

MAPPING

marimba-datamodel

marimba-rml

marimba-mapping

001 → bima0000007176

005 → 20150511

008 → 910805s1605 sp spa

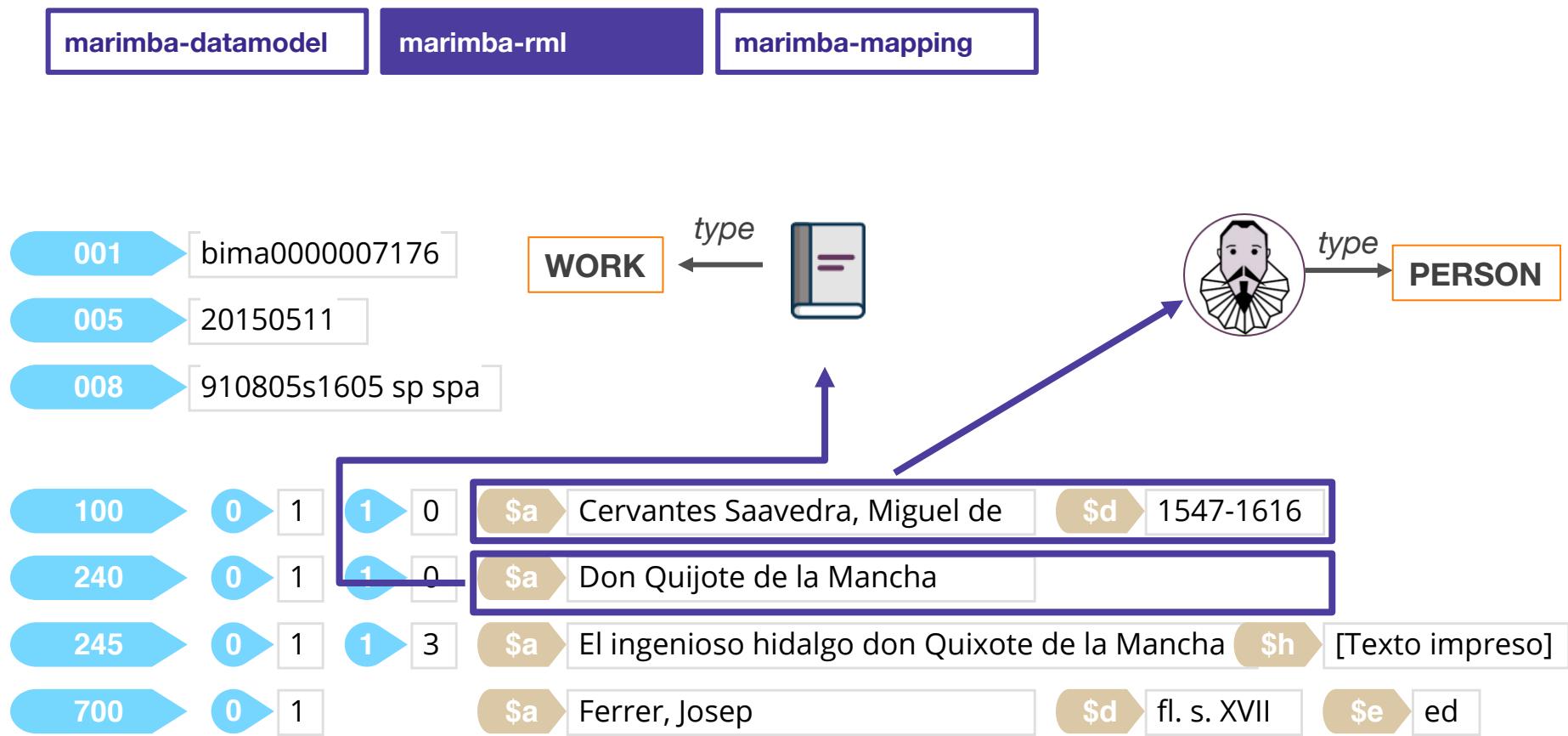
100 → 0 1 1 0 \$a Cervantes Saavedra, Miguel de \$d 1547-1616

240 → 0 1 1 0 \$a Don Quijote de la Mancha

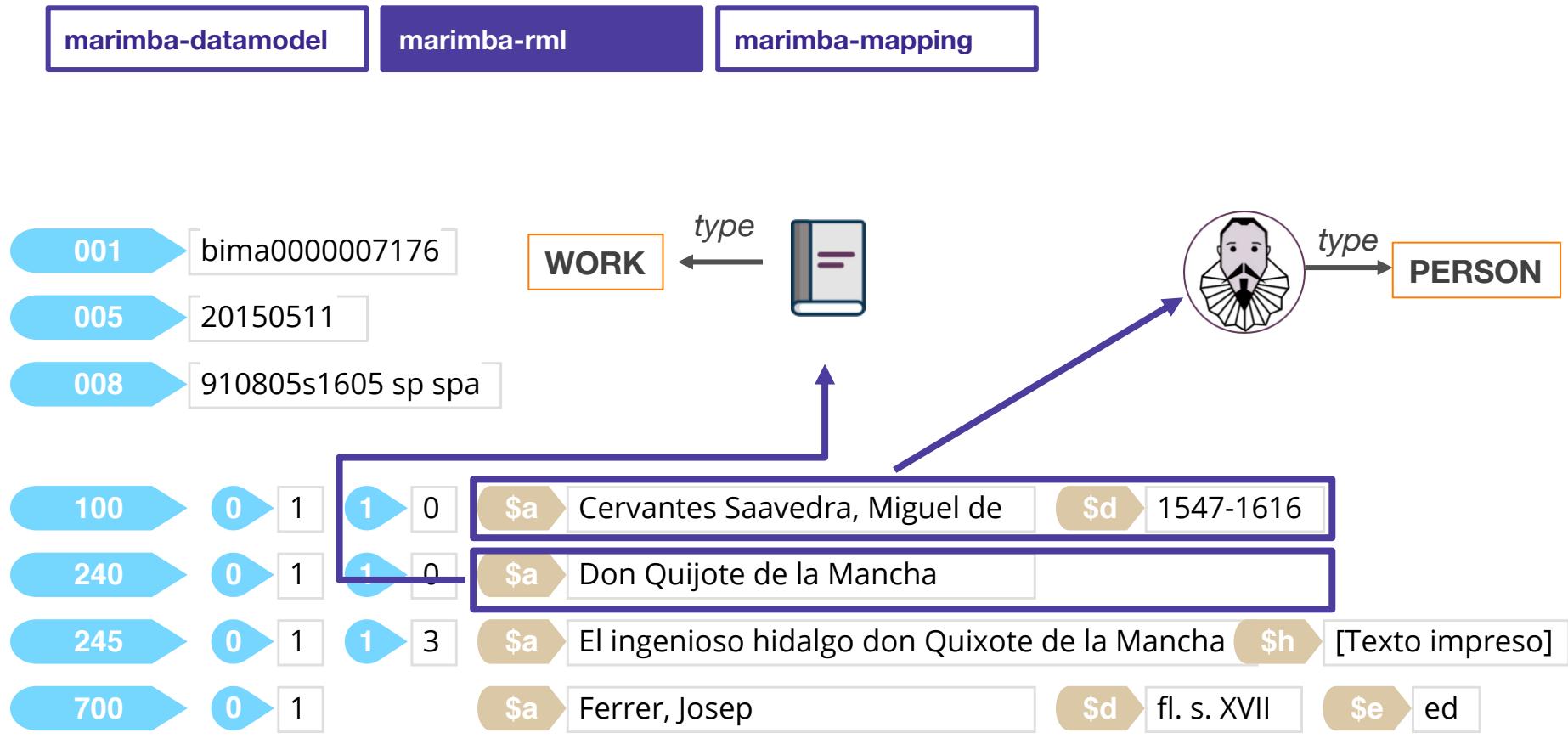
245 → 0 1 1 3 \$a El ingenioso hidalgo don Quixote de la Mancha \$h [Texto impreso]

700 → 0 1 \$a Ferrer, Josep \$d fl. s. XVII \$e ed .

MAPPING



MAPPING

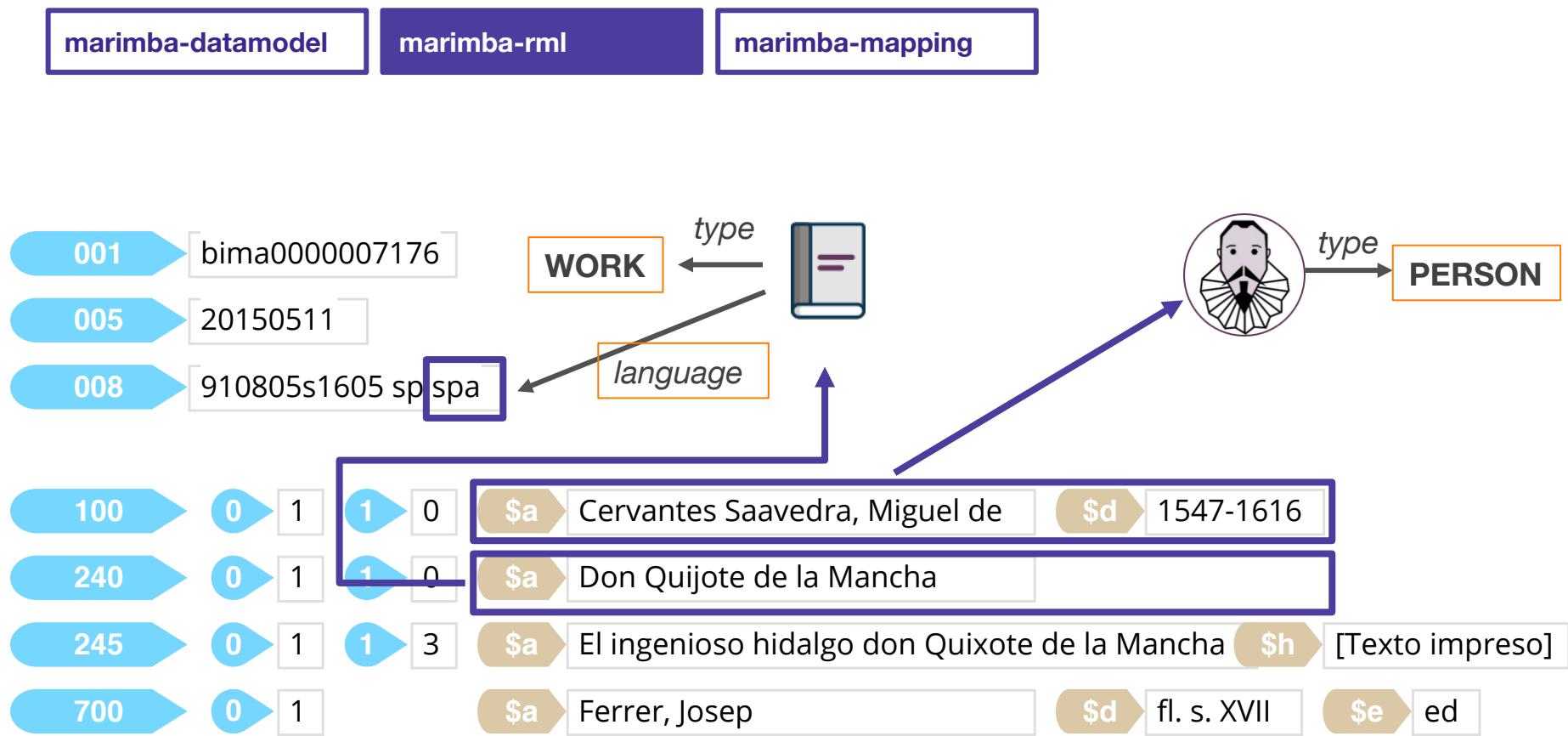


First modification of R2RML (*R2RML views*)

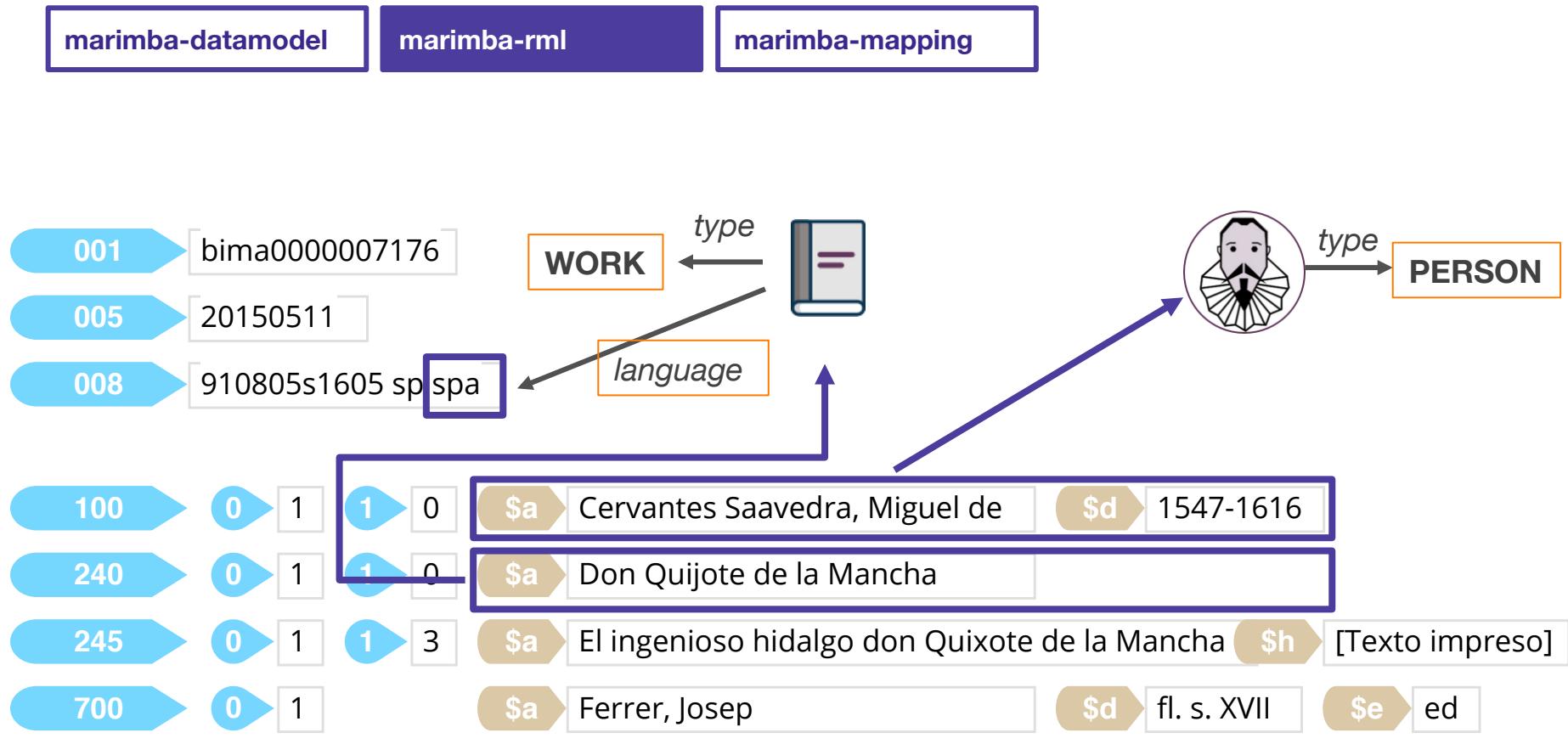
rr:sqlVersion → marimba-sql IRI

marimba-sql queries for processing data in nested relations.

