Maximising (Re)Usability of Library metadata using Linked Data

Asunción Gómez-Pérez
Ontology Engineering Group
Universidad Politécnica de Madrid

asun@fi.upm.es

@asungomezperez

Acknowledgments:
Daniel Vila Suero (Library linked data and …… of datos.bne.es
Victor Rodríguez Doncel (licensed linked data)
Jorge Gracia (linguistic linked data)

Context – Ontology Engineering Group

Directors: A. Gómez-Pérez, O. Corcho
Position: 8th in the UPM ranking (200 groups)
Founded: 1994

- **Research Group (30 people)**
- **Experience on**
  1. Ontologies, Semantic Web, Linked Data, Open Data
  2. Semantic E-science
  3. Multilingualism
- **ODI Madrid : Madrid Node of the Open Data Institute**
- **Projects**
  - 27 EU projects (7 as coordinator)
  - 54 National Projects
  - 27 contracts with companies
- **Standardization activities**
  - >25 @ W3C, ISO, OASIS, etc.
- **Impact of publications H-index (scholar)**
  - Asunción Gómez-Pérez (h:50, citations 14852)
  - Oscar Corcho García (h: 36, citations 8152)
- **Services to the Spanish community**
  - **esDbpedia**
  - **linkeddata.es**
  - **vocab.linkeddata.es**

DBpedia español

http://www.oeg-upm.net/

https://github.com/oeg-upm

@oeg-upm

170+ Past Collaborators
50+ Past Visitors

License

- This work is licensed under the Creative Commons Attribution – Non Commercial – Share Alike License
- You are free:

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

- Under the following conditions

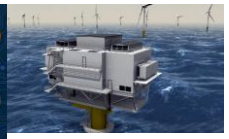
  - Attribution — You must attribute the work by inserting
    - "[source http://www.oeg-upm.net/]" at the footer of each reused slide
    - a credits slide stating: "Maximising (Re)Usability of Library metadata using Linked Data" by A. Gómez-Pérez
  - Non-commercial
  - Share-Alike

Table of Content

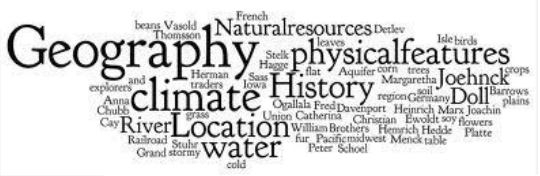
- Motivation
- How to deal with
  - Multilingualism and Language
  - Provenance
  - License
- Linked Data Process
- Uses of library linked metadata

