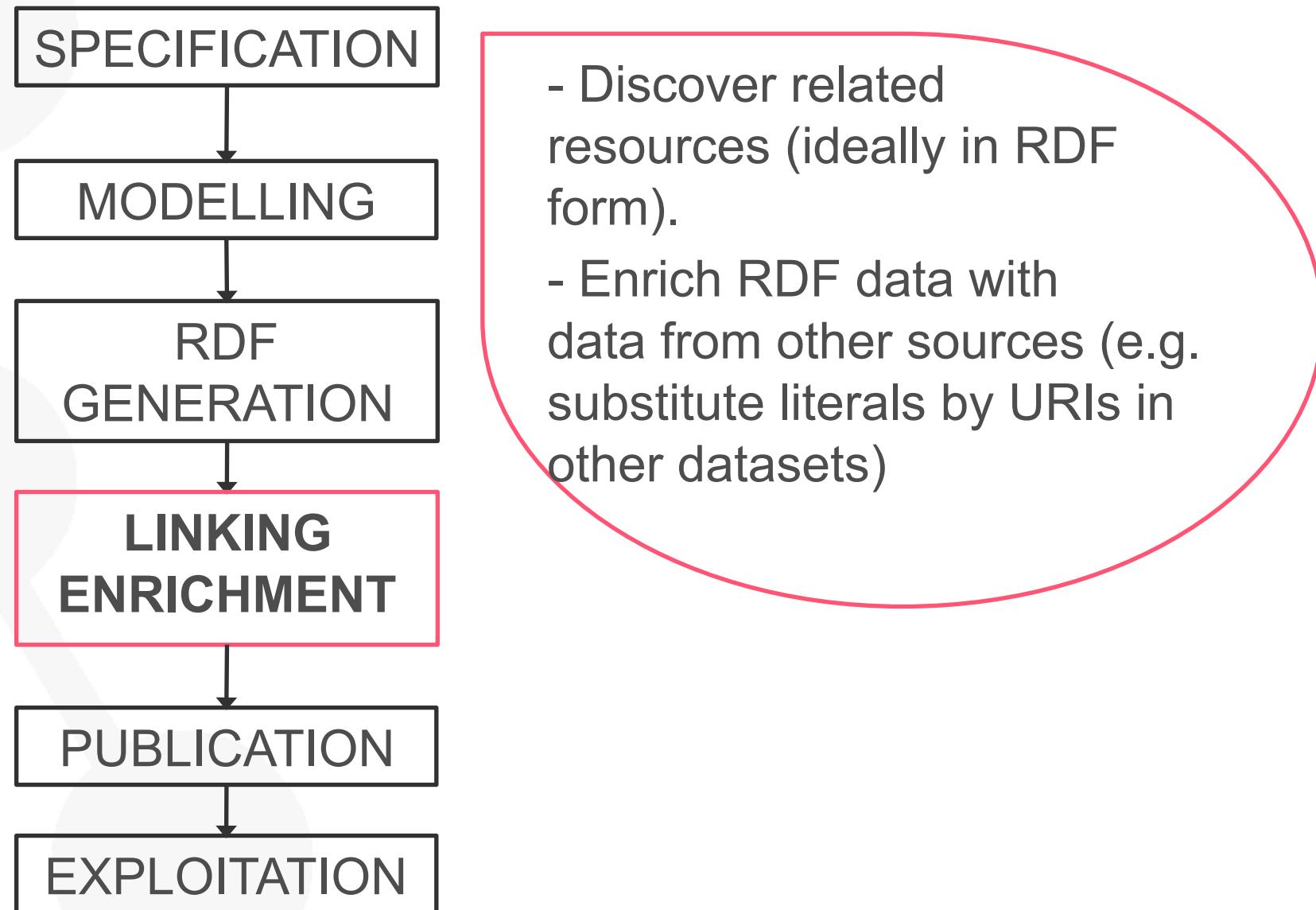
A (Library) Linked Data Life-cycle



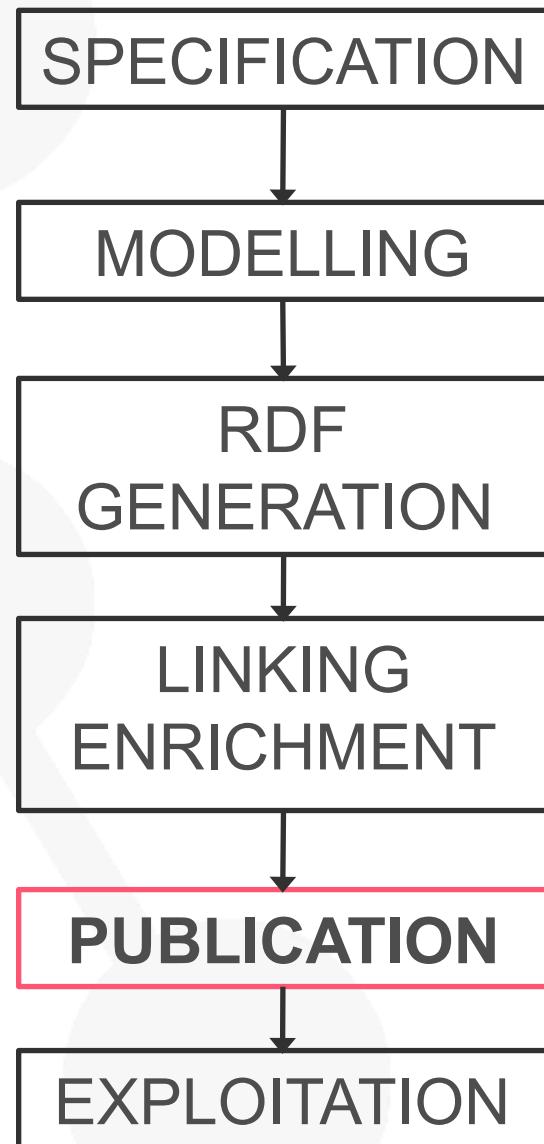
Taking source data, and vocabularies:

Mapping (cross-walk) source data to produce RDF

A (Library) Linked Data Life-cycle

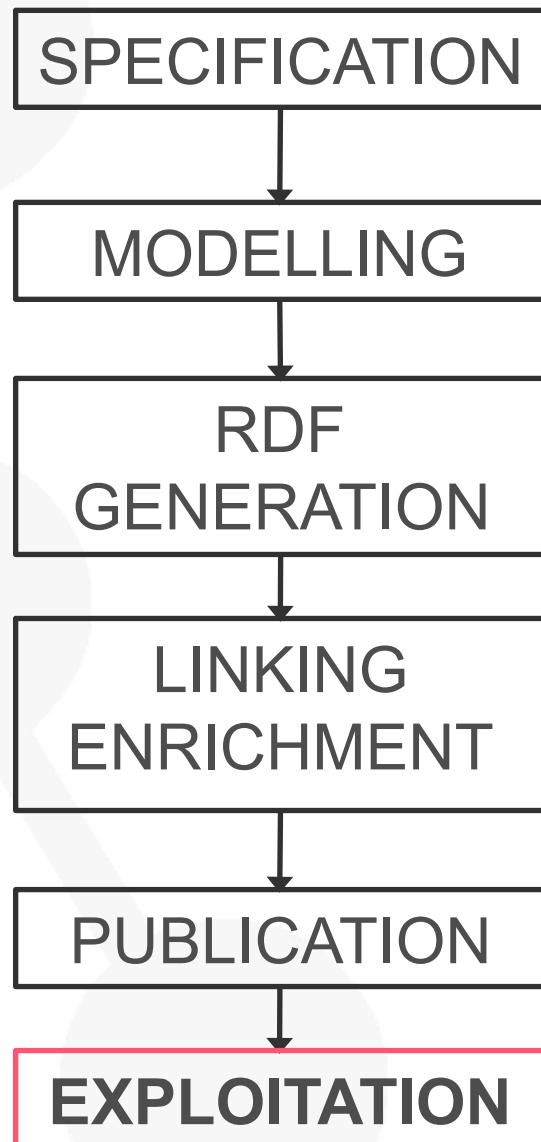


A (Library) Linked Data Life-cycle

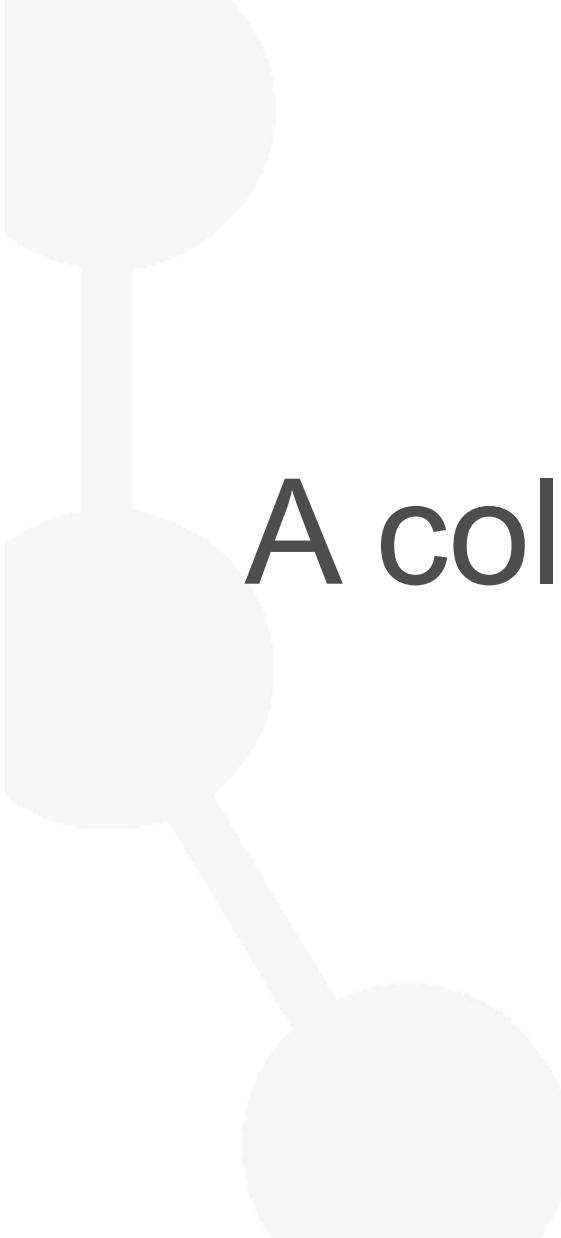


- Setup the infrastructure to expose your data to the Web (SPARQL, APIs, dumps).
- Enable discovery of your data (sitemap, void)