MAPPING



MAPPING

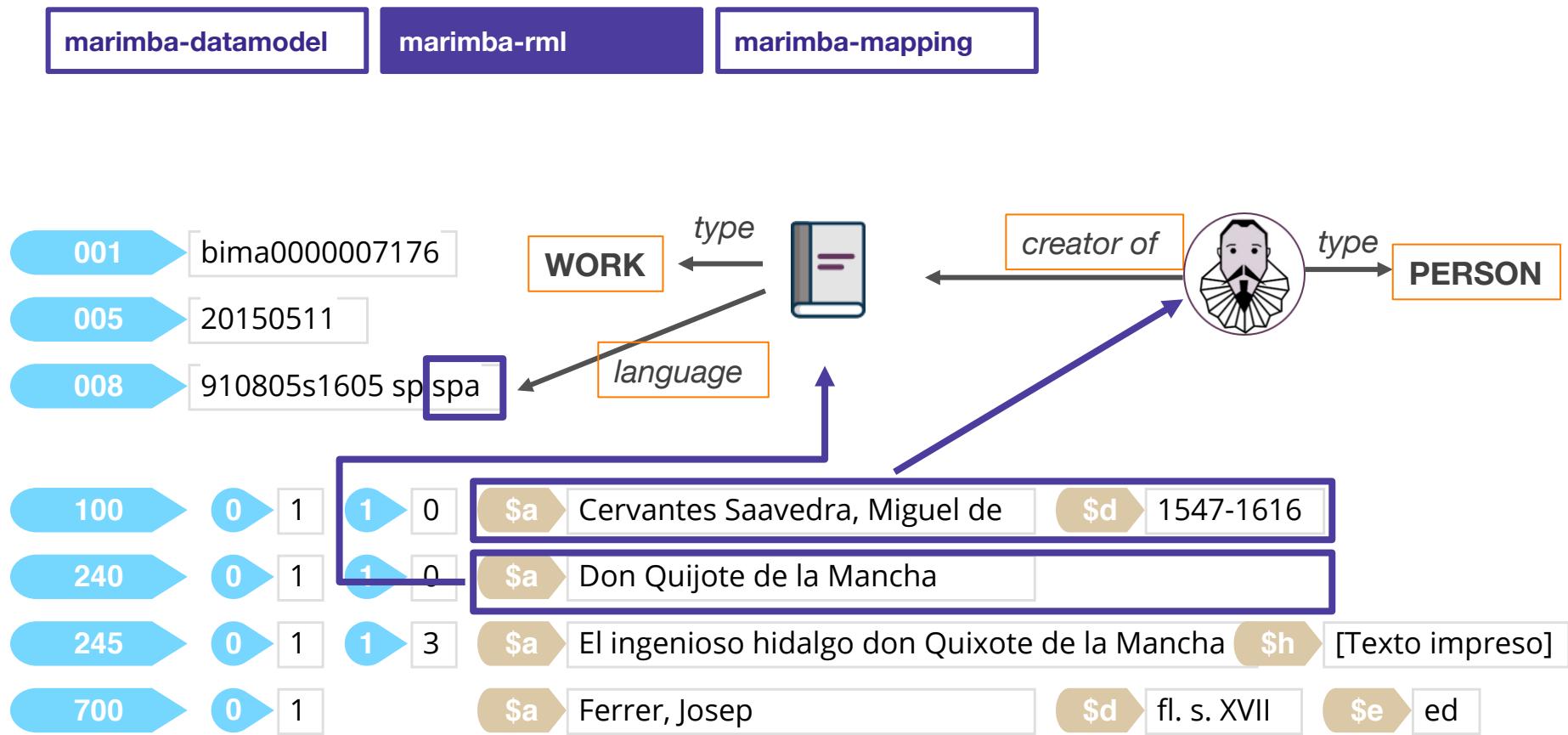


Second modification of R2RML (*rr:template and rr:column*)

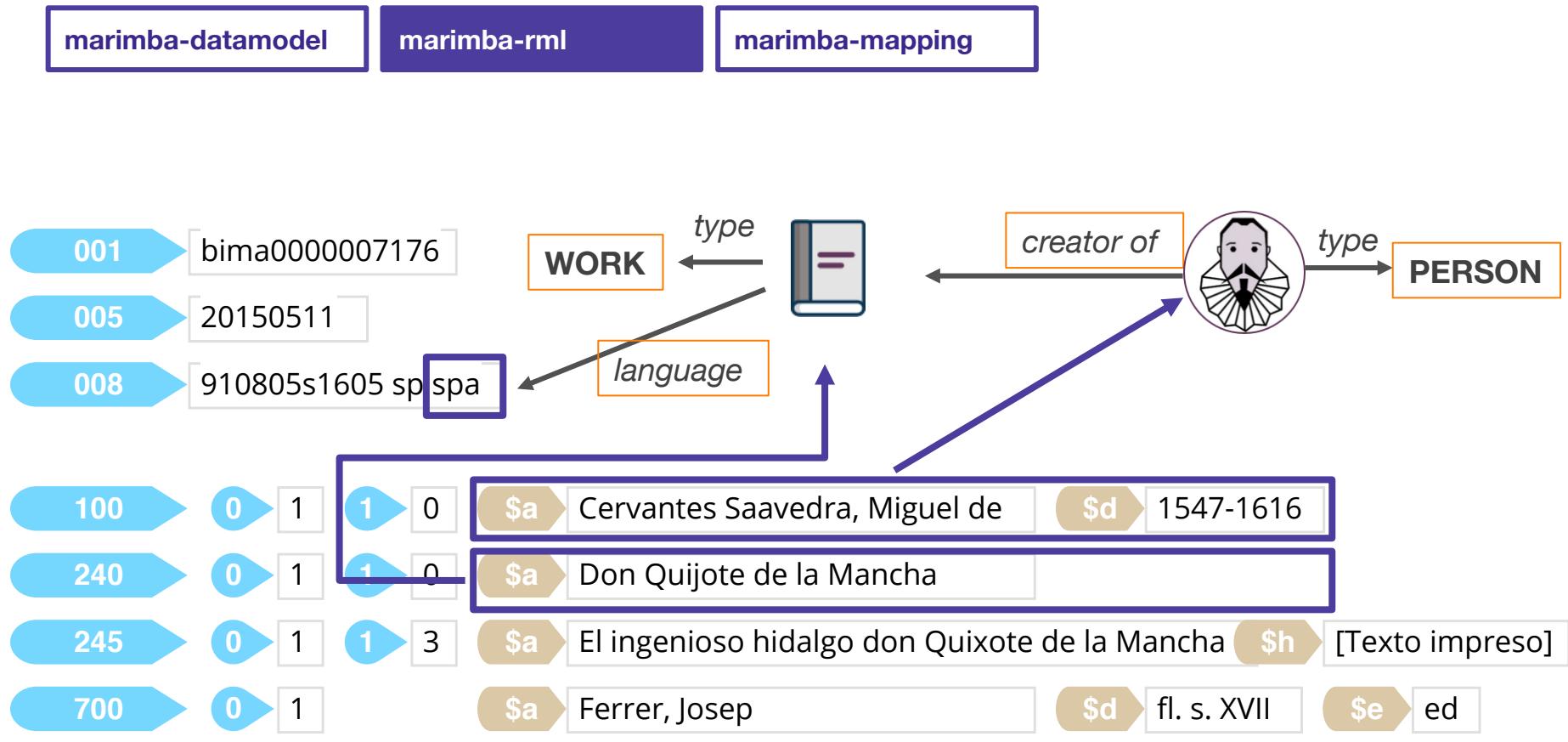
R2RML column → marimba-sql query

Generate **values** from attributes in nested relations.

MAPPING



MAPPING



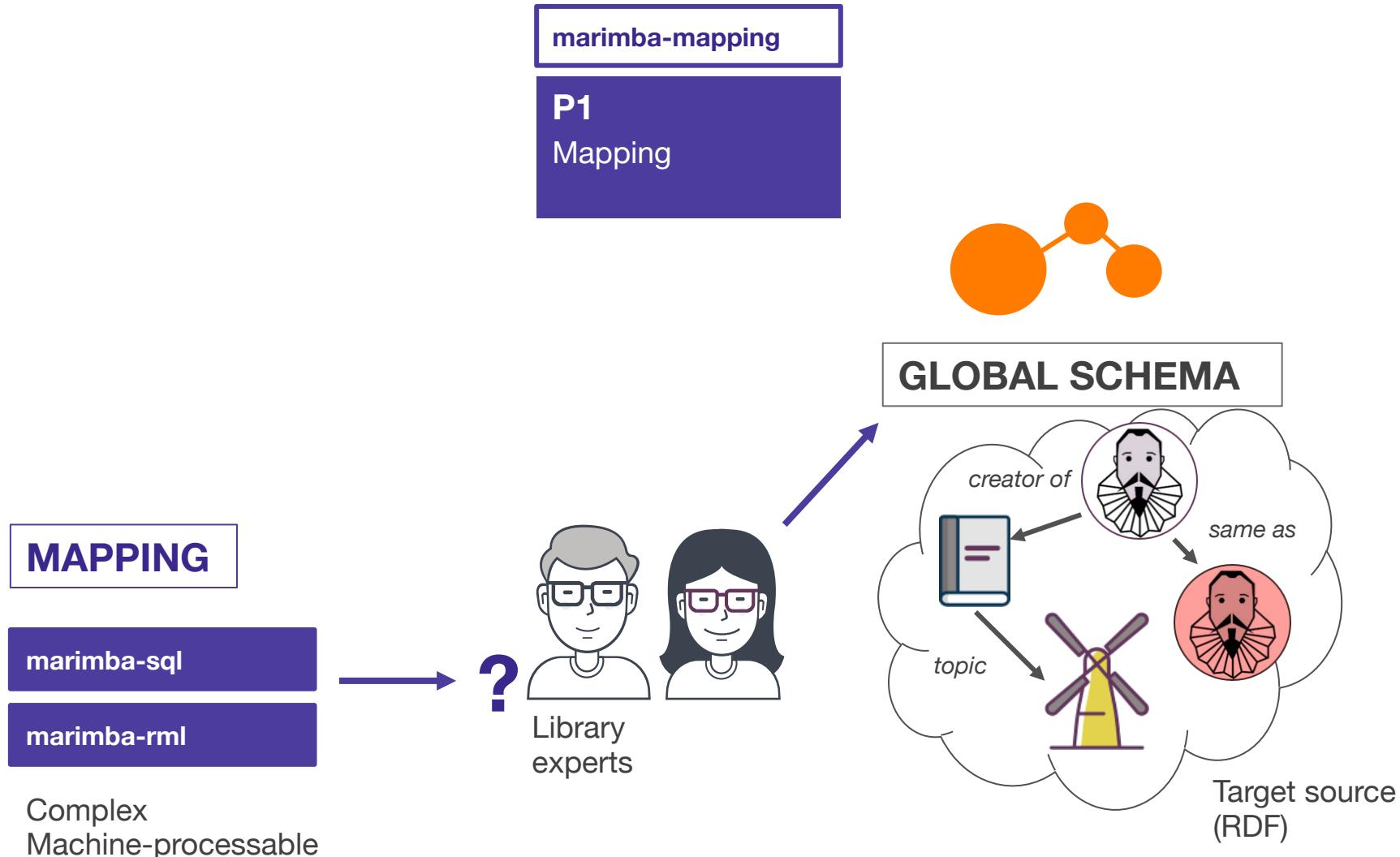
Third modification of R2RML (*rr:refObjectMap*)

R2RML column → marimba-sql query

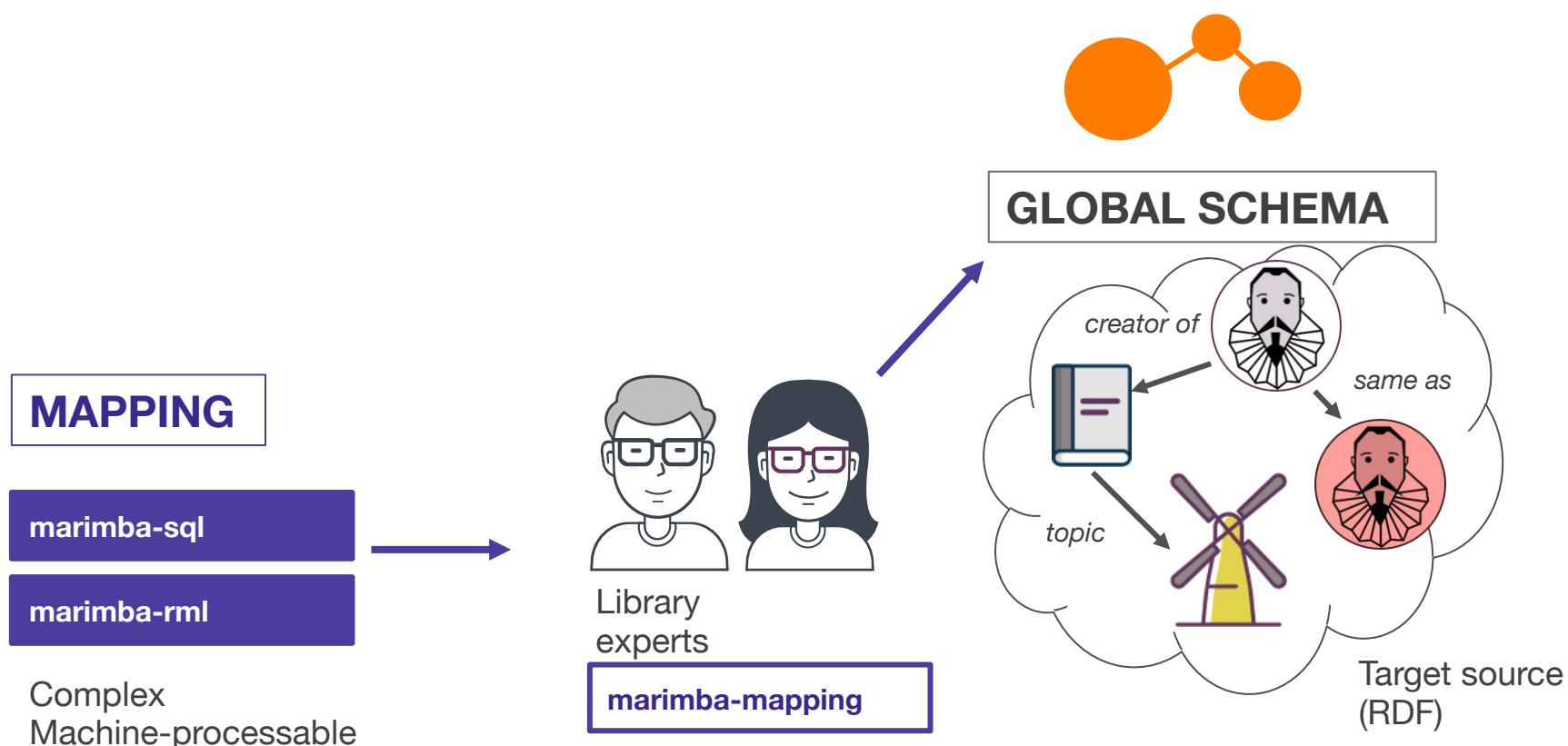
Generate ***joins*** across nested relations



Defined operational semantics
Three core modifications



MAPPING



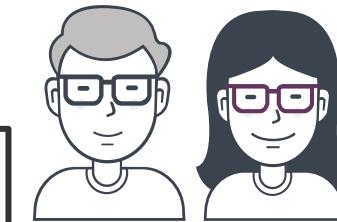
marimba-datamodel

marimba-rml

marimba-mapping

Generation of mapping templates

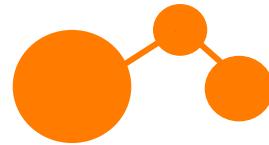
- Support the **mapping and ontology development process by library experts**
- **Full coverage** of data sources (e.g., at BNE more than 4,000 patterns)
- Library experts write simple mappings that are translated into **marimba-rml**



MARC 21 metadata	Record count	Class IRI
100adt	1,222,400	
100ad	999,789	
100adtl	567,534	
100ae	1,658	
100ac	20,768	

SCHEMA	STATISTICS	40
--------	------------	----

OUTLINE



MARC 21 metadata	Record count	Class IRI
100adt	1,222,400	
100ad	999,789	
100adtl	567,534	
100ae	1,658	
100ac	20,768	

P1
Mapping

BNE ontology

marimba-topicmodel

marimba-modelling

P2
Ontology
development
process

Empirical studies

datos.bne.es

P3
Library
applications

Library
data sources

Ontology-based
Library data (RDF)