A world of digital data

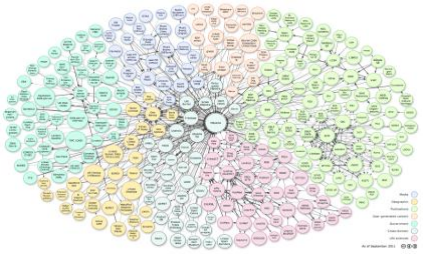
Providers

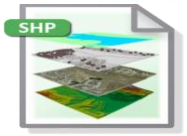


Domains

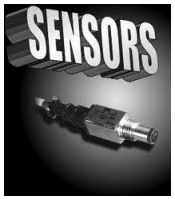


Heterogeneous Formats





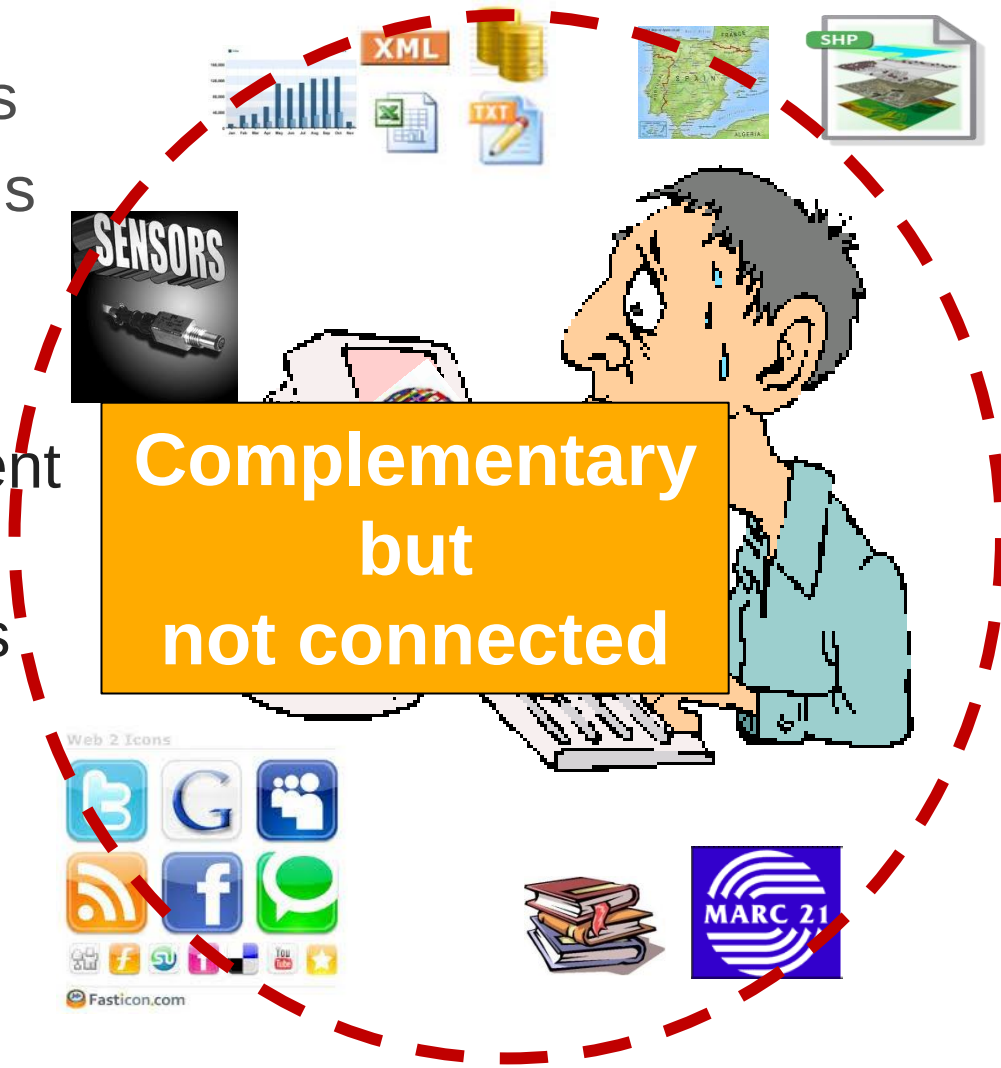
Languages



Licenses

Lack of interoperability: Language, Syntax, Semantic &Technical

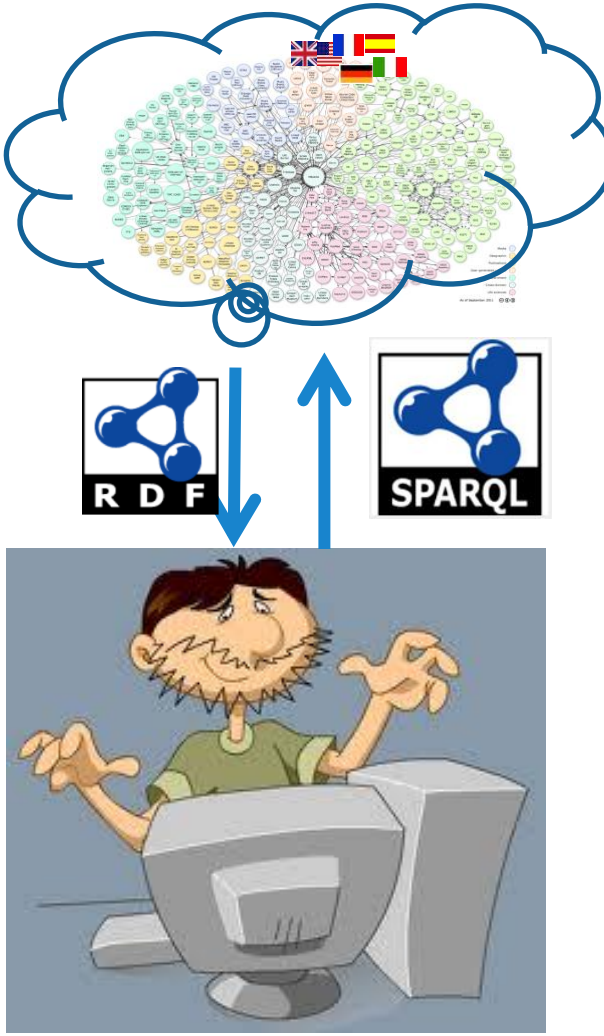
- Ecosystem of
  - Open Resources in silos
  - Complementary domains
  - Heterogeneous formats
  - Different languages
  - Repositories with different metadata
  - Many APIs and services for querying



Data Overload

Linked Data allows uniform access

1. **Agree on vocabularies** for describing
   - metadata
   - domain data
2. Unified and standardized language for describing resources (**RDF(S)**)
3. Unified and standardized **query language** (SPARQL)
4. Standardized non-proprietary APIs
5. Links to other resources



Questions of the end-user …

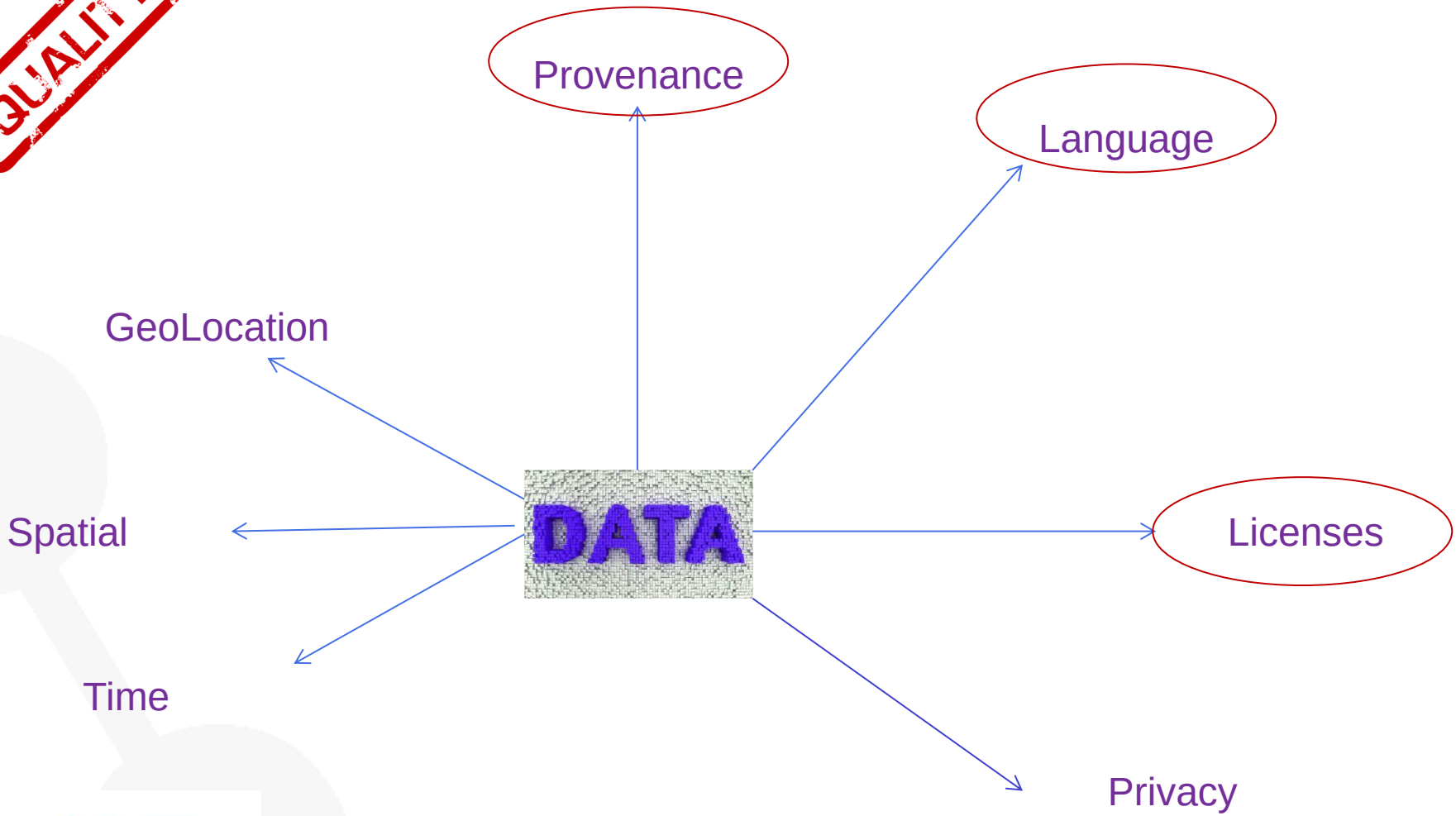
- Who generated the data set
- When the dataset was created?
- How the dataset was built?
- Is that dataset the last version?
- Is license information clearly stated?

Licenses

- Is there any condition that prevents my organization to use the data set?
- In which formats the data are delivered?
- Are data monolingual or multilingual?