- Include data provenance, license, etc.

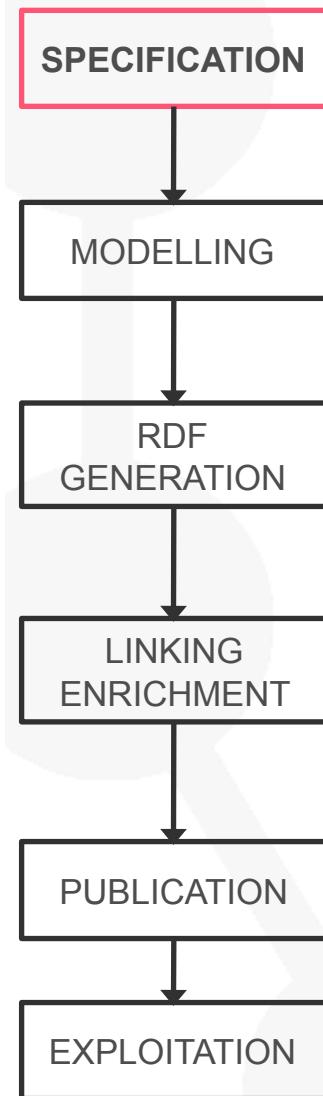
A (Library) Linked Data Life-cycle



- Produce user interfaces that integrate your data (and other sources)
- Provide innovative services on top of the data
- Integrate with existing services (e.g. patron services), etc.