Library
online applications

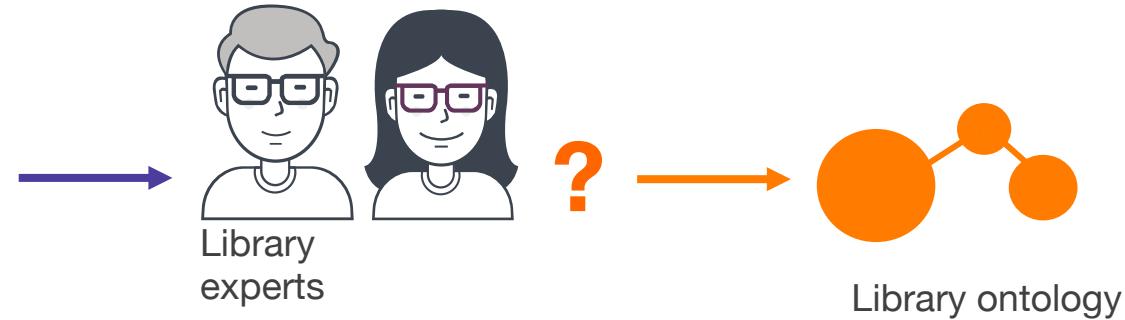


ONTOLOGY DEVELOPMENT



MARC 21 metadata	Record count	Class IRI
100adt	1.222.400	
100ad	999.789	
100adtl	567.534	
100ae	1.658	
100ac	20.768	

Mapping templates

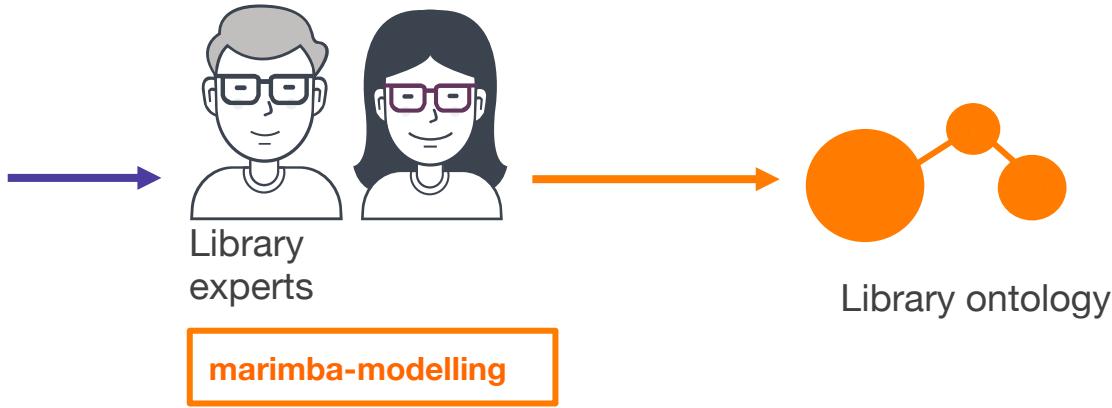


ONTOLOGY DEVELOPMENT



MARC 21 metadata	Record count	Class IRI
100adt	1.222.400	
100ad	999.789	
100adtl	567.534	
100ae	1.658	
100ac	20.768	

Mapping templates



ONTOLOGY DEVELOPMENT

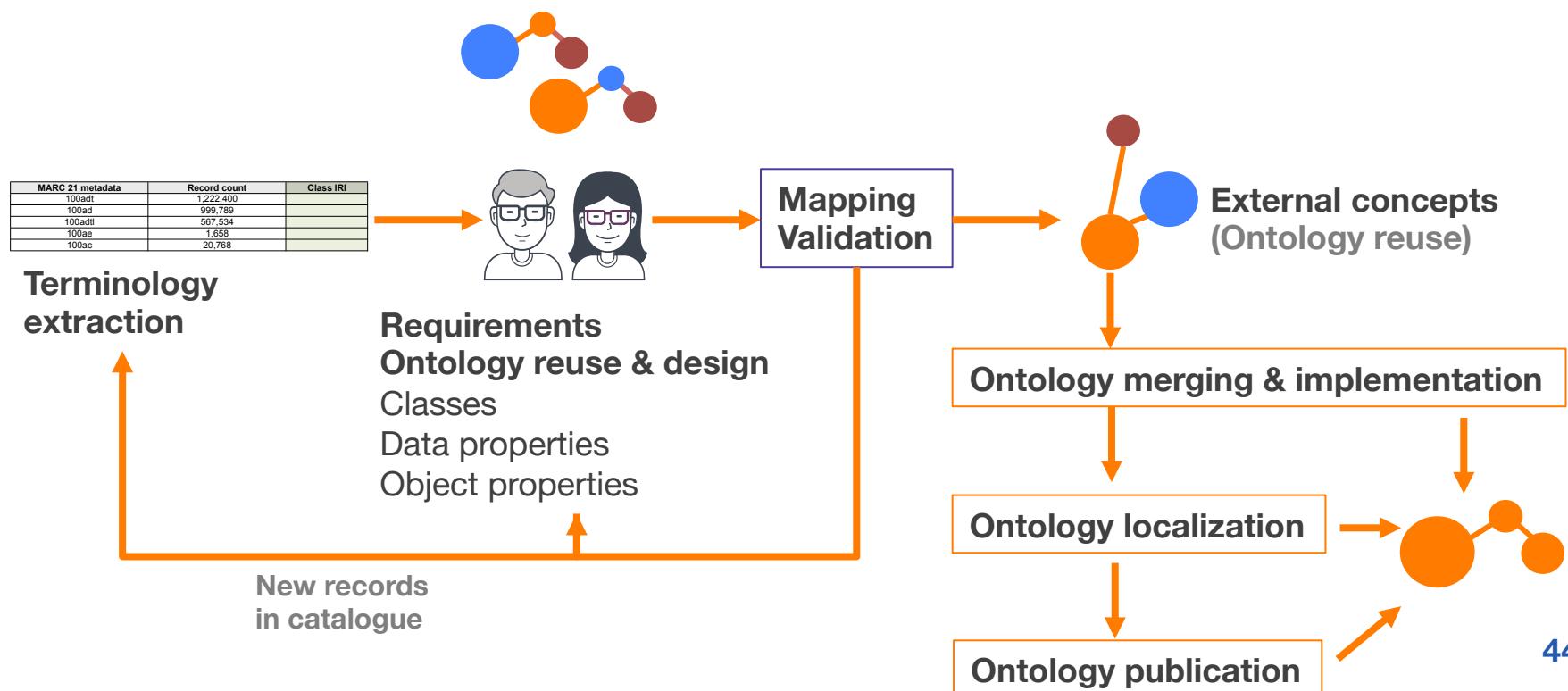


marimba-modelling

marimba-topicmodel

BNE ontology

- **Life-cycle model** based on **NeOn** methodology (Suarez-Figueroa et al. 2015).
- **Terminology extraction and ontology design activities** using **mapping templates**.
- **Ontology publication activity** added to NeOn.
- Two phases: (1) **Ontology reuse**, and (2) **ontology merging**



ONTOLOGY DEVELOPMENT



marimba-modelling

marimba-topicmodel

BNE ontology

ONTOLOGY DESIGN USING MAPPING TEMPLATES

MARC 21 metadata	Class IRI
100adt	frbr:C1001
100ad	frbr:C1005
100adtl	frbr:C1002
100ae	frbr:C1005
100ac	frbr:C1005

Classification template

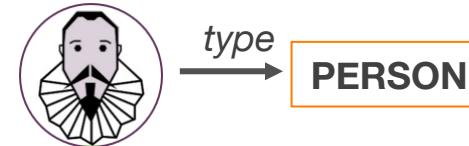


CLASSES



WORK

← type



type →

PERSON

ONTOLOGY DEVELOPMENT



marimba-modelling

marimba-topicmodel

BNE ontology

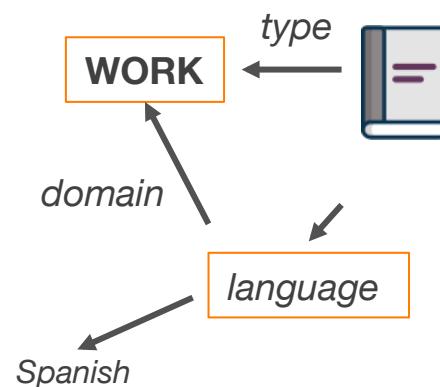
ONTOLOGY DESIGN USING MAPPING TEMPLATES

MARC 21 metadata	Datatype property IRI	Domain IRI
245n	isbd:P3033	frbr:C1003
321a	isbd:P3032	frbr:C1003
110a	frbr:P6001	frbr:C1006
400a	frbr:P5012	frbr:C1005

Annotation template



DATATYPE PROPERTIES
RDFS DOMAINS



ONTOLOGY DEVELOPMENT



marimba-modelling

marimba-topicmodel

BNE ontology

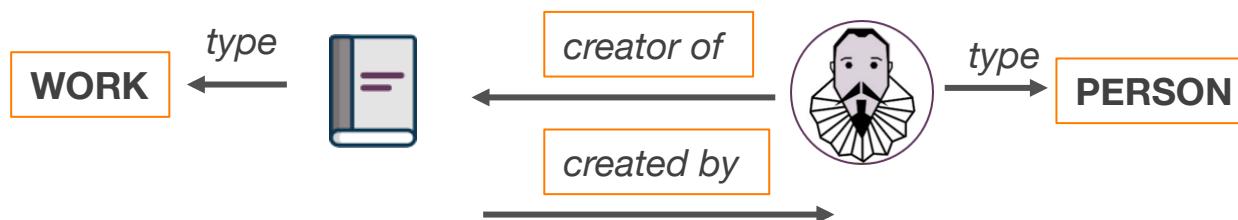
ONTOLOGY DESIGN USING MAPPING TEMPLATES

MARC 21 metadata	Object property IRI	Inverse property IRI	Domain IRI
t	frbr:P5001	frbr:OP1001	frbr:C1005
l	frbr:P1002	frbr:OP2002	frbr:C1001
n	frbr:OP1003	frbr:OP1004	frbr:C1001