Language

Metadata matters



Provides vocabularies for representing these dimensions

www.lider-project.eu

Multilingualism and Language in Linked Data

Libraries store multilingual data

Cervantes Saavedra, Miguel de 1547-1616

Cervantes Saavedra, Miguel de

Cervantes x Saavedra, Miguel de, 1547-1616

Más información

Otros nombres utilizados

Fuente consultada

Información encontrada

100 1 _ ‡a Cervantes Saavedra, Miguel de, ‡d 1547-1616

100 2 _ ‡a Cervantes Saavedra, Miguel de, ‡d 1547-1616

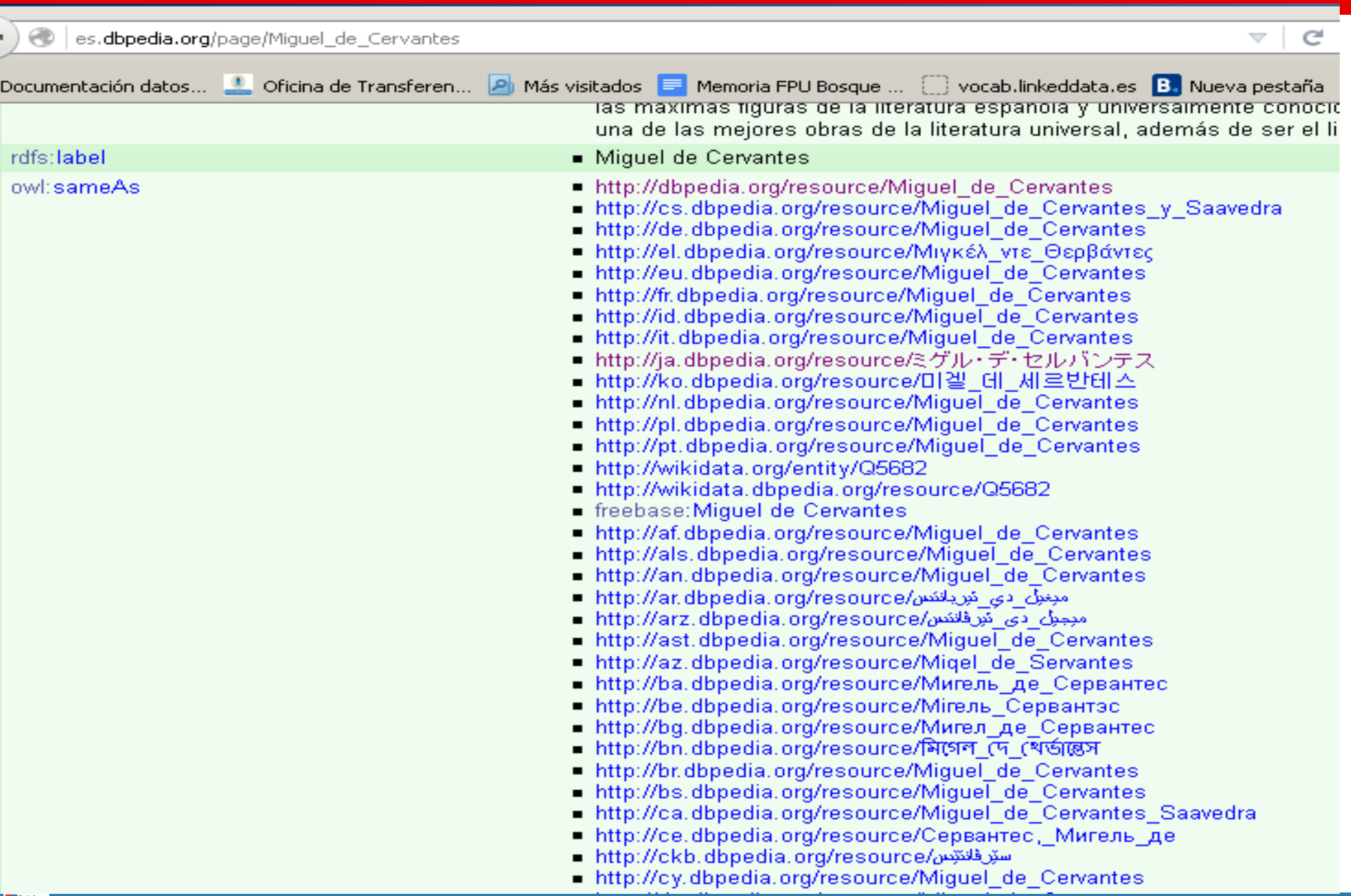
100 1 _ ‡a Cervantes Saavedra, Miguel de, ‡d 1547-1616

Autres formes du titre

data.bnf.fr

- El ingenioso hidalgo Don Quijote de la Mancha *(espagnol; castillan)*
- Don Quijote *(espagnol; castillan)*
- Don Quichotte *(français)*
- Histoire de Don Quichotte de la Manche *(français)*
- Histoire de l'admirable Don Quichotte de la Manche *(français)*
- L'histoire des admirables Don Quichotte et Sancho Pança *(français)*
- L'ingénieux chevalier Don Quichotte de la Manche *(français)*
- L'ingénieux hidalgo Don Quichotte de la Manche *(français)*
- Le noble gentilhomme Don Quichotte de la Manche *(français)*
- O epitídeios eugenís don Kisótīs tīs Mántsas
- Ho epitídeios eugenìs dòn Kisótīs tĩs Mántsas
- Ο επιτήδειος ευγενής δον Κισότης της Μάντσας
- Ὁ ἐπιτήδειος εὐγενὴς δὸν Κισότης τῆς Μάντσας
- Don Kisótīs tīs Mántsas
- Dòn Kisótīs tĩs Mántsas
- Δον Κισότης της Μάντσας
- Δὸν Κισότης τῆς Μάντσας
- L'ingéniós gentilòme dòn Quichòt de la Mancha *(occitan (après 1500); provençal)*
- Dòn Quichòt de la Mancha *(occitan (après 1500); provençal)*

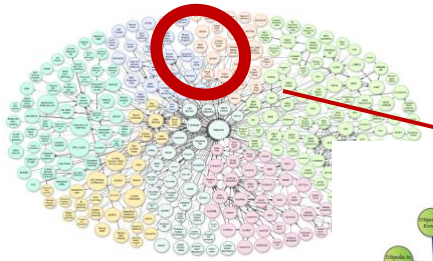
The Web of Data links



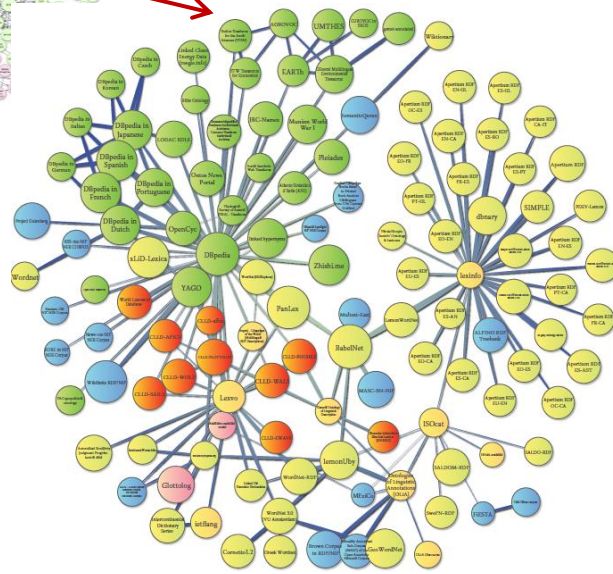
Linguistic Linked Data Cloud

Linguistic Linked Data Cloud

- Subset of LOD
- Linguistic domain
- Many type of resources
- Interconnected with other Language Resources
- Enables the lexicalization of data on the web, not necessarily data in the LD format
- Helps with Multilingual Search
- Enables a new generation of LD-aware NLP and MT Services