A collection of current limitations



- MANY source formats, encodings and schemas:

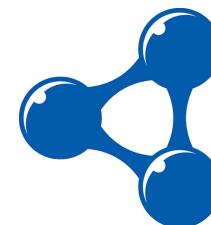


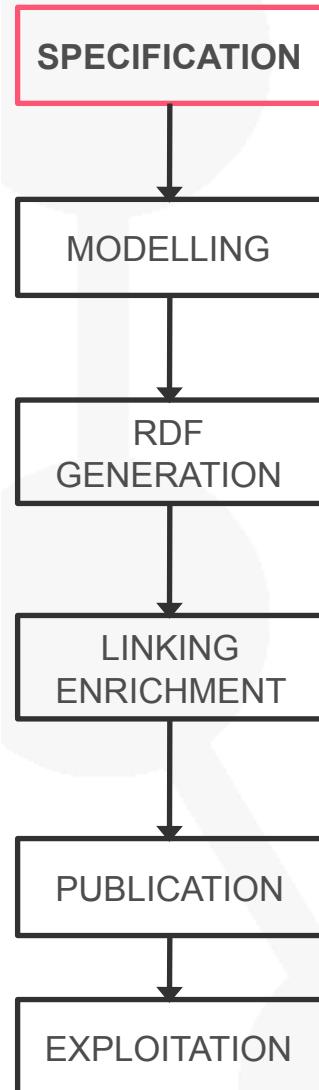
PICA+

MAB

CSV

<XML>

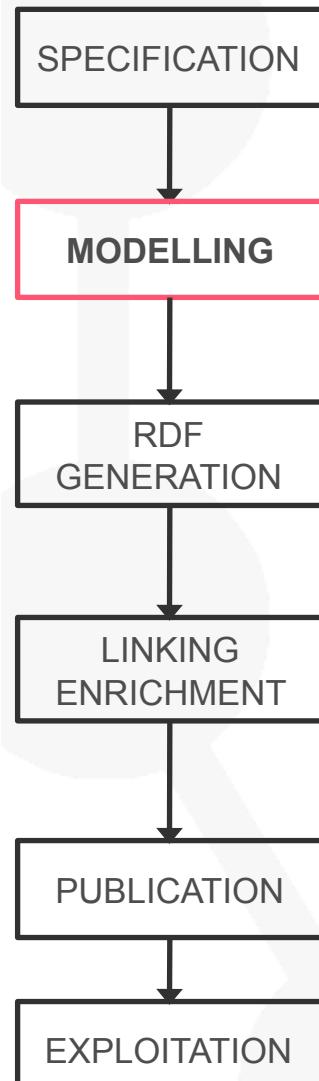




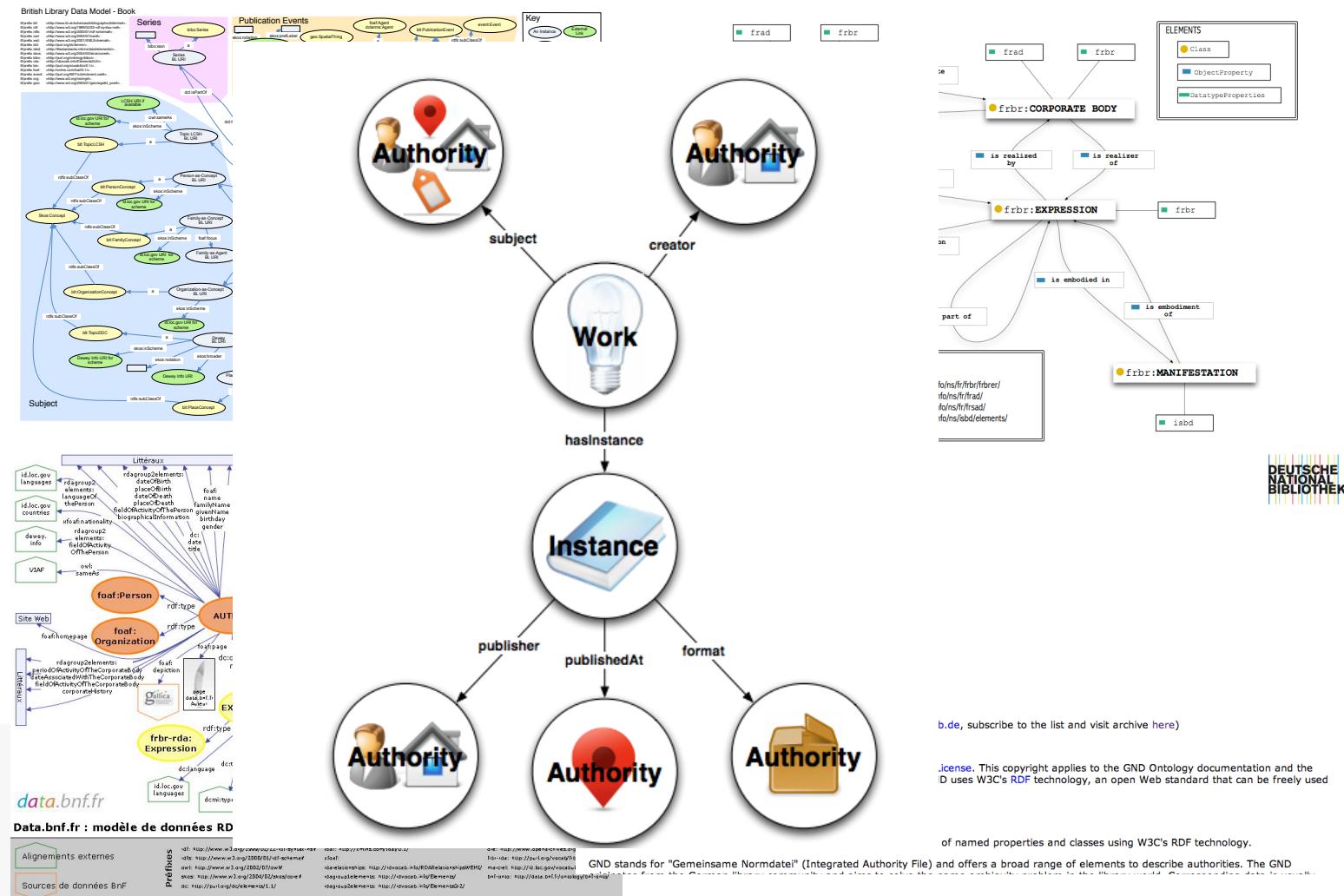
LIMITATIONS:

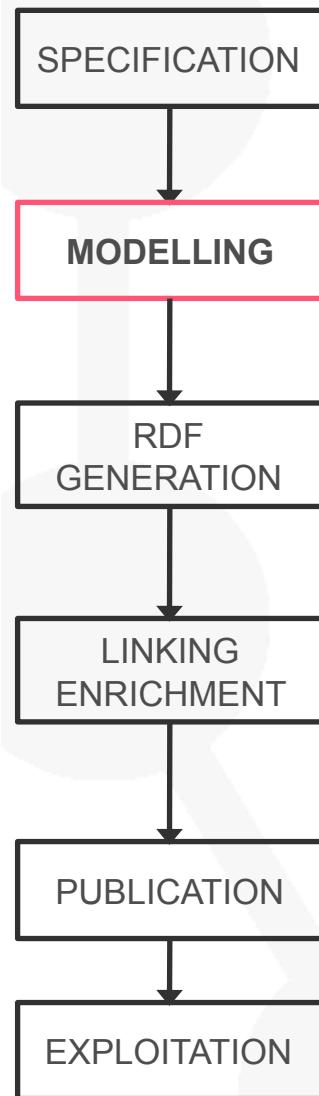
1. Lack of principled *methods, techniques and tools* to deal with **heterogeneous source formats, schemas and encodings**
😊 metastream, metamorph APIs from culturegraph
2. Need for **analysis of the semantics of metadata schemas** and the **variation of their usage** accross libraries:
😊 **MARC 21: "Marc21 as Data: A start".**
Karen Coyle's code4lib's article*