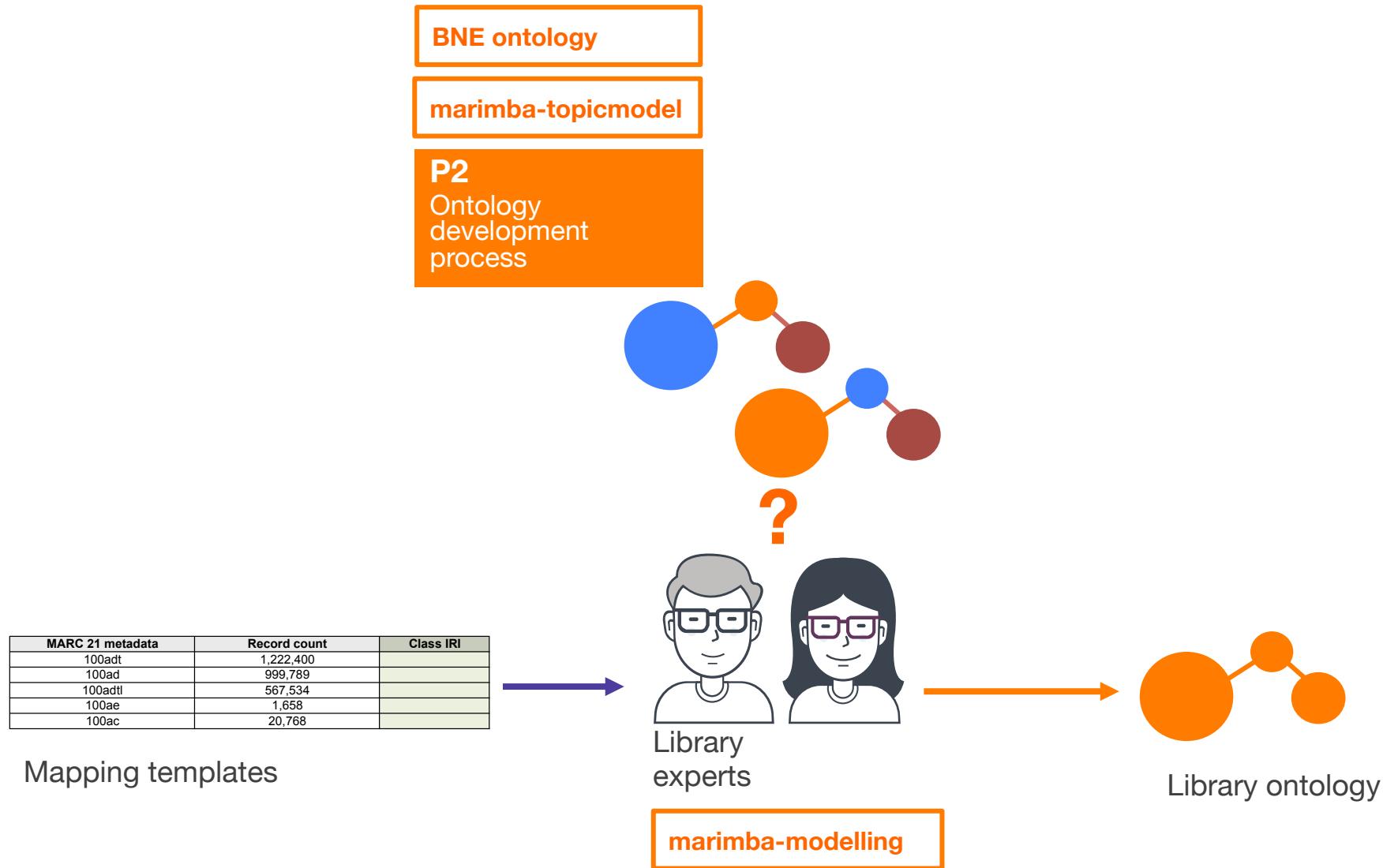
Relation extraction template



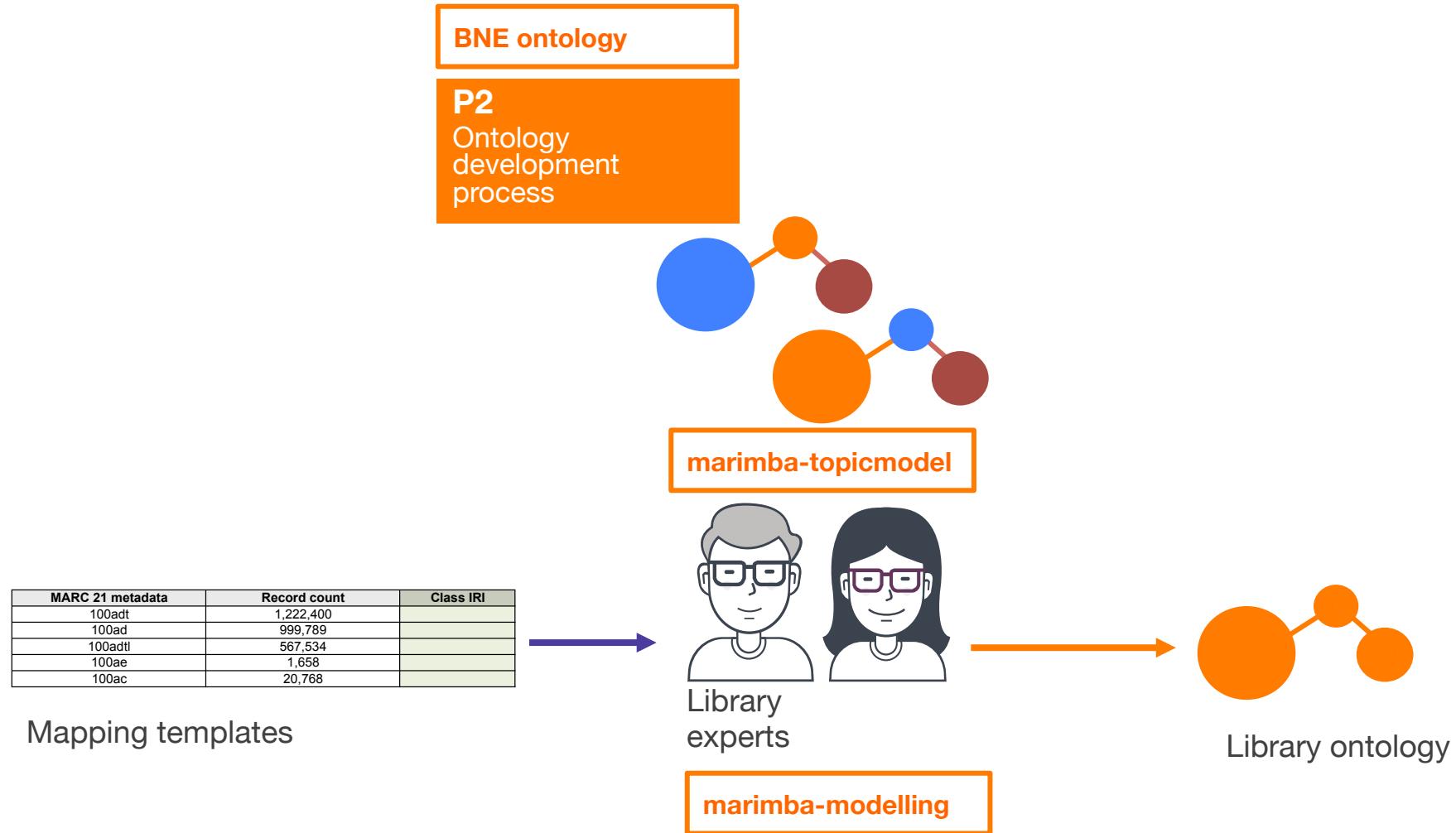
OBJECT PROPERTIES
RDFS DOMAINS
INVERSE PROPERTIES



ONTOLOGY DEVELOPMENT



ONTOLOGY DEVELOPMENT



marimba-topicmodel

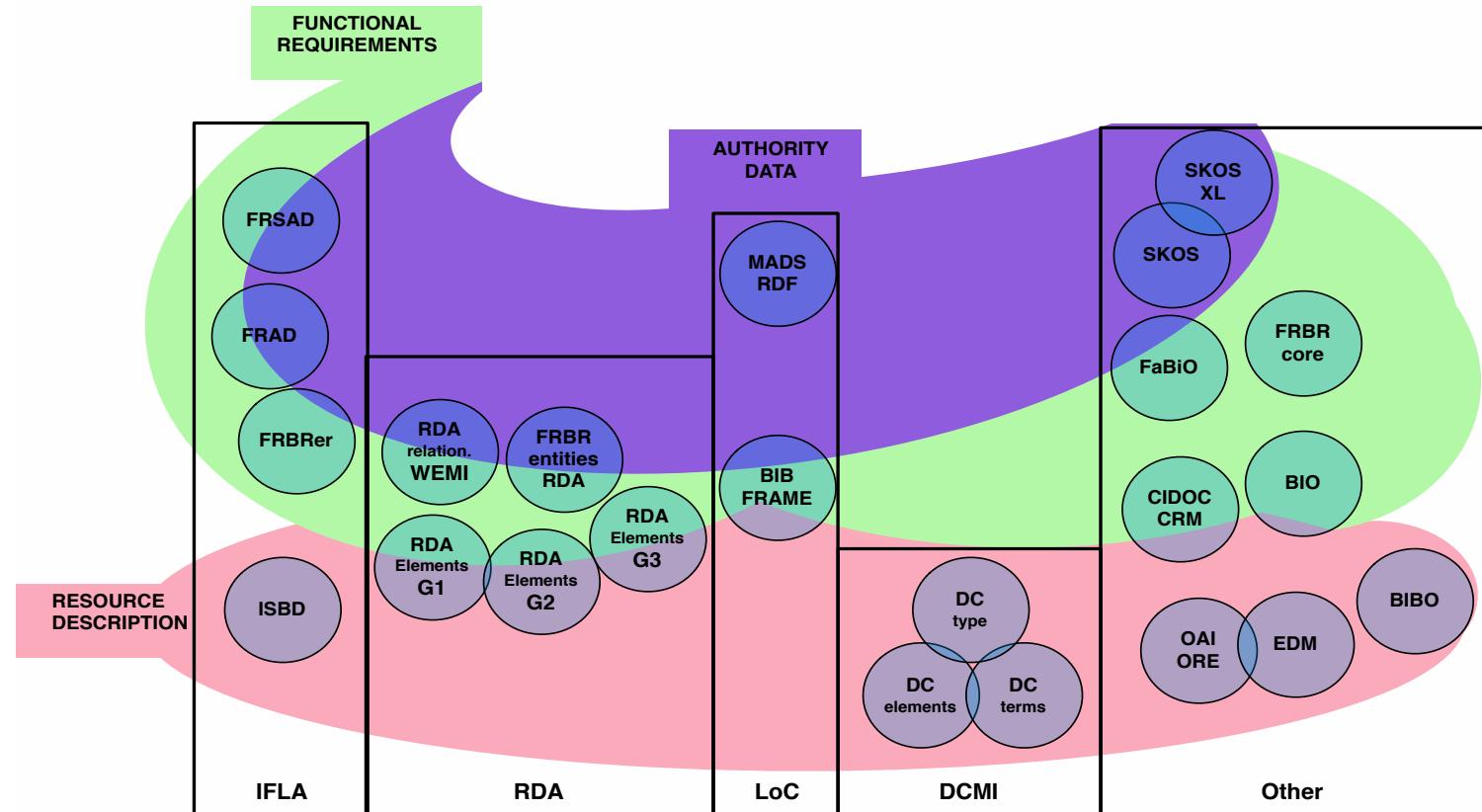
Evaluation

Application

MOTIVATION: MANUAL CLASSIFICATION OF LIBRARY ONTOLOGIES

- Using **ontology repositories** (LOV and metadataregistry.org)
- Very **costly** and **not scalable** (e.g., new ontologies?)

Can we do this analysis automatically?

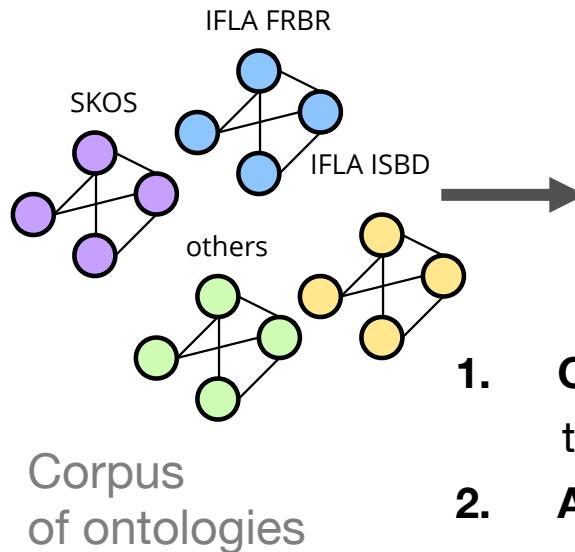


marimba-topicmodel

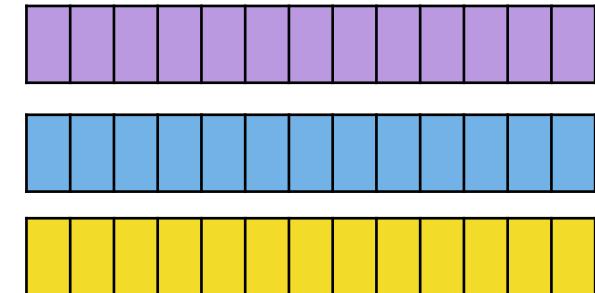
Evaluation

Application

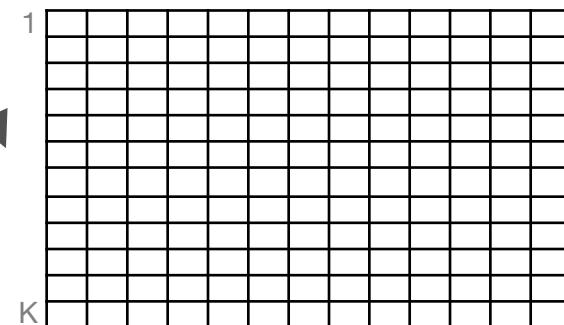
MARIMBA-TOPICMODEL: OVERVIEW



1. **Ontology documents** from textual descriptions
2. **Annotate with external word senses**
3. Train a **probabilistic topic model**



Ontologies as probability distributions over **K** topics



K topics as probability distributions over **S** word senses

Topic-based ontology similarity measures:

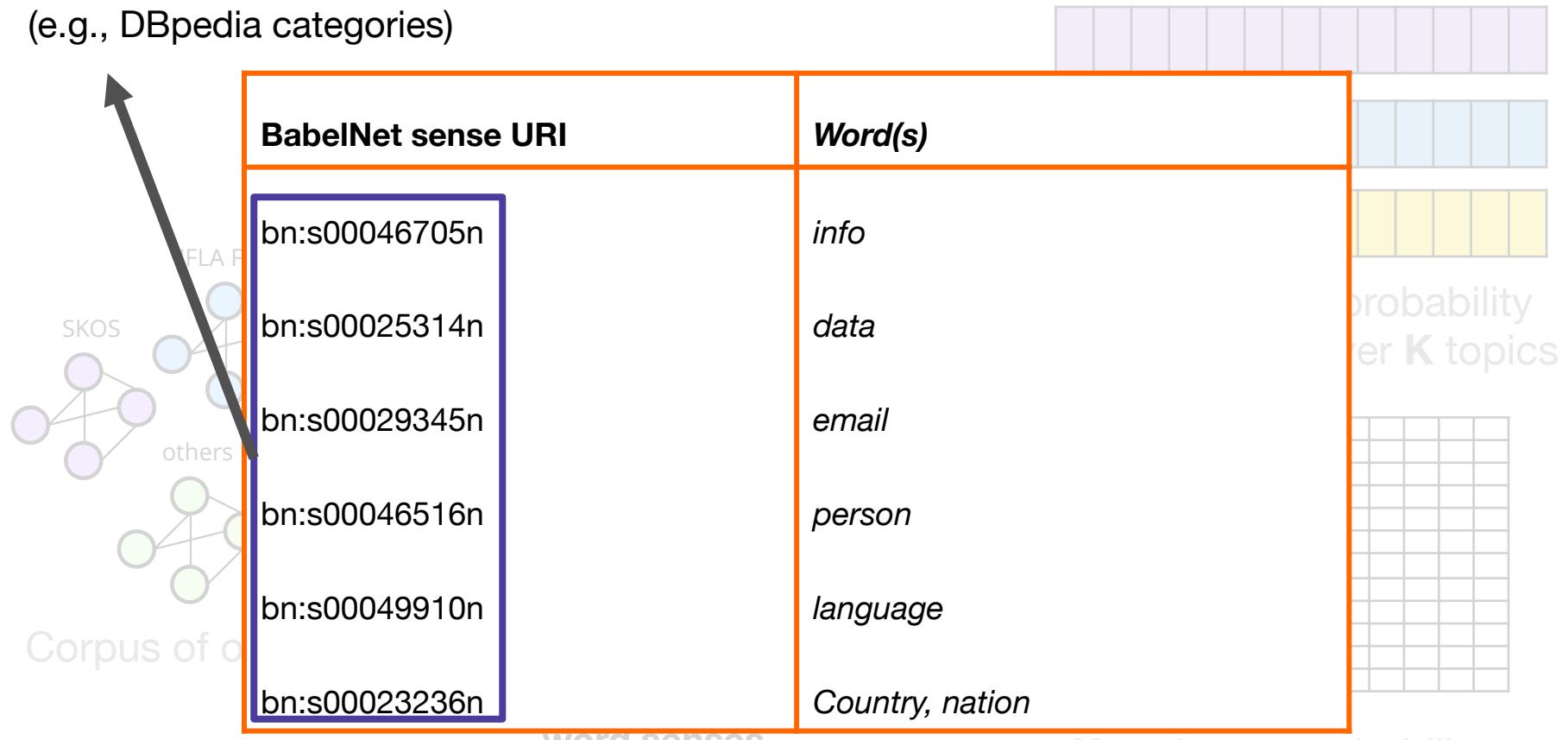
JSD: Jensen-Shannon Divergence

JSM: Jensen-Shannon Metric

marimba-topicmodel

Evaluation**Application**

LOD resources
(e.g., DBpedia categories)



3. Train a probabilistic topic model

TOPIC EXAMPLE: PERSONAL INFORMATION

HIGH PROBABILITY: FOAF, BIBO, FABIO

marimba-topicmodel

Evaluation

Application

EXPERIMENT 1: TOPIC COHERENCE

1. External senses improve the coherence of topics?
2. Short-text oriented topic models produce more coherent topics than classical topic models?

Measure: Mimmo et al. [2011]

METHODS

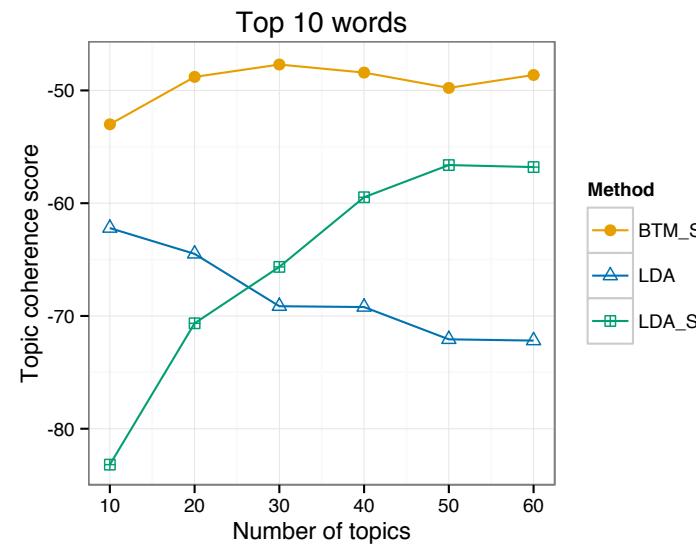
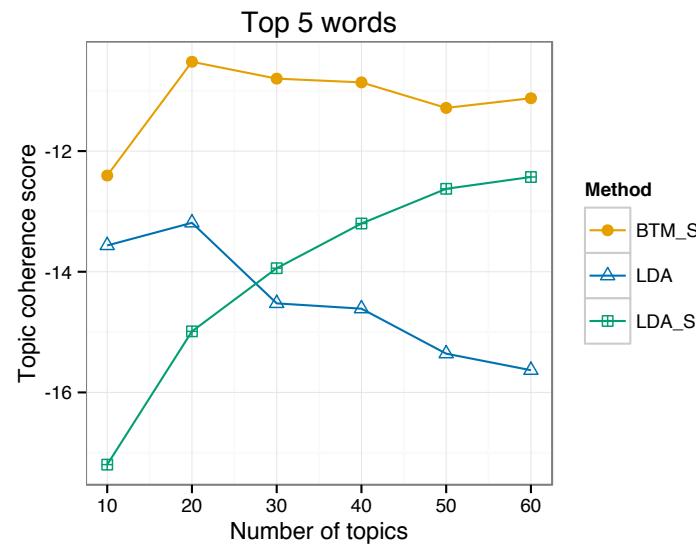
Classical LDA (Latent Dirichlet Allocation)

LDA_S: LDA with sense-annotated documents

BTM_S (marimba-topicmodel): BTM with sense-annotated documents

EXPERIMENT 1: TOPIC COHERENCE

1. External senses improve the **coherence** of topics?
2. Short-text oriented topic models produce **more coherent** topics than classical topic models?



1. Annotation with **senses increase topic coherence** ($\text{BTM}_S > \text{LDA}_S > \text{LDA}$)
2. **Short-text topic models increase coherence**
(BTM_S consistently outperforms LDA and LDA_S).

marimba-topicmodel

Evaluation

Application

EXPERIMENT 2: ONTOLOGY CLUSTERING EXPERIMENT

Does **marimba-topic model** produce **more precise clustering** than classical search methods?