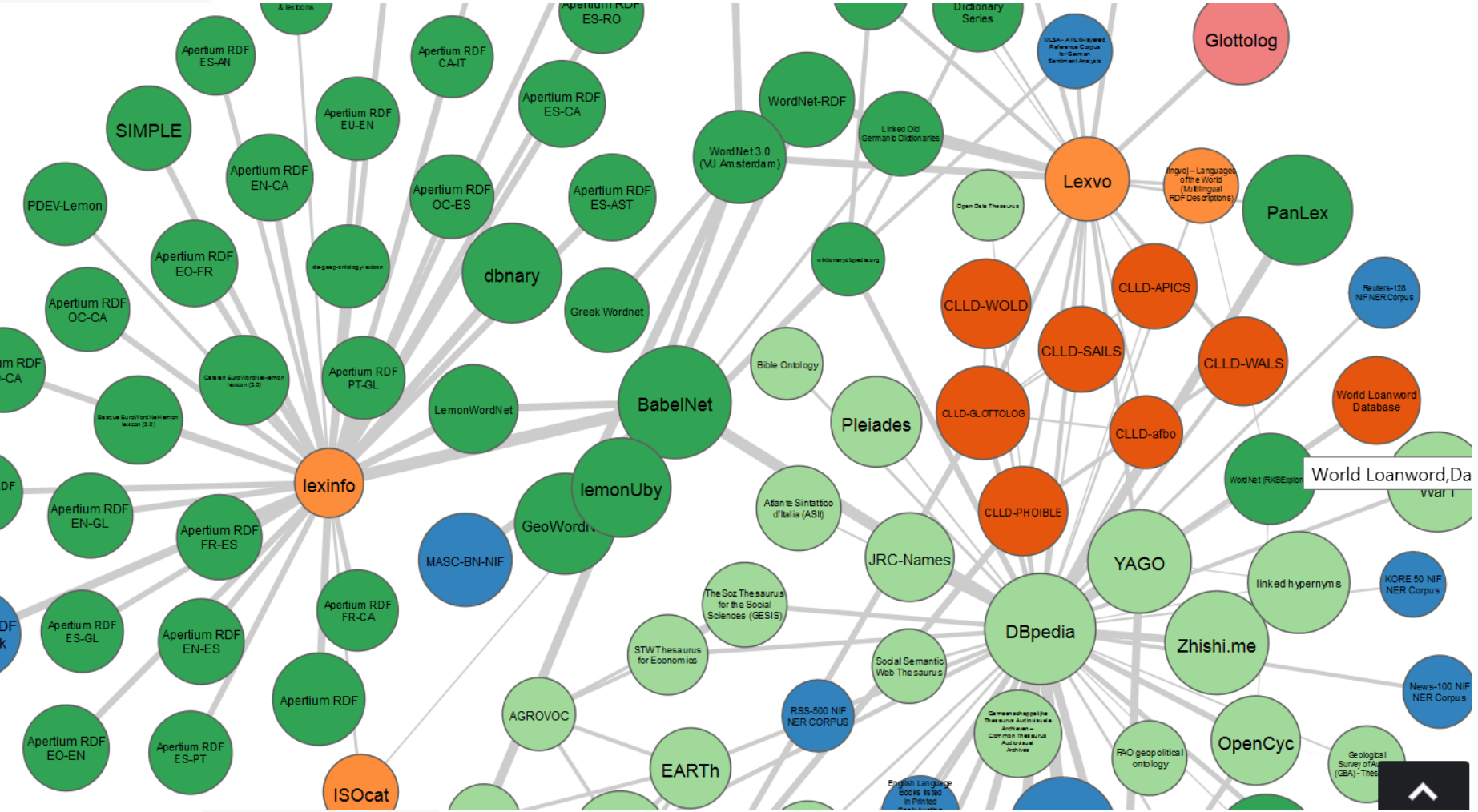
Linguistic Linked Data Cloud



- Corpora
- Terminologies, Thesauri and Knowledge Bases
- Lexicons and Dictionaries
- Linguistic Resource Metadata
- Linguistic Data Categories
- Typological Databases

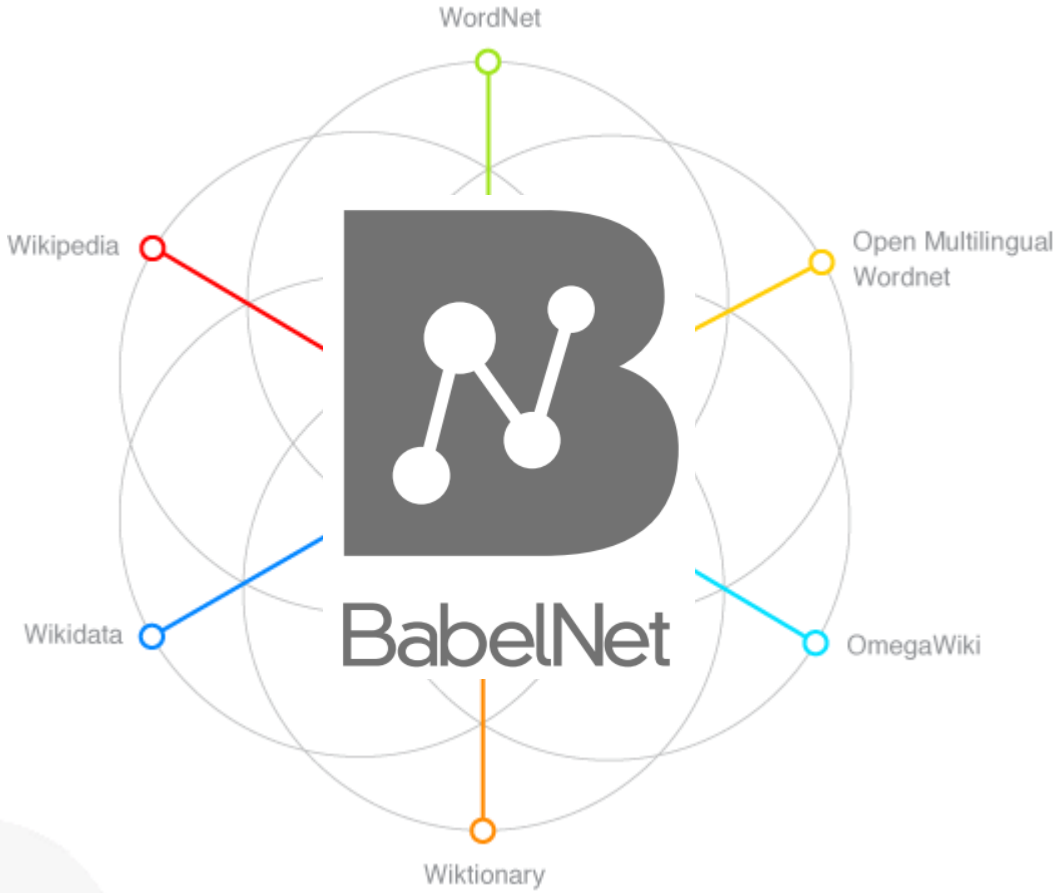
Linguistic Linked Data helps to translate the same term (book title, author name, place, …) in different languages and to deal with acronyms

The core of the Linguistic Linked Open Data cloud!