* <http://journal.code4lib.org/articles/5468>



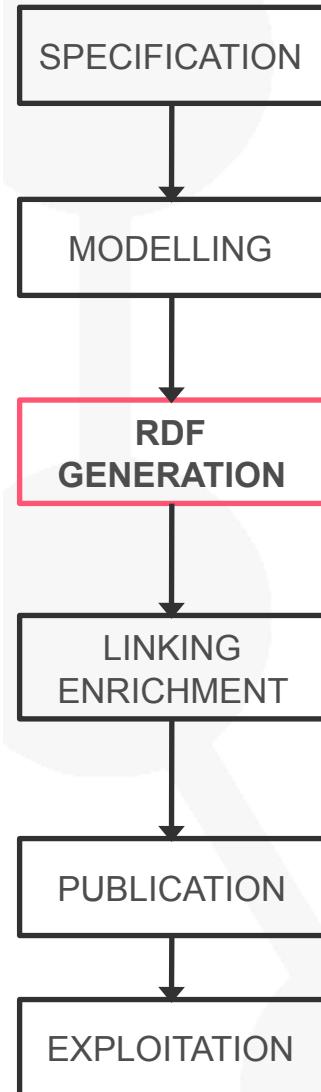
- Many different models and approaches



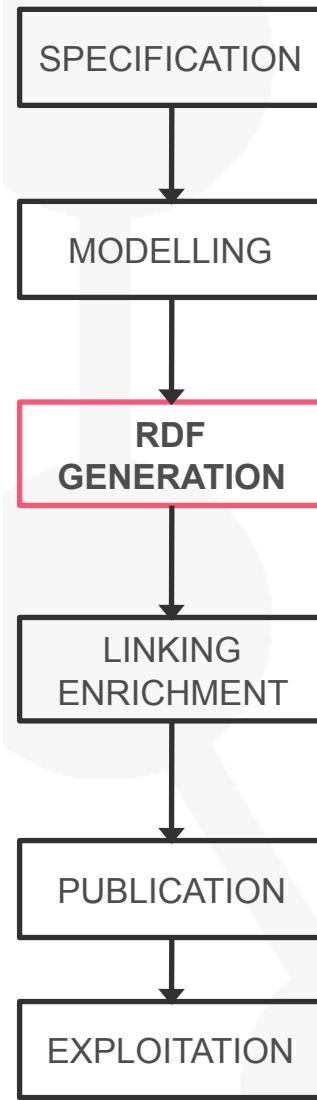


LIMITATIONS:

1. Difficulties in **using vocabularies that adapt to**
 - Past and current cataloguing practices
 - Past and current library formats and schemas
2. **Mapping and managing vocabularies**
 1. Lack of mapping across vocabularies
 2. High manual effort and costly process
 - ☺ GND ontology,
Dunsire et al "Linked Data vocabulary management"*
3. **Lack of multilingual vocabulary elements description**
 - ☺ IFLA Namespace Multilingual efforts



- Several tools, APIs and systems
- Almost each new project follows its own approach → it is still a costly process
- Participation of library experts is crucial (those who know the formats, e.g. MARC 21)
- There is not a generic mapping from source metadata schemas (due to the differences in use across libraries, different cataloguing practices)



LIMITATIONS:

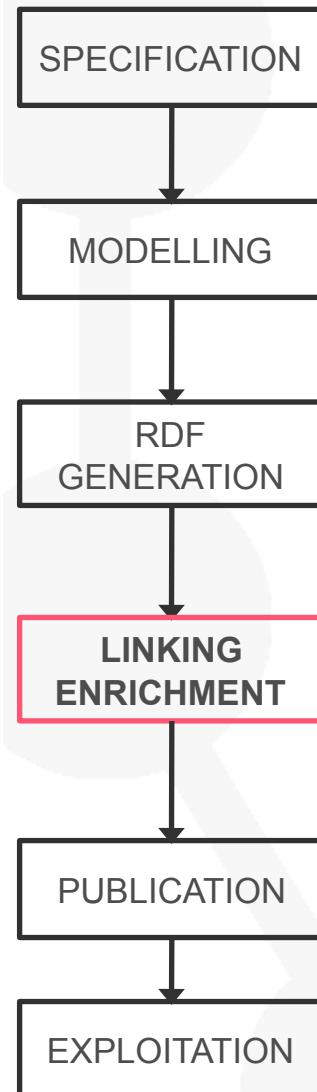
1. Lack of tools and services that are **easy to use by non-technical users** (experts in library formats)

2. Lack of **integrating tools and services** that provide a **full view** of the mapping and RDF generation process
 - **Analytics of the source data** (e.g. how often is a specific MARC subfield used, how many records will an specific transformation rule affect..)
 - **Dashboard-like** generation that helps to understand how the RDF data has been produced and allows to refine the mappings, find errors, etc.