- **How to measure?**
 - **H-score:** IntraCluster distance < Intercluster distance
- **Methods with different distance measures and metrics:**
 - **2 TF-IDF:** Baseline method (Cosine and Euclidean distance)
 - **2 Marimba-topicmodel** (JSD and JSM)
- **Gold-standards** (annotated by humans)
 1. **LODTHEMES:** LOD cloud themes
 2. **LOVTAGS:** Topic tags of LOV repository

marimba-topicmodel

Evaluation

Application

EXPERIMENT 2: ONTOLOGY CLUSTERING EXPERIMENTAL RESULTS

	tf-idf cosine	marimba-cosine	marimba-jsd
LOVTAGS	0.924	0.861	0.859
LODTHEMES	0.842	0.722	0.683

	tf-idf eucli	marimba-jsm
LOVTAGS	0.808	0.998
LODTHEMES	0.723	0.741

* **Lower** values mean
better precision
 (low intracluster,
 high intercluster)

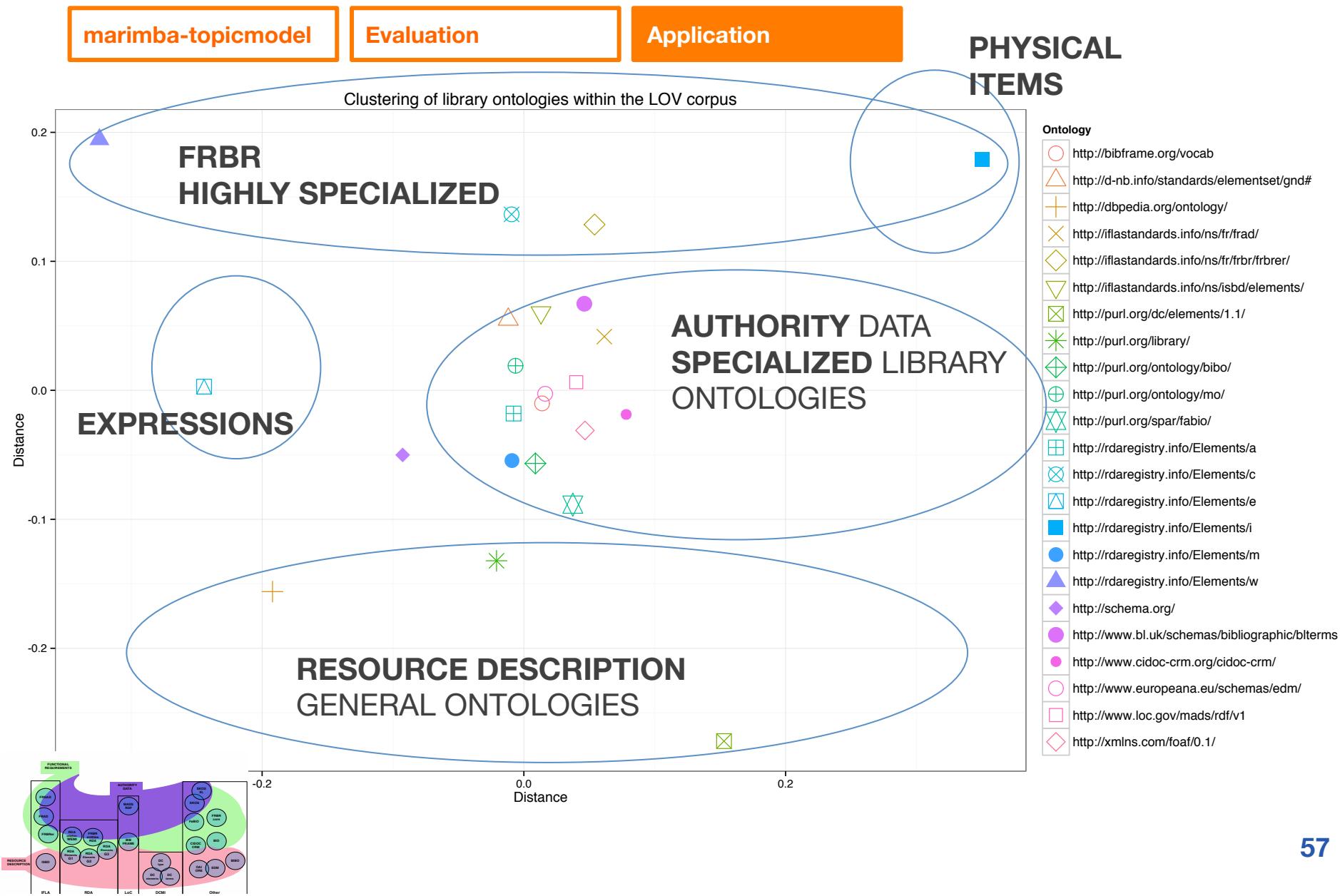
Does **marimba-topic model** produce **more precise clustering** than classical search methods

H4

Topic-based ontology similarity measures perform better than existing methods used in ontology search for **clustering widely used LOD ontologies** (LOV repository and LOD cloud).

But, what about **library ontologies** and **more qualitative results?**

ONTOLOGY DEVELOPMENT



ONTOLOGY DEVELOPMENT



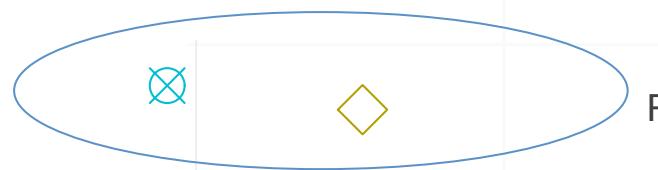
marimba-topicmodel

Evaluation

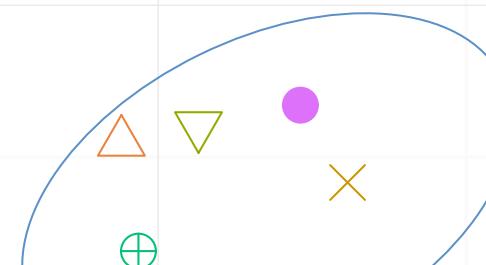
Application

Ontology

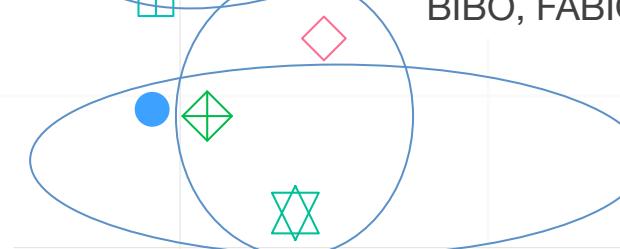
- (○) http://bibframe.org/vocab
- (△) http://d-nb.info/standards/elementset/gnd#
- (+) http://dbpedia.org/ontology/
- (×) http://iflastandards.info/ns/fr/frad/
- (◊) http://iflastandards.info/ns/fr/frbr/frbrer/
- (▽) http://iflastandards.info/ns/isbd/elements/
- (◻) http://purl.org/dc/elements/1.1/
- (*) http://purl.org/library/
- (◇) http://purl.org/ontology/bibo/
- (⊕) http://purl.org/ontology/mo/
- (✖) http://purl.org/spar/fabio/
- (■) http://rdaregistry.info/Elements/a
- (⊗) http://rdaregistry.info/Elements/c
- (□) http://rdaregistry.info/Elements/e
- (■) http://rdaregistry.info/Elements/i
- (●) http://rdaregistry.info/Elements/m
- (▲) http://rdaregistry.info/Elements/w
- (◆) http://schema.org/
- (●) http://www.bl.uk/schemas/bibliographic/
- (●) http://www.cidoc-crm.org/cidoc-crm/
- (○) http://www.europeana.eu/schemas/edm/
- (□) http://www.loc.gov/mads/rdf/v1
- (◊) http://xmlns.com/foaf/0.1/



RDA AND IFLA FRBR CLASSES



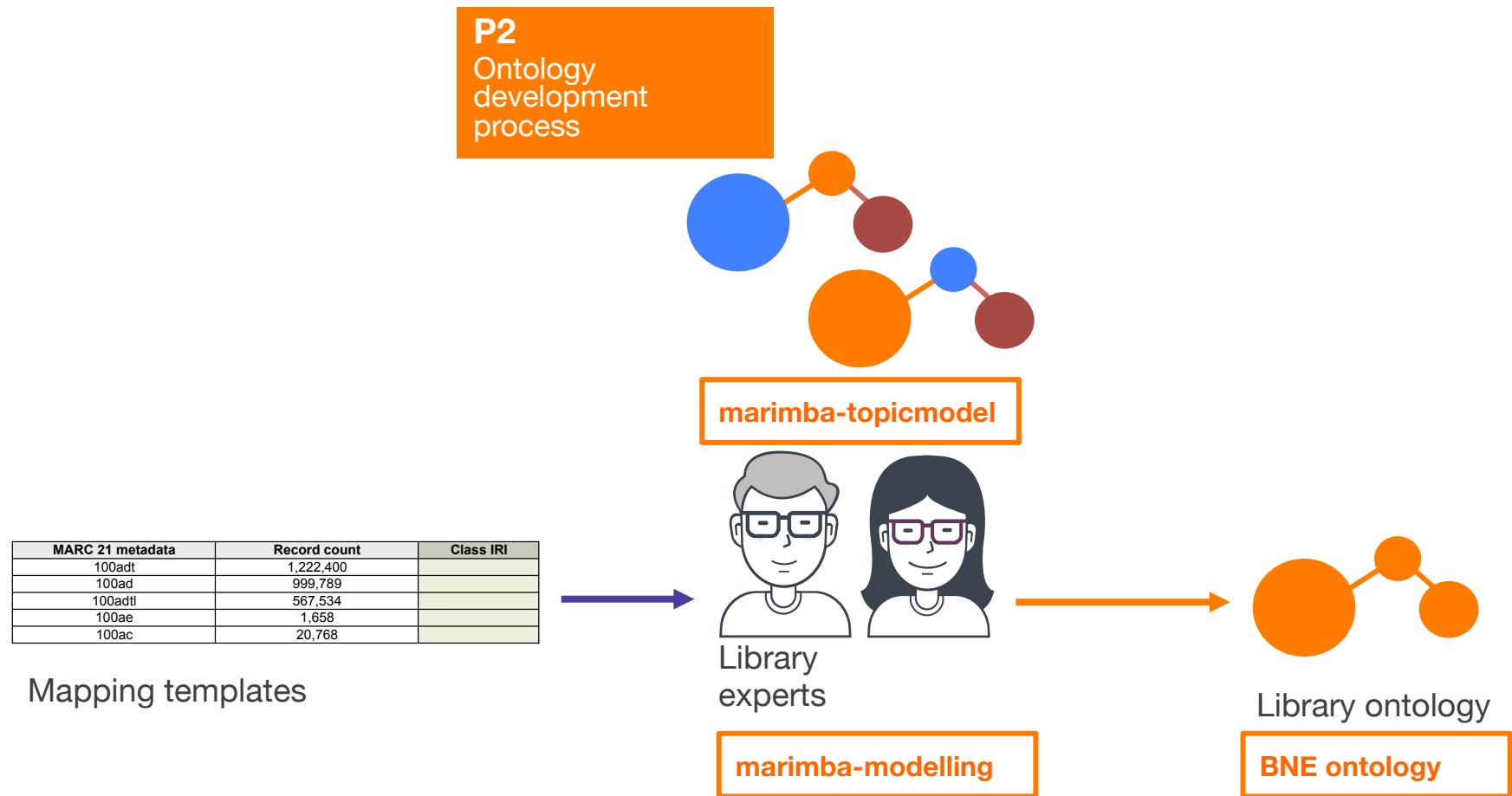
RDA, EDM, BIBFRAME, MADS
FRAD: AUTHORITY DATA AND
CONCEPT OF AGENT AND EVENT



BIBO, FABIO, FOAF PERSON INFORMATION

BIBO, RDA MANIFESTATIONS
FABIO COVERING EDITION
INFORMATION

ONTOLOGY DEVELOPMENT



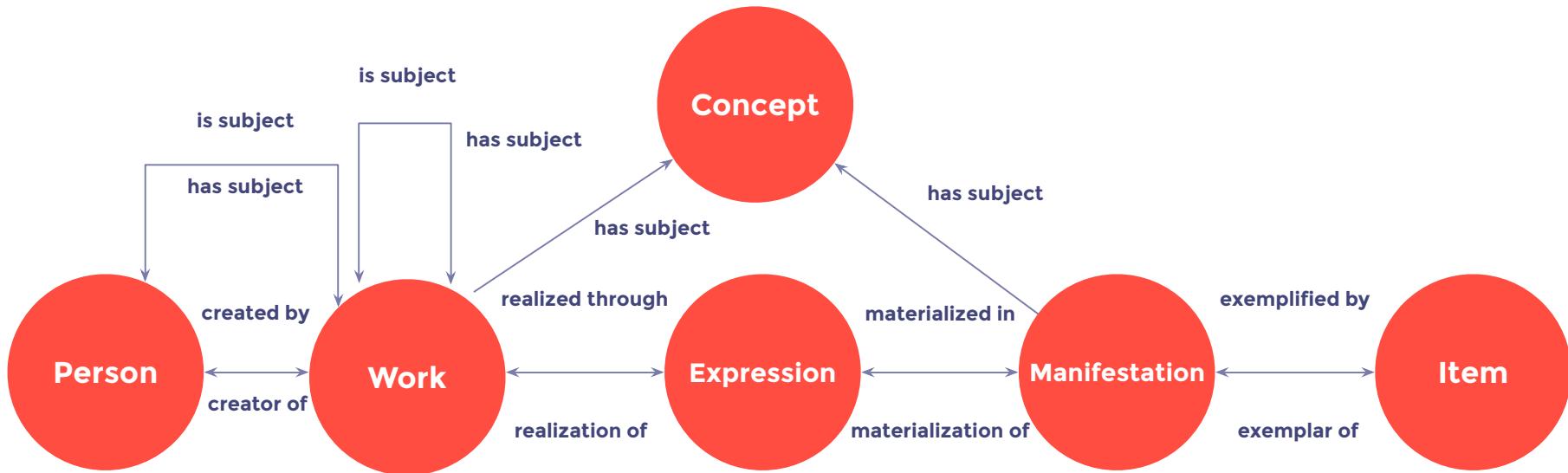
ONTOLOGY DEVELOPMENT



marimba-modelling

marimba-topicmodel

BNE ontology



- Includes **alignments** to library ontologies (ISBD, IFLA FRBR, RDA)
- **6** classes, **33** object properties, **200** datatype properties
- **Ontology localization:** Spanish and English
- **Ontology publication:** HTML doc, RDF formats

marimba-modelling

marimba-topicmodel

BNE ontology

H3

Analytical data and the feedback of library experts can be used to develop a library **ontology** with sufficient **quality** with respect to **state of the art evaluation metrics**.

Evaluation from two perspectives:

1. **Topology based:** OOPS! (Poveda-Villalón et al. [2014]):
0 critical pitfalls, only 3 important.
2. **Application-based:** Task-based experiment with 72 participants.