What is BabelNet?

- A merger of resources of different kinds:



From Roberto Navigli (Sapienza University)

Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 272 languages and 14 million entries!
  - **6M** concepts and **7.7M** named entities
  - **119M** word senses
  - **378M** semantic relations (27 relations per concept on avg.)
  - **11M** images associated with concepts
  - **41M** textual definitions
  - **2M** concepts with domains associated

From Roberto Navigli (Sapienza University)

ENTRAR REGISTRARSE

BabelNet

miguel de cervantes saavedra | SPANISH | TRADUCIR AL... | BUSCAR

miguel de cervantes

Spanish Abkhaz

bn:00017382
nuevos, Novelistas...

Miguel
Miguel de

Miguel de Cerv

Traducciones

AB Мигель де Сервантес, *Миге́ль де Серва́нтес*

Miguel de Cervantes, *cervantes, miguel de cervantes saavedra, cervantes saavedra*

Miguel de Cervantes, *miguel de cervantes saavedra, cervantes, cervantes saavedra*

ALS Miguel de Cervantes

ሚጌል ዴ ሴርቫንቴስ

ميغيل دي ثيربانتس, ميغيل دي ثيرفانتس, سرفانتس, ميغيل دي سرفانتس, *Miguel de Cervantes*, ثيربانتس, سرفانتس, سيربانتس, سيرفانتس, ميجيل دي سيرفاتس, ميجيل دي ثيربانتس, ميخيل دي ثيرفانتس, ميغويل دي سيرفانتس, ميغيل دي ثربانتس, ميغيل دي ثيرفانتيس, ميغيل دي ثيرفانتس, ميغيل دي سيرفانتس, سرفانتس ساڤيدرا, ميغيل دي سرفانتس ساڤيدرا

Miguel de Cervantes, *Miguel de Cervantes Saavedra, Miguel de Cervantes y Saavedra, Miguel de Zerbantes*

HY Միգել դե Սերվանտես, Միգել դը Սերվանտես, Միգել Սերվանտես, *Միգել դը Սերվանտես Սաավեդրա, Cervantes, Miguel de Cervantes, Դը Սերվանտես, Մ., Դը Սերվանտես, Միգել, Դը Սերվանտես-Մ., Դը Սերվանտես Մ., Դը Սերվանտես Միգել, Մ. դը Սերվանտես, Սերվանտես*

ROA Miguel de Cervantes

AST Miguel de Cervantes, *Cervantes, Miguel de Cervantes Saavedra*

AY Miguel de Cervantes

Migel de Servantes, Miqel de Servantes

Мигель де Сервантес, Сервантес, Мигель де

Miguel de Cervantes, *Miguel de Cervantes Saavedra, Mikel Tzerbantesko*

BAR Miguel de Cervantes, *Cervantes*

Мігель Сервантэс, *М. Сервантэс, Мігель дэ Сервантэс*

BE_ Мігель Сэрвантэс

মিগেল দে থের্ভান্তেস, *Miguel de cervantes, Don Miguel Cervantes y Saavedra, Miguel de Cervantes Saavedra, Miguel de Servantes, Miguel Saavedra de Cervantes, Cervantes, Cervantes Saavedra, Cervantes Saavedra, Miguel de, Cervantes Saavedra, Miguel de—Writer, Cervantès, De Cervantes, Don Miguel de Cervantes y Saavedra, Miguel Cervantes, Miguel de Cervantes-Saavedra, Miguel de Cervantes y Saavedra, Saavedra Cervantes, Servantes, মিগেল দি সের্ভান্তিস, মিগেল দে থের্ভান্তেস সাভেদ্রা, সের্ভান্তিস*

g/synset?w

Ontology Engineering Group

Linguistic Linked Licensed Data

www.lider-project.eu

3LD

Linguistic Linked Licensed Data

Language resources such as:

- Lexica

- Corpora

- Dictionaries ..

*Using **RDF** and **standard data models** (vocabularies):*

- Lexica

- Corpora

NIF
NLP Interchange Format

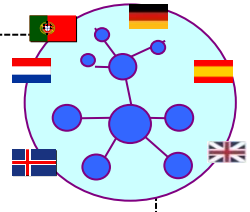
*Published along with a **machine-readable license**.*

ODRL
Open Digital Rights Language

How to represent in Linked Data ...

- Traditional annotation properties to represent language

bne:XX1718747
rdfs:label ”Θερβάντες, Μιγκέλ ντε"@gr.
”Miguel de Cervantes"@es.
“Cervantes di Saavedra, Michele"@it.