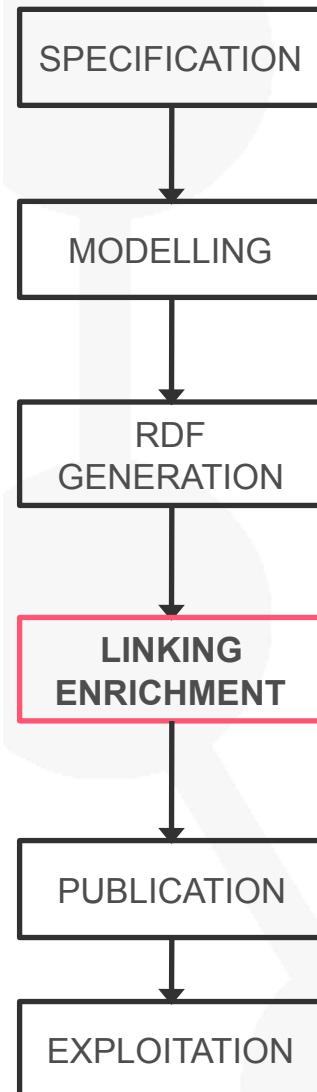
<http://marimba4lib.com>

Linking and enrichment



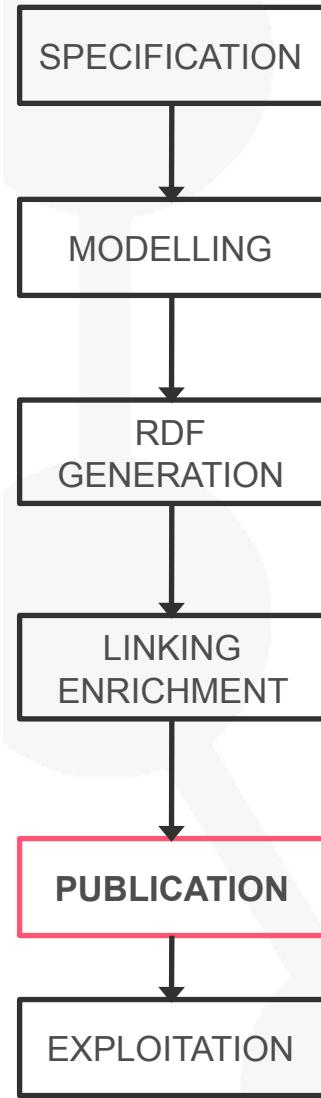
- Already a **high number of links at the authority level** (mostly Persons and Corporate Bodies)
 - VIAF
 - Culturegraph Authorities
 - DNB, BNF, BNE, BL
- Very positive efforts:
 - National library cataloguers adding links during the cataloguing process (e.g. DNB, BNE)
 - Cross-library collaboration for linking (British Library and DNB*)

* <http://www.culturegraph.org/Subsites/culturegraph/DE/Home/news4.html>



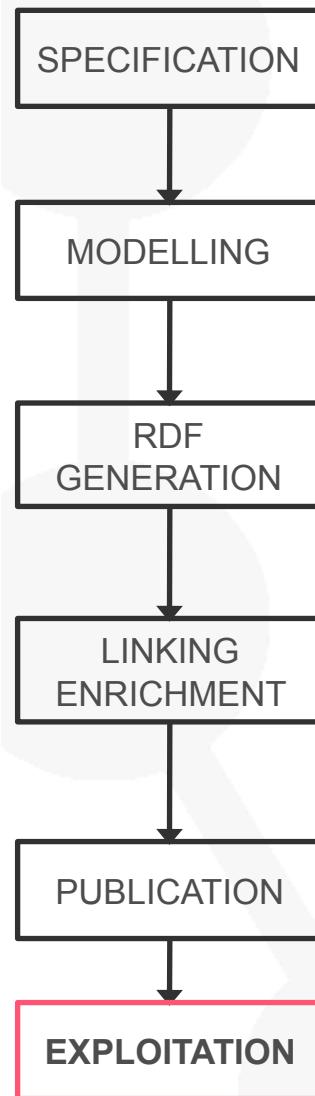
LIMITATIONS:

1. Very **limited linking at the bibliographic level**
2. Lack of **cross-lingual mechanisms** to link resources accross libraries
3. Lack of the **a solid infrastructure to enable links sharing and exchange** accross libraries
4. No **semantic enrichment of the content** (textual, sound and visual) and **linkage** of this content **to the URIs** that represent the real-world entities



LIMITATIONS:

1. No extensive use of mechanisms to indicate provenance, license, last-update, etc. in a per-resource basis
2. Low usage of mechanisms to enable efficient discovery and usage like **void descriptions**
3. Need for scalable and generic infrastructure to facilitate consumption of linked data:
 1. Most of the APIs are not extensively documented
 2. Not every dataset provides search over the data
 3. W3C Linked Data platform WG proposition is promising (REST access to resources and containers: paging, etc.)



LIMITATIONS:

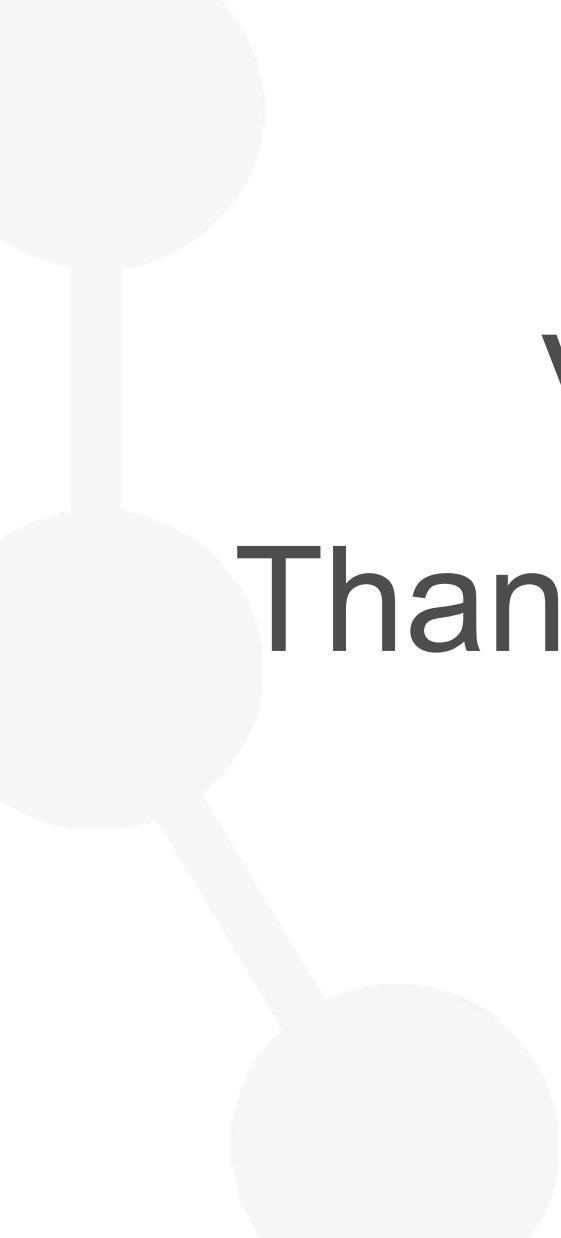
1. Lack of integration of library linked data in
 - Library curation and cataloguing workflows
 - Existing library systems (e.g. digital library systems, patron services)
2. Need of end-user interfaces providing enriched information spaces
 - that integrate several LD sources,
 - allowing for serendipitous discovery and enhanced navigation



CultureSampo,
<http://uilld2013.linkeddata.es> Call for participation!

Conclusions

- There still exist **some barriers**:
 - **Organizational**: Infrastructure costs, integration of linked data into library processes, cross-library collaboration
 - **Language**: Multilingual services, cross-lingual linking and data integration
 - **Data formats and modalities**: Different digital objects and representations are rarely interlinked, cope with heterogeneity of formats and vocabularies
- **But we are on the right path!**



Vielen Dank!

Thank you very much!

Questions?