OUTLINE



P1
Mapping

P2
Ontology
development
process

Empirical studies

datos.bne.es

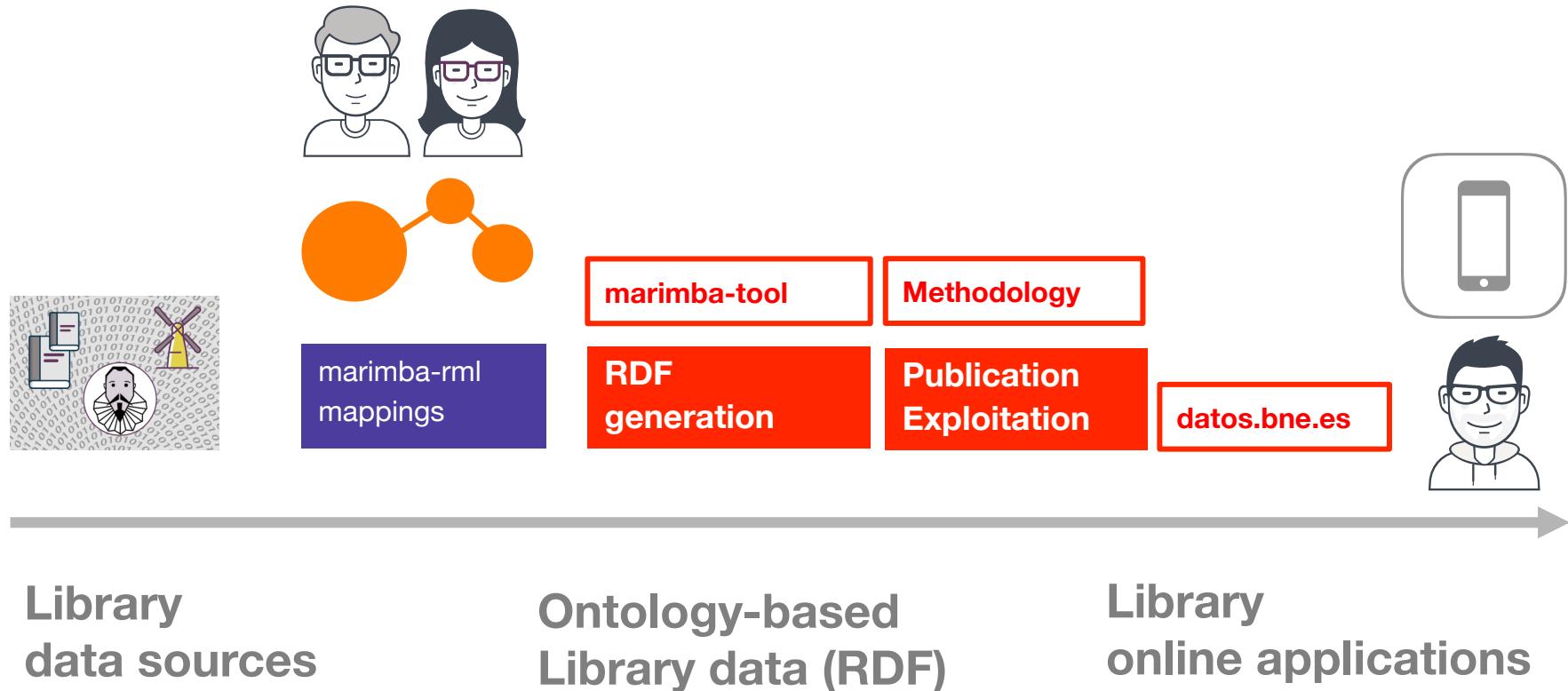
P3
Library
applications

Library
data sources

Ontology-based
Library data (RDF)

Library
online applications

The screenshot shows a search interface for library data. At the top, there is a search bar with the text "ramon y cajal". Below it, a button says "Filtrar por tipo (1192 resultados)". Underneath are three red buttons labeled "1134 Obras", "50 Entidades", and "8 Personas". A red button labeled "L 177 Ediciones" is also present. The text "21 de 1192" indicates the current page. Below this, there are three cards: 1) "Ramón y Cajal, Santiago (1852-1934)" with a portrait of him; 2) "Santiago Ramón y Cajal" with the text "Ramón y Cajal, Santiago" and a red "Obra" button; 3) "Ramón y Cajal (Vídeo)" with a red "Obra" button. To the right of the cards is a large purple arrow pointing right.



datos.bne.es

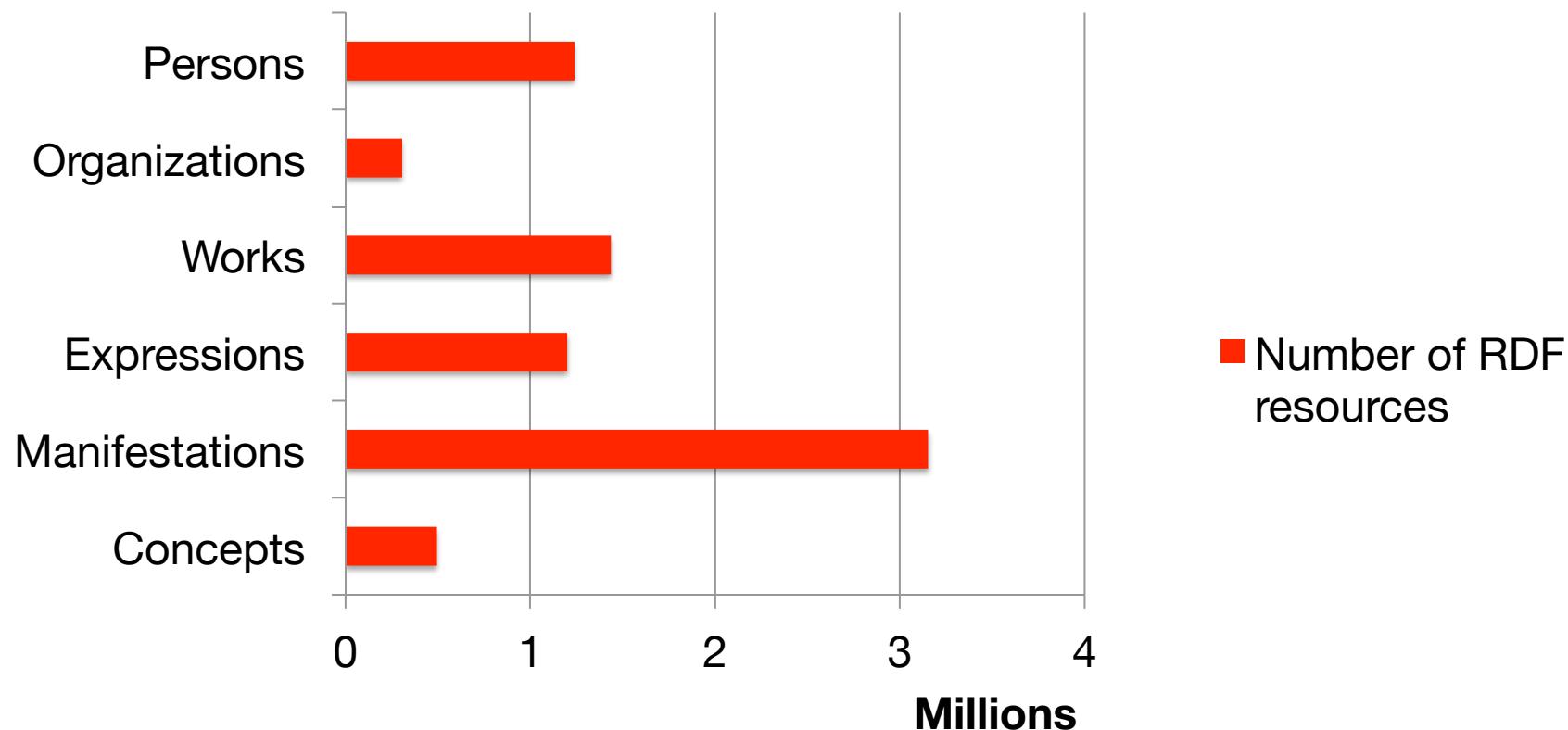
Empirical studies

Task-based experiment

Usability study

PUBLICATION: +7 MILLION RDF RESOURCES, +1M SAMEAS LINKS

Number of RDF resources per class (June 2016)



LIBRARY ONLINE APPLICATIONS

datos.bne.es
[Empirical studies](#)
[Task-based experiment](#)
[Usability study](#)

DATOS·BNE·es *beta

[Inicio](#) [Personas](#) [Entidades](#) [Obras](#) [Temas](#) [Ayuda](#)


1. MULTILINGUAL SEARCH



Filtrar por tipo (663 resultados)

[654 Obras](#) [7 Personas](#) [2 Entidades](#)

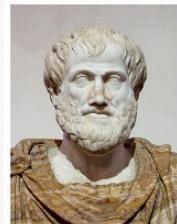
L [907 Ediciones](#)

21 de 663

2. ONTOLOGY ENTITIES

Aristóteles

Aristóteles, en griego antiguo Ἀριστοτέλης Aristotélēs, fue uno de los más influyentes filósofos de la antigüedad, de la historia de la filosofía occidental y considerado por muchos como ...



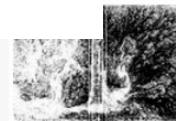
[Obras destacadas](#)

Poética
Ética a Nicómaco
Política
Metafísica

4. DATA ENRICHMENT

Aristóteles

[Obra](#) Disponible en digital



Política

Aristóteles

[Obra](#) Disponible en digital



Ética a Nicómaco

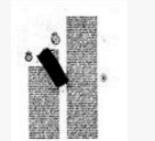
Aristóteles

[Obra](#) Disponible en digital

Metafísica

Aristóteles

[Obra](#) Disponible en digital



datos.bne.es

Empirical studies

Task-based experiment

Usability study

Main questions:

H5

1. Do semantic technologies increase **user satisfaction** and **efficiency**?
2. **Holistic evaluation** of contributions in the thesis

Comparing 2 large-scale services of BNE:

1. Online Public Access Catalogue: **catalogo.bne.es**
2. Linked data-based service: **datos.bne.es**

Two experiments:

1. **Task-based experiment**: 72 participants
2. **Usability and user satisfaction study**: 36 participants

datos.bne.es

[Empirical studies](#)

[Task-based experiment](#)

[Usability study](#)

EXPERIMENTAL SETTING

- **Between-group** experiment:

72 participants of library and information sciences.

- User's goal: Complete an information retrieval **scenario of 10 tasks**.
- **Scenario:**
Documentation about **Spanish scientist** “Santiago Ramón y Cajal”
 - **Designed with professors of Library and information sciences**
 - **Realistic tasks** reproducing a documentation scenario.
 - **Tasks: interrelated and with different levels of complexity.** E.g.,:
“Find dates related to the author”,
“Find the title of three works”,
“Find works by topic”

LIBRARY ONLINE APPLICATIONS

[datos.bne.es](#)

[Empirical studies](#)

[Task-based experiment](#)

[Usability study](#)

EVALUATION APPLICATION

datosbne

Tarea 1. Encuentra información sobre Ramón y Cajal.

Encuentra y anota los años de nacimiento, fallecimiento de Santiago Ramón y Cajal, así como su nombre completo tal y como aparecen en el catálogo.

9
Minutos

37
Segundos

Tiempo restante
■ ¿Has terminado? Puedes aquí



The screenshot shows a task-based experiment interface. At the top, a yellow bar displays a timer: "9 Minutos" on the left and "37 Segundos" on the right. Below the timer is a button labeled "■ ¿Has terminado? Puedes aquí". The main area is titled "DATOS-BNE.es" with a "beta" badge. It features a navigation menu with links to Inicio, Personas, Entidades, Obras, Temas, Ayuda, and the BNE logo. A banner below the menu reads: "El portal de datos bibliográficos de la Biblioteca Nacional de España". Below the banner, a text block explains the purpose of the site: "Datos.bne.es propone al usuario un nuevo modo de acercarse a las colecciones y recursos de la Biblioteca Nacional de España. Es un proyecto de publicación de datos como Linked Open Data, basado en tecnologías y estándares de la Web. Consulta la ayuda para las instrucciones y consejos de búsqueda." A search bar contains the query "Buscar una persona, grupo, entidad, el título de una obra o un tema. Ej.: Leonardo da Vinci, ONU, Grapes of Wrath, /". Below the search bar are two categories: "Personas" and "Obras", each with a thumbnail image. The "Obras" category is currently selected, indicated by a red underline.

OPAC

Tarea 1. Encuentra información sobre Ramón y Cajal.

Encuentra y anota los años de nacimiento, fallecimiento de Santiago Ramón y Cajal, así como su nombre completo tal y como aparecen en el catálogo.

9
Minutos

48
Segundos

Tiempo restante
■ ¿Has terminado? Puedes aquí



The screenshot shows a task-based experiment interface for an OPAC. At the top, a yellow bar displays a timer: "9 Minutos" on the left and "48 Segundos" on the right. Below the timer is a button labeled "■ ¿Has terminado? Puedes aquí". The main area is titled "Catálogo BNE" with a "beta" badge. It features a navigation menu with links to Inicio, Colecciones especiales, Autoridades, Bibliografía Española, Recursos electrónicos, Volver, Ayuda, and Terminar. A banner below the menu reads: "Descubre una nueva forma de acceso a nuestras colecciones y recursos a través de DATOS-BNE.es". Below the banner, there are two search input fields: "Búsqueda sencilla" (with radio buttons for "Palabra clave", "Listado alfabético", and "Exacta") and "Búsqueda avanzada" (with a "Todos los campos" dropdown and a "Buscar" button). To the right of these fields is a sidebar titled "Búsquedas" with links to "Búsqueda avanzada", "Búsqueda alfabética", "Búsqueda por signatura", and "Búsqueda de obras digitalizadas". A note at the bottom left provides instructions for users: "Para cualquier duda consulte en la parte superior la Ayuda. Si lo desea, puede efectuar una Búsqueda por la signatura del documento que desea localizar. En las Colecciones Especiales podrá buscar sobre colecciones destacadas de la Biblioteca Nacional. Si quiere información sobre la descarga de registros puede utilizar la ayuda para la descarga de registros. Si tiene más preguntas o dudas sobre el Catálogo, puede contactarnos en el buzón del Catálogo." Below the note are buttons for "Volver", "Ayuda", and "Terminar".

- **Measures:**
 - Task completion time (Kelly [2009])
 - Visited pages (Su [1992])
 - User satisfaction (1-5 scale) (Brooke [2013])

LIBRARY ONLINE APPLICATIONS

datos.bne.es

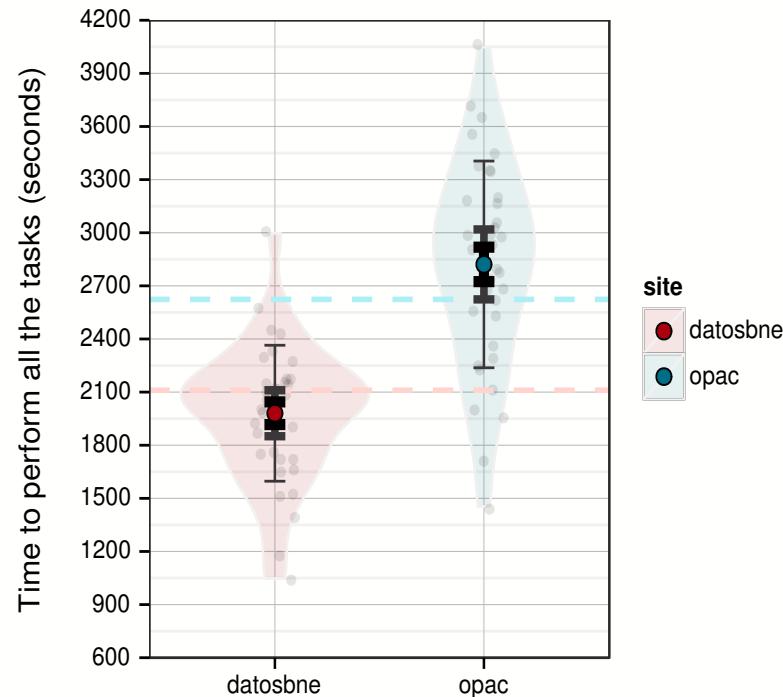
Empirical studies

Task-based experiment

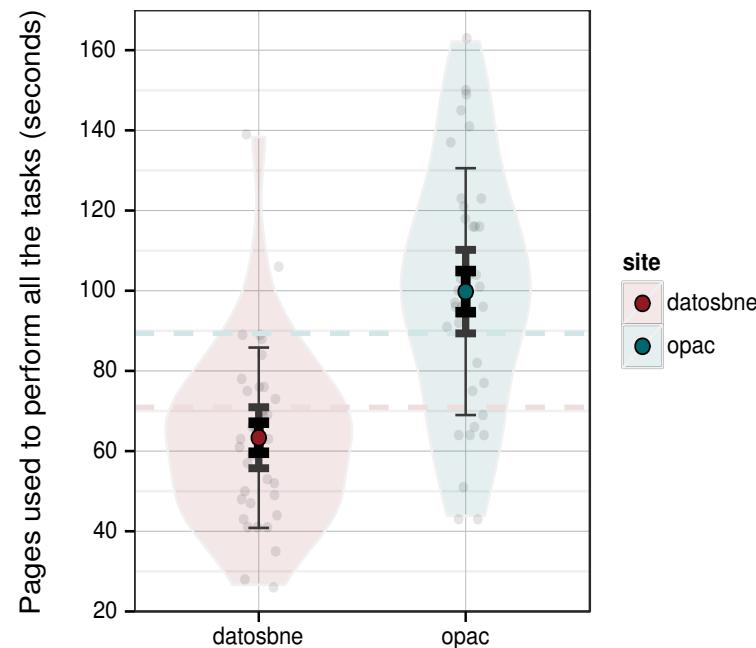
Usability study

Efficiency

- **32% less time** on avg. to complete scenario (**>15 min**)



- **40% less pages** on avg. to complete scenario (**>40 pages**)



H5

Users are significantly more **efficient** in completing information retrieval tasks.