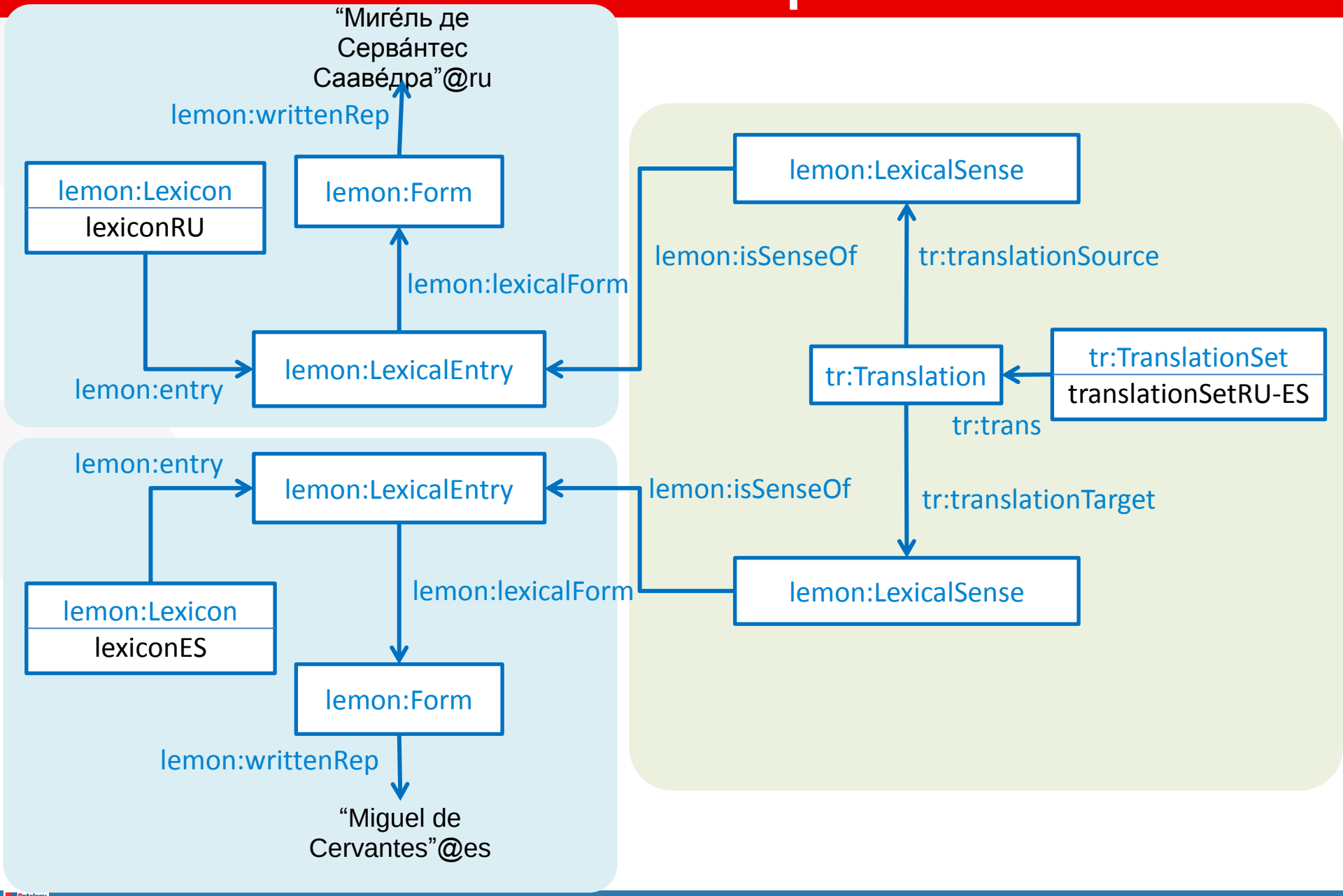
- Richer models to represent linguistic information for more demanding applications

Association of the vocabulary to an external lexicon model: Is author of (Femenine and Masculine)

lemon

```
# LEMON
Bne:OP5001 lemon:isReferenceOf [lemon:isSenseOf :author_of].
:author_of a lemon:LexicalEntry;
lemon:form [lemon:writtenRep “es autor de”@es;
isocat:grammaticalGender isocat:masculine];
lemon:form [lemon:writtenRep “es autora de”@es;
isocat:grammaticalGender isocat:feminine].
isocat:grammaticalGender rdfs:subPropertyOf lemon:property.
```

How to represent translations



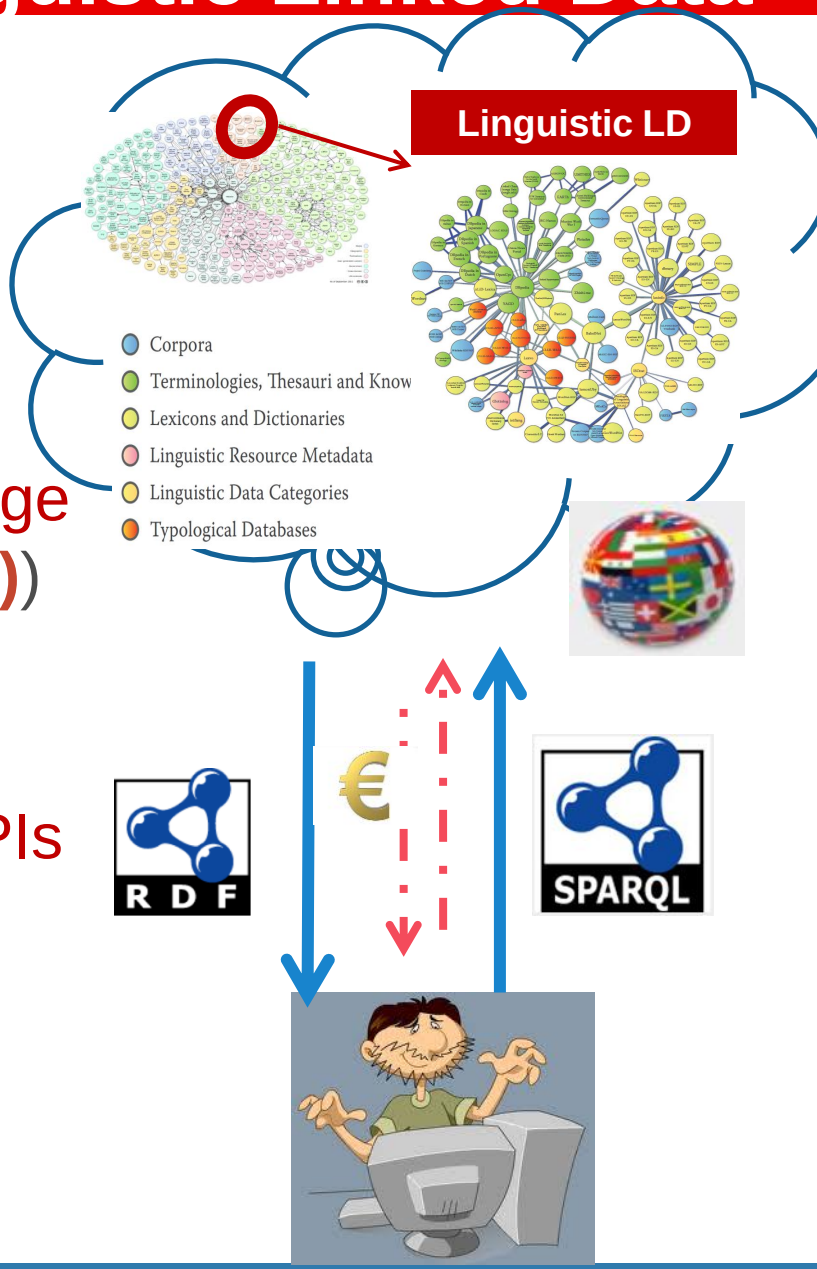
How to add the language to the dataset description

VOID: W3C Vocabulary of Interlinked Datasets
DCAT: W3C Data Catalog vocabulary

```
# VoiD description

:bne a void:Dataset;
dcterms:language <http://id.loc.gov/vocabulary/iso639-1/es> .


# DCAT description
:bne a dcat:Dataset;
dcterms:language <http://id.loc.gov/vocabulary/iso639-1/es>
```

Linked Data and Linguistic Linked Data

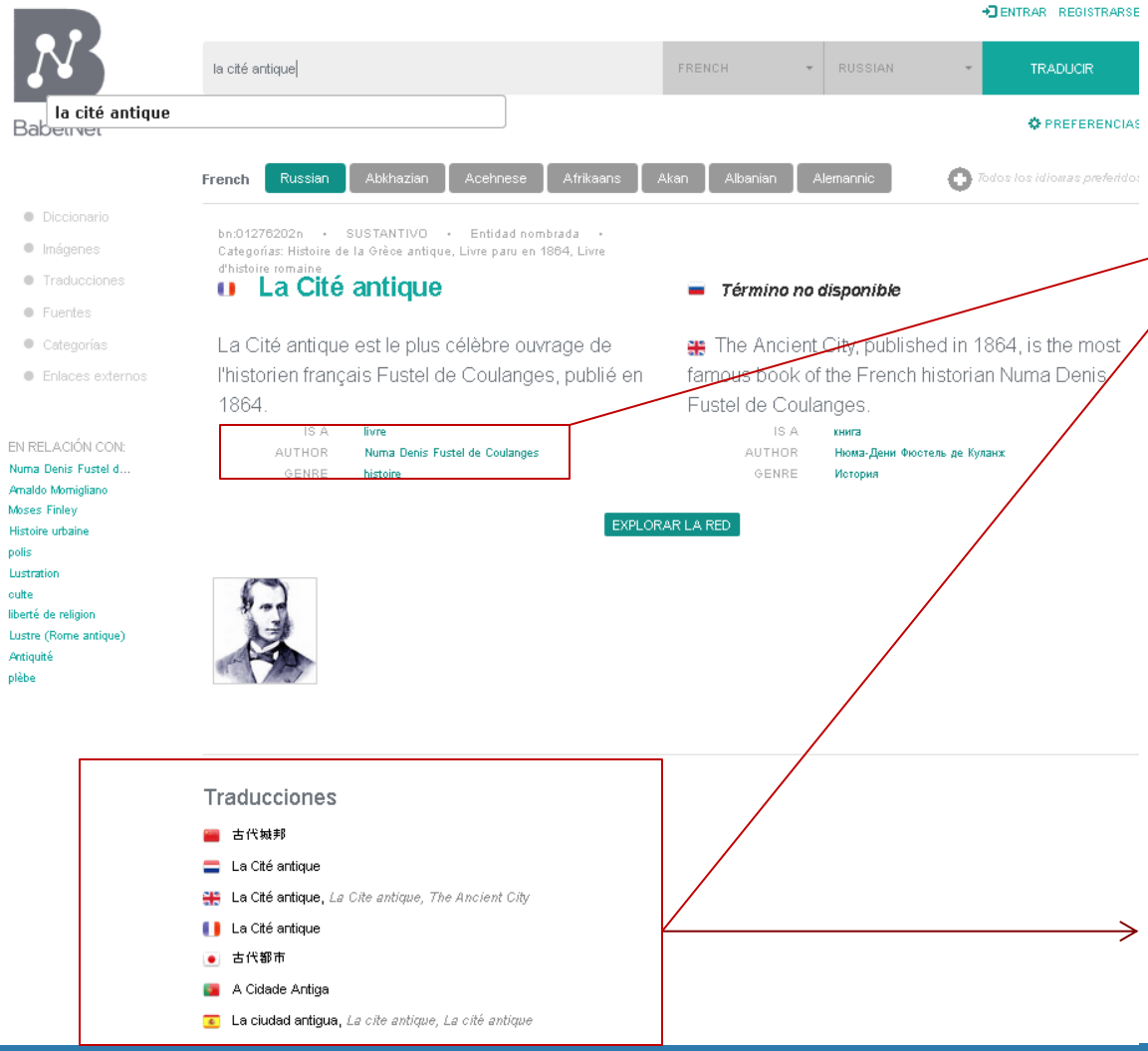
1. **Agree on vocabularies** for describing
   - Domain vocabularies
   - LR metadata and content (Lemon-Ontolex, NIF, …)
2. Unified and standardized language for describing resources (**RDF(S)**)
3. Unified and standardized **query language** (SPARQL)
4. Standardized non-proprietary APIs
5. Links to other resources



Multilingual search: retrieve a book in german

- Servantes Saavedra, Migel
- Therbantes, Minkel nte
- Cervantes di Saavedra, Michele
- سرفنتس ساڤدرا، ميجيل دى
- Zerbantes eta Saabedra, Mikel
- Θερβάντες, Μιγκέλ ντε
- Cervantes
- Sirfantis Saafedrā, Mīgīl dī
- Сервантес Сааведра, Мигель де
- Sewantisi Saweidela, Migai'er de
- 塞万提斯·萨维德拉, 米盖尔 德
- Cervantes, Miguel de

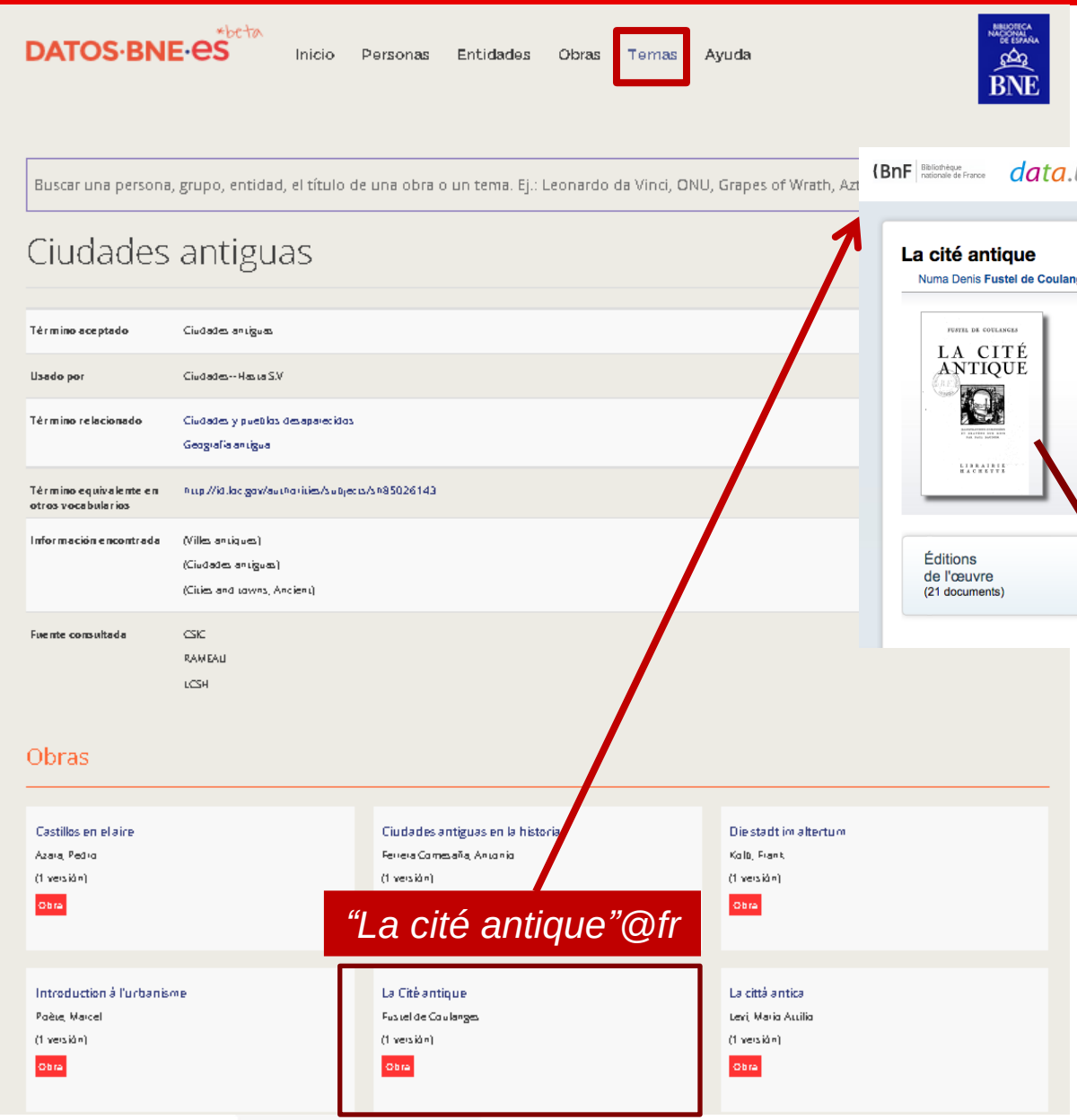
Expand the information based on language

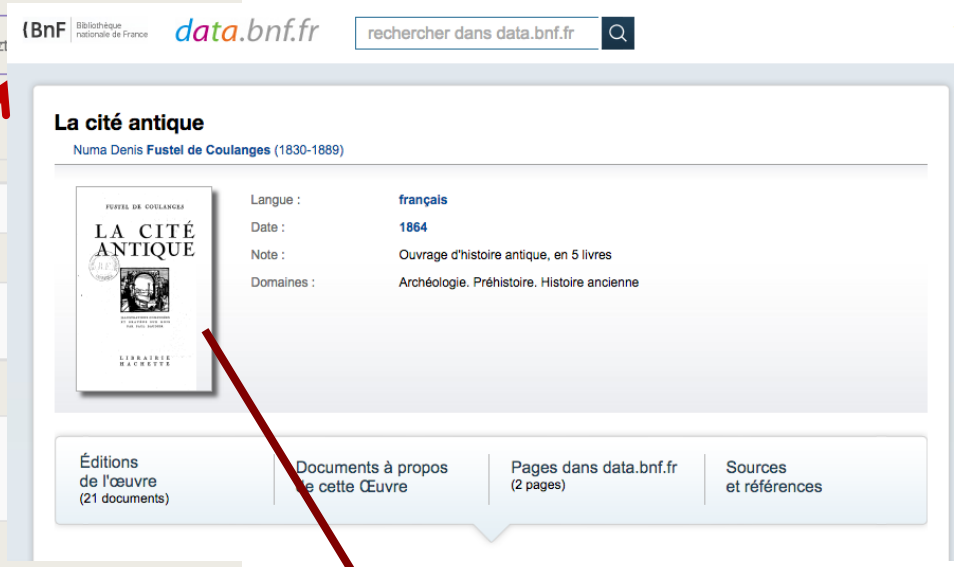


Dbpedia Information about the author in English

Multilingual titles Links (not available in BNF and BNE)

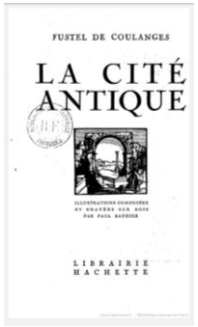
Expand the information based on language





Digital resource available

"La cité antique"@fr



Multilingual Content Aggregation

Full text

LIVRE II.

LA FAMILLE.

CHAPITRE PREMIER.

LA RELIGION A ÉTÉ LE PRINCIPE CONSTITUTIF DE LA FAMILLE ANCIENNE.

Si nous nous transportons par la pensée au milieu de ces anciennes générations d'hommes, nous trouvons dans chaque maison un autel et autour de cet autel la famille assemblée. Elle se réunit chaque matin pour adresser au foyer ses premières prières, chaque soir pour l'invoquer une dernière fois. Dans le courant du jour, elle se réunit encore auprès de lui pour le repas qu'elle se partage pieusement après la prière et la libation. Dans tous ses actes religieux, elle chante en commun des hymnes que ses pères lui ont légués.

Hors de la maison, tout près, dans le champ voisin, il y a un tombeau. C'est la seconde demeure de cette famille. Là reposent en commun plusieurs générations d'ancêtres; la mort ne les a pas séparés. Ils restent groupés dans cette seconde existence, et continuent à former une famille indissoluble (01).

Entre la partie vivante et la partie morte de la famille, il n'y a que cette distance de quelques pas qui sépare la maison du tombeau. A certains jours, qui sont déterminés pour chacun par sa religion domestique, les vivants se réunissent auprès des ancêtres. Ils leur portent le repas funèbre, leur versent le lait et le vin, déposent les gâteaux et les fruits, ou brûlent pour eux la chairs d'une victime. En échange de ces offrandes, ils réclament leur protection; ils les appellent leurs dieux, et leur demandent de rendre le champ fertile, la maison prospère, les cœurs vertueux.

Le principe de la famille antique n'est pas uniquement la génération. Ce qui le prouve, c'est que la sœur n'est pas dans la famille ce qu'y est le frère, c'est que le fils émancipé ou la fille mariée cesse complétement d'en faire partie, ce sont enfin plusieurs dispositions importantes des lois grecques et romaines que nous aurons l'occasion d'examiner plus loin.

Le principe de la famille n'est pas non plus l'affection naturelle. Car le droit grec et le droit romain ne tiennent aucun compte de ce sentiment. Il peut, exister au fond des cœurs, il n'est, rien dans le droit. Le père peut chérir sa fille, mais non pas lui léguer son bien. Les lois de succession, c'est-à-dire parmi les lois celles qui

DBpedia

“La cité antique”

More links

Link to full text:
http://remacle.org/bloodwolf/livres/Fustel/intro.htm

Multilingual and complementary information

DATOS·BNE·es *beta

“La cité antique”@es

Work by F. de Coulanges

Subject: “Ciudades antiguas”@es

Only available non-digitized manifestations in Spanish (title “La ciudad antigua”).

BabelNet

“La cité antique”

Translations: “The Ancient City”@en, ” 古代城邦”@zh