LIBRARY ONLINE APPLICATIONS

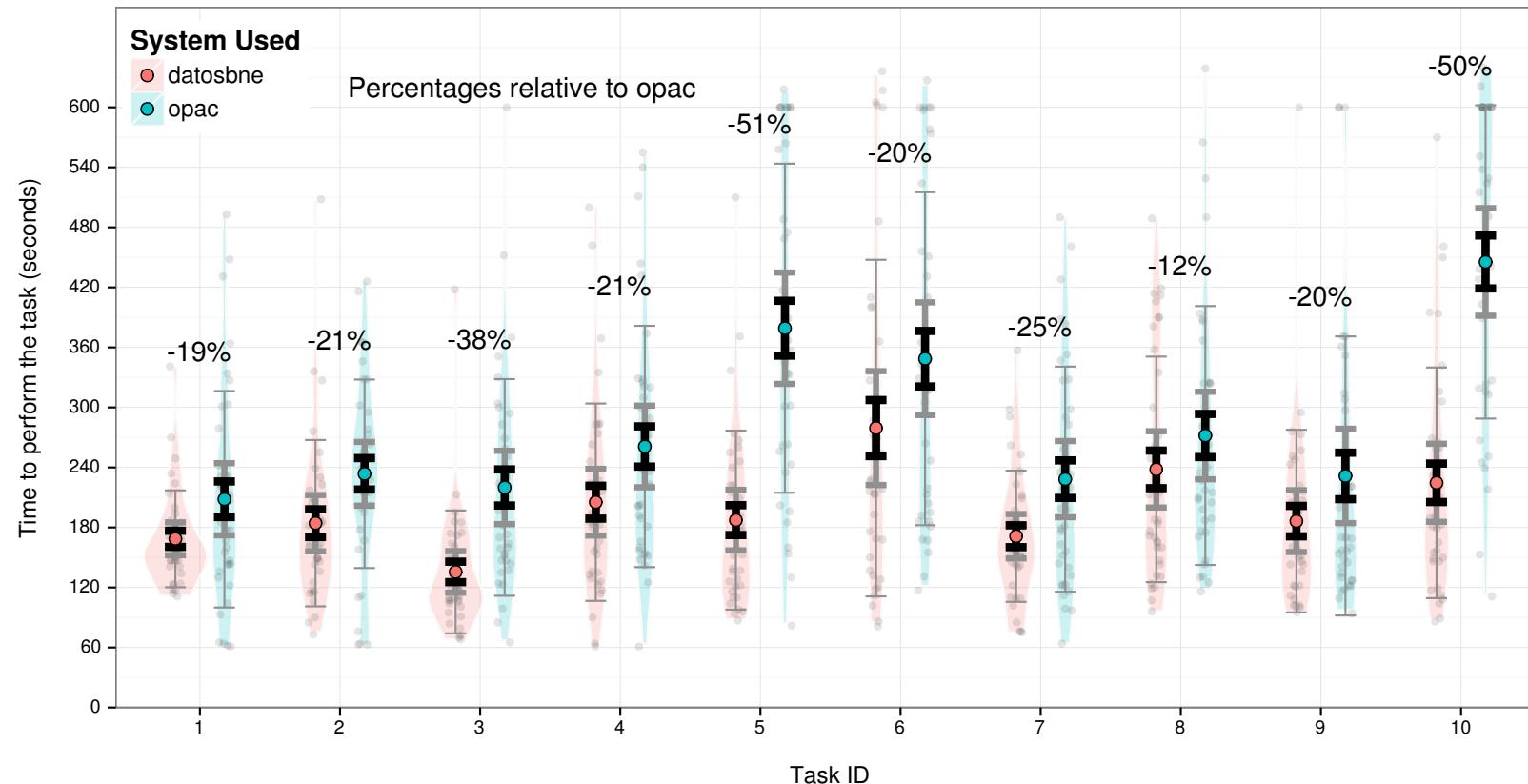
datos.bne.es

Empirical studies

Task-based experiment

Usability study

Time completion per task



- More complex tasks: difference > 50% → Complex navigation of ontology
 - Finding the title of a **translation** into English (Task 5)
 - Finding the **topic** and **author** of a work (Task 10)

datos.bne.es

Empirical studies

Task-based experiment

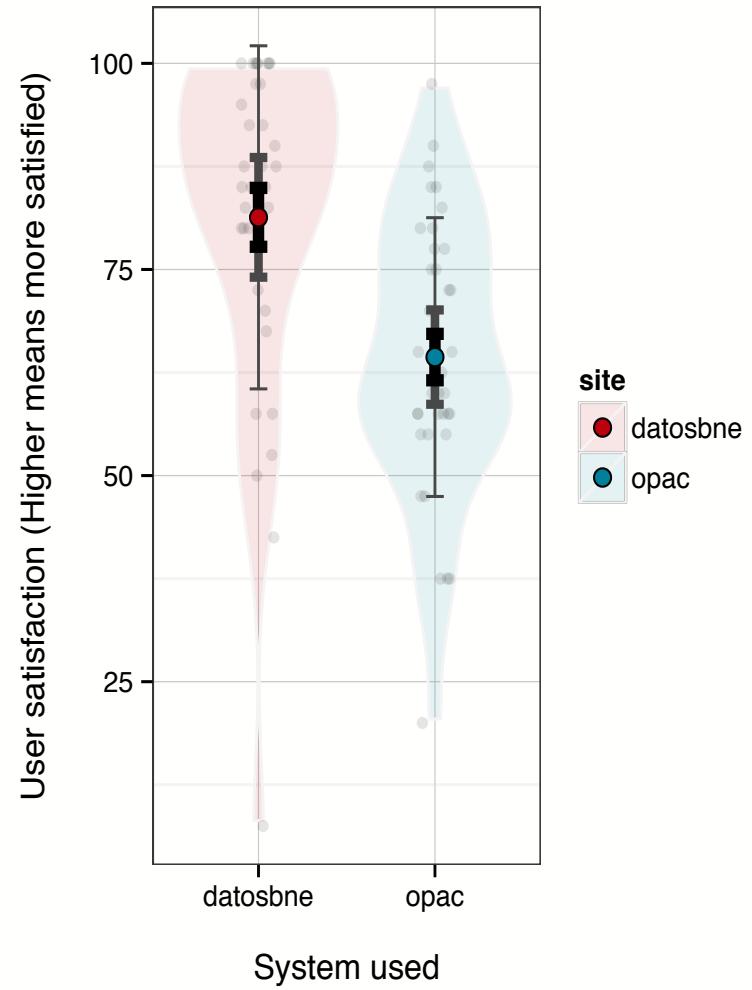
Usability study

Task User Satisfaction

- **datos.bne.es** average score = 81
- **17% higher** than catalogo.bne.es

H5

Significant increase in **user satisfaction** for information retrieval tasks



[datos.bne.es](#)

[Empirical studies](#)

[Task-based experiment](#)

[Usability study](#)

USABILITY STUDY: EXPERIMENTAL DESIGN

- **36 participants** (18 for each system)
- **No time measurement and constraints**
- **No task solving**
- Usability test “Practical Heuristics for Usability Evaluation” (Perlman [1997])

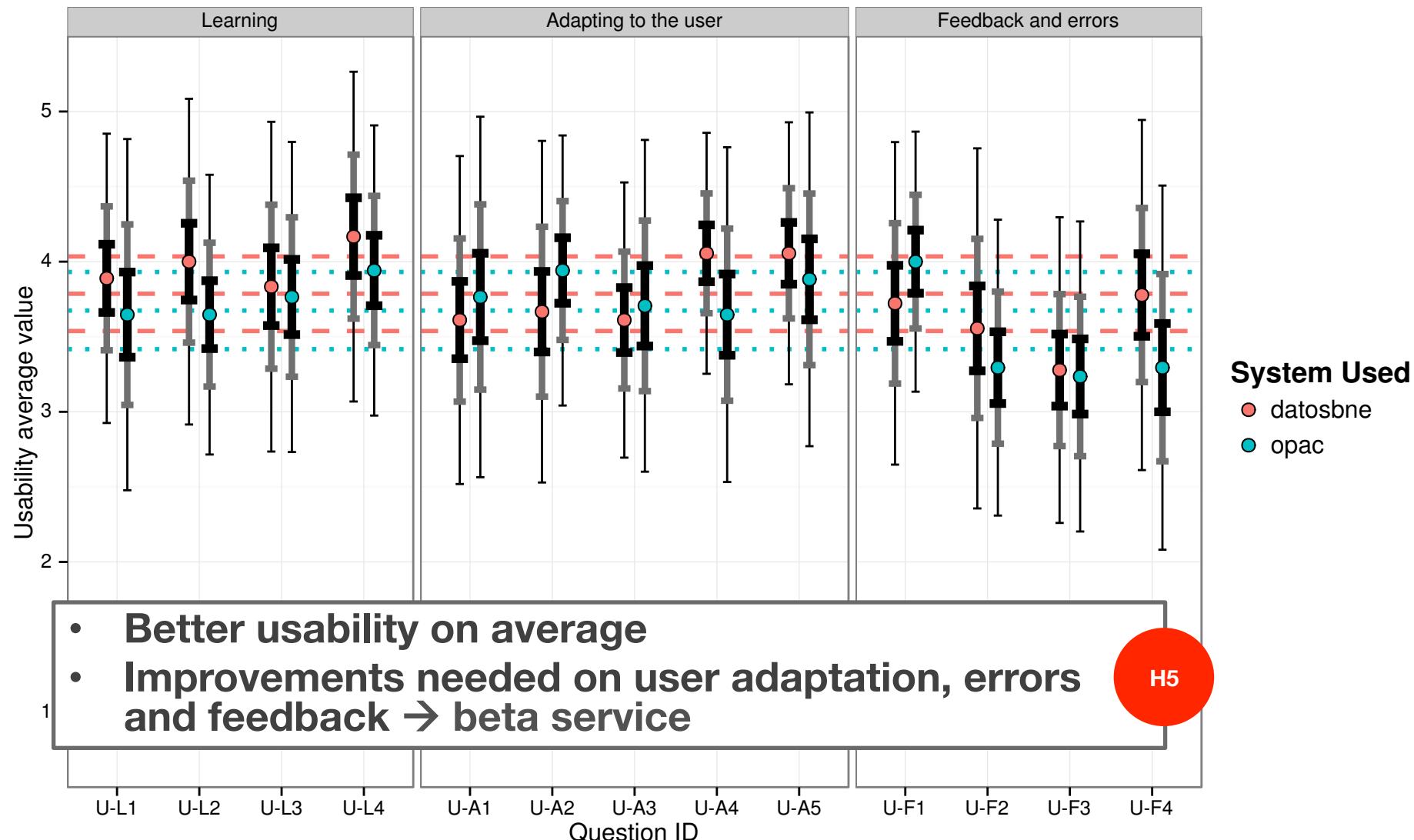
LIBRARY ONLINE APPLICATIONS

datos.bne.es

Empirical studies

Task-based experiment

Usability study



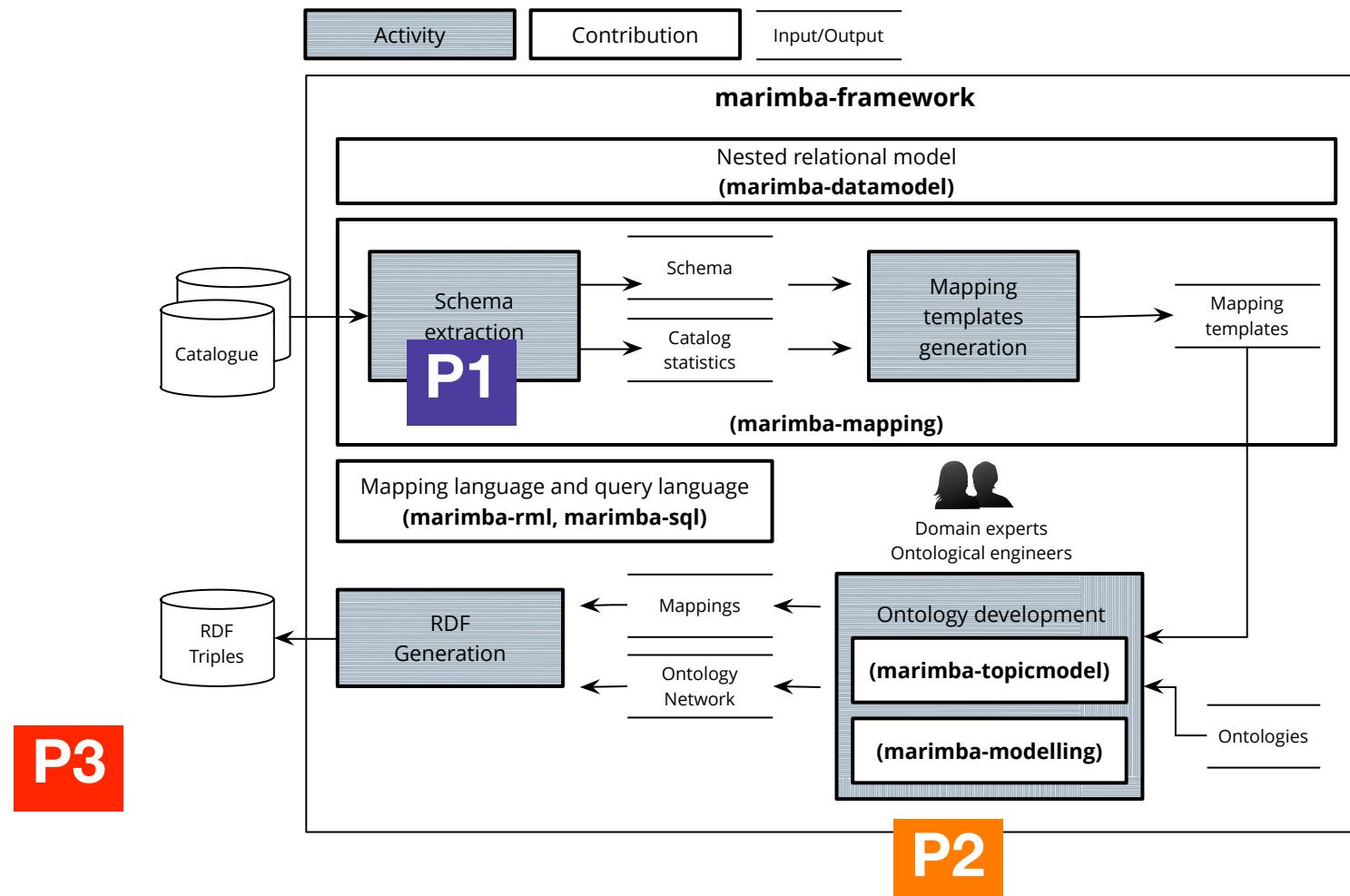
CONCLUSION AND FUTURE DIRECTIONS



CONCLUSION

THEORETICAL, TECHNOLOGICAL AND EMPIRICAL RESULTS

End-to-end solution for ontology-based library data generation, access and exploitation



OTHER RESULTS

DATOS·BNE·es^{*beta}

- **Discoverability:**
 - **45,000 visitors/month**, +1M visitors,
 - **Access through search engines to BNE catalogue has doubled.** Specially “long-tail” queries
- **Reusability:**
 - Used by **+10 third-party library projects**
 - **Wikidata**
 - **Street detective:** Fujitsu Laboratories of Europe Innovation Award 2015
- International **reference library LOD implementation**
- Library science **courses at Universities.**

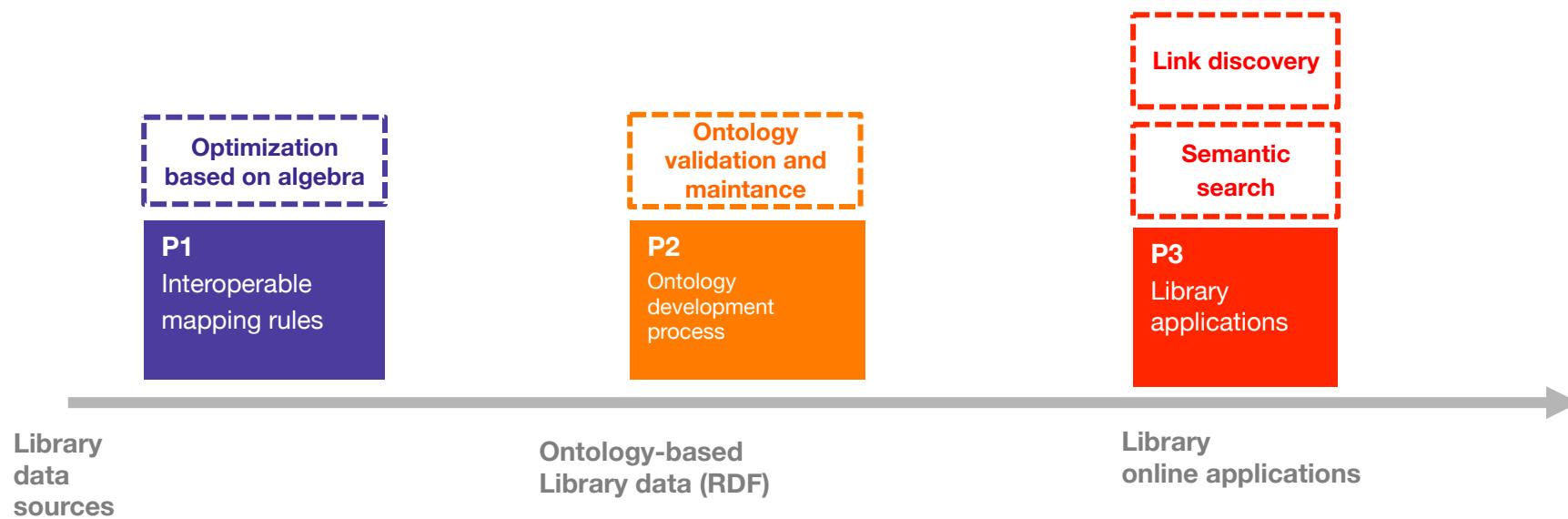
OTHER RESULTS

- **Technology transfer:**
 - **Commercial product:** deployed and licensed to BNE.
 - Requested by **+20 libraries world-wide**.
 - **Diploma** to one of the **best business ideas UPM 2016** (ActuaUPM).
- **Scientific value:**
 - Reference **methodological guidelines**.
 - Topic of **+10 invited talks** and **keynotes** at major library events (e.g., LODLAM summit, EuropeanaTech, SWIB 2012 and 2015).

CONCLUSION



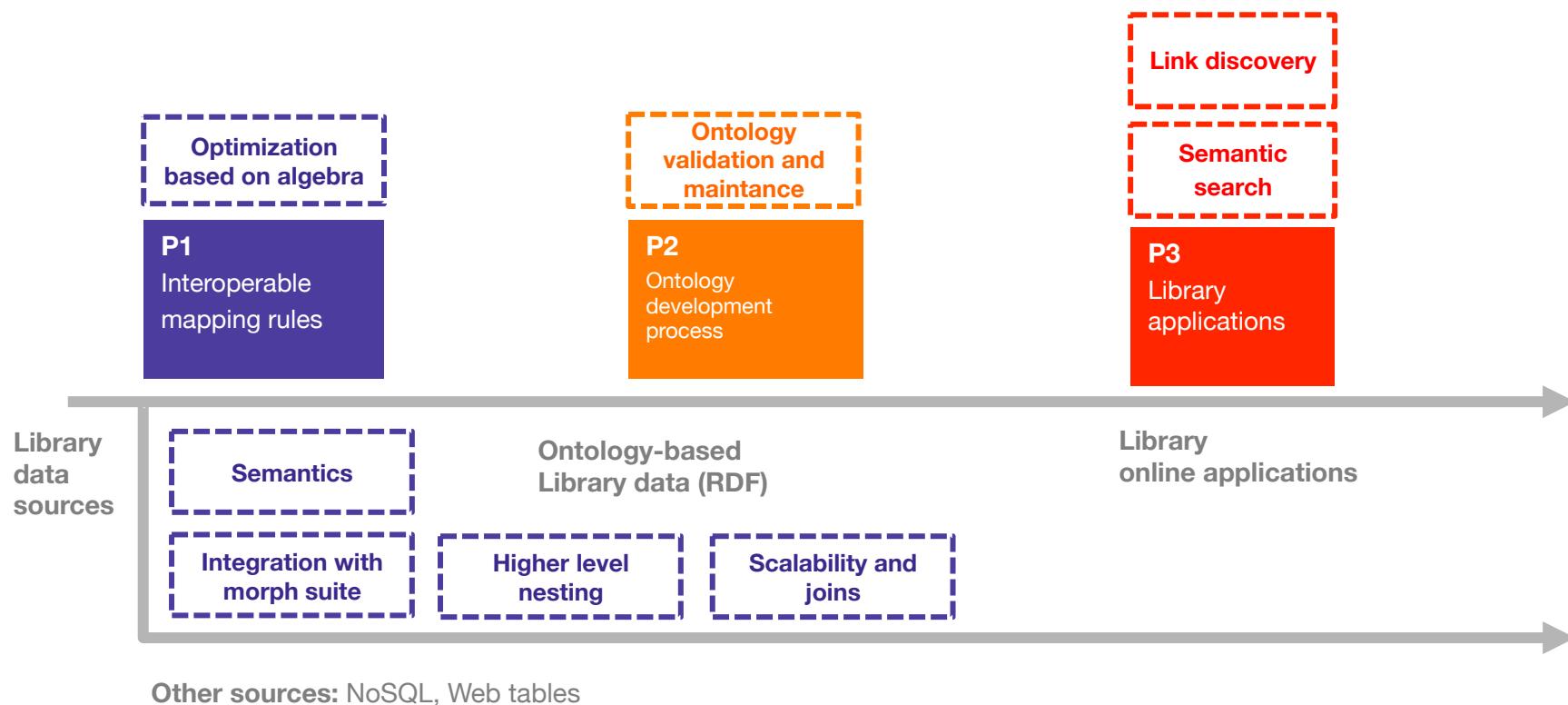
FUTURE DIRECTIONS



CONCLUSION



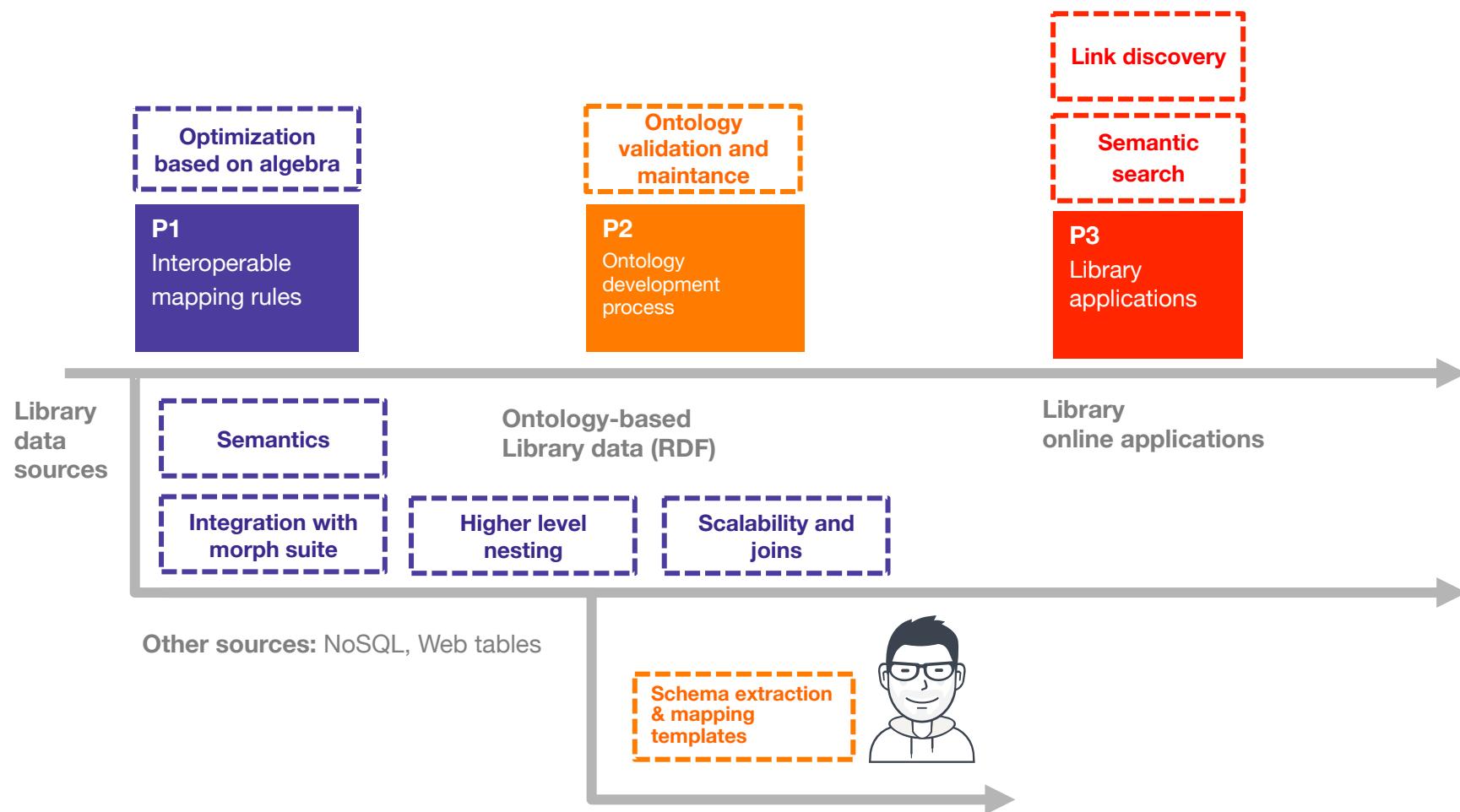
FUTURE DIRECTIONS



CONCLUSION



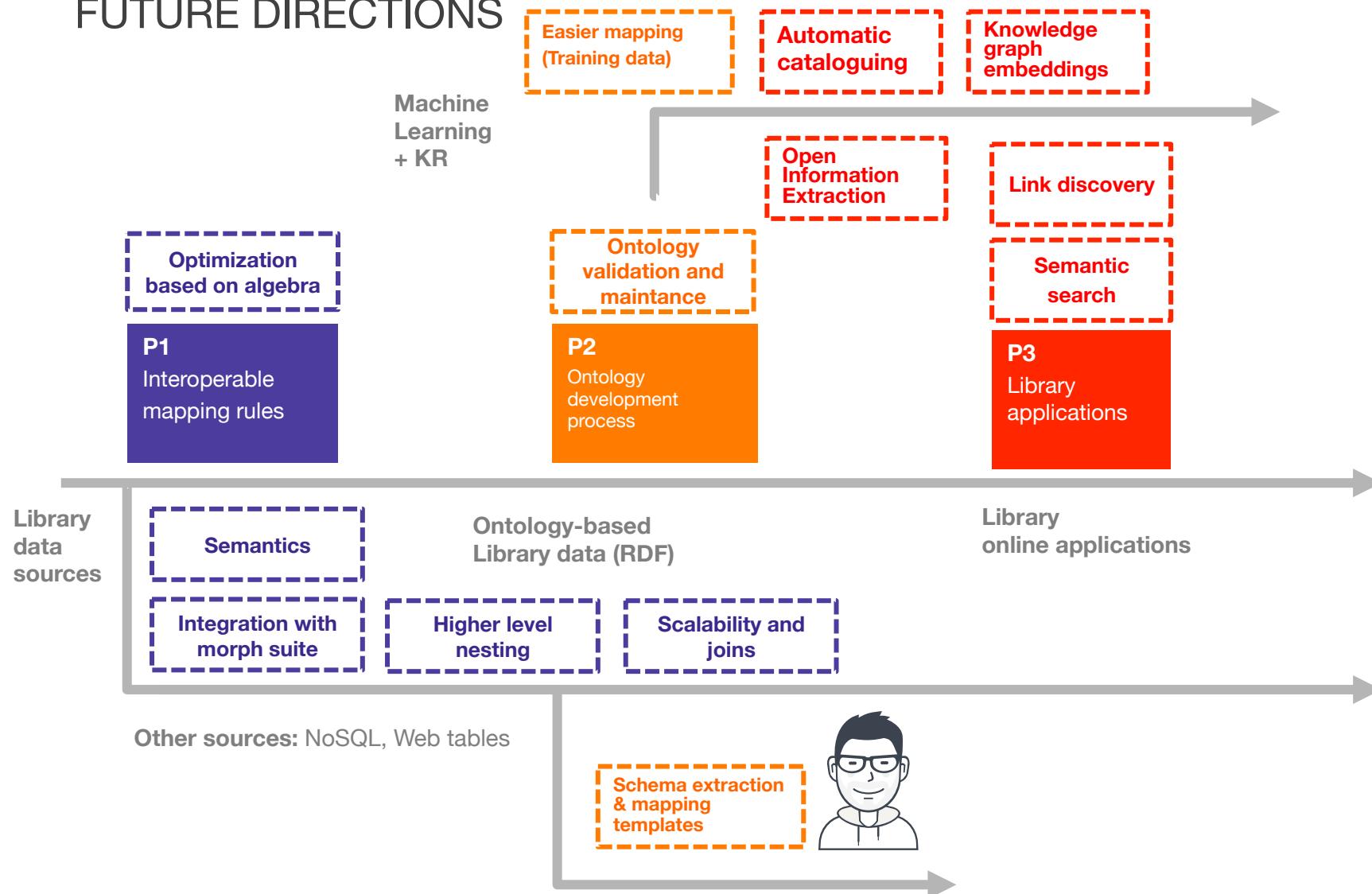
FUTURE DIRECTIONS



CONCLUSION



FUTURE DIRECTIONS



CONCLUSION



SELECTED PUBLICATIONS

• JOURNAL ARTICLES

- Daniel Vila-Suero and Asunción Gómez-Pérez. datos.bne.es and marimba: an insight into library linked data. Library Hi Tech, 31(4):575–601, 2013.
- Daniel Vila-Suero, Boris Villazón-Terrazas, and Asunción Gómez-Pérez. datos.bne.es: A library linked dataset. Semantic Web Journal, 4(3):307–313, 2013.

• INTERNATIONAL CONFERENCES

- Ricardo Santos, Ana Manchado, and Daniel Vila-Suero. Datos.bne.es: a LOD service and a FRBR-modelled access into the library collections. In IFLA World Library International Conference. Cape Town, South Africa, 2015.
- Daniel Vila-Suero, Víctor Rodríguez-Doncel, Asunción Gómez-Pérez, Philipp Cimiano, John McCrae, and Guadalupe Aguado-de Cea. 3ld: Towards high quality, industry-ready linguistic linked linguistic data. European Data Forum 2014, 2014.
- Asunción Gómez-Pérez, Daniel Vila-Suero, Elena Montiel-Ponsoda, Jorge Gracia, and Guadalupe Aguado-de Cea. Guidelines for multilingual linked data. In 3rd International Conference on Web Intelligence, Mining and Semantics, WIMS '13, Madrid, Spain, June 12-14, 2013, page 3.

• OTHER

- Daniel Vila-Suero, Asunción Gómez-Pérez, Elena Montiel-Ponsoda, Jorge Gracia, and Guadalupe Aguado-de Cea. Publishing linked data on the web: The multilingual dimension. In Paul Buitelaar and Philipp Cimiano, editors, Towards the Multilingual Semantic Web, pages 101–117. Book chapter. Springer Berlin Heidelberg, 2014.
- Daniel Vila-Suero. W3C Library Linked Data Incubator Group: Use cases Report. 2011.

RESEARCH STAYS

- INRIA 2012 (3 months) and 2013 (2 months)
- Knowledge Management Institute, Open University, UK: 2012, 1 month working on linked data applications



A FRAMEWORK FOR ONTOLOGY-BASED LIBRARY DATA GENERATION, ACCESS AND EXPLOITATION

Daniel Vila Suero

Advisors:

Prof. Dr. Asunción Gómez-Pérez and Dr. Jorge Gracia del Río

PhD in Artificial Intelligence,
Defense, Madrid, 27th of July 2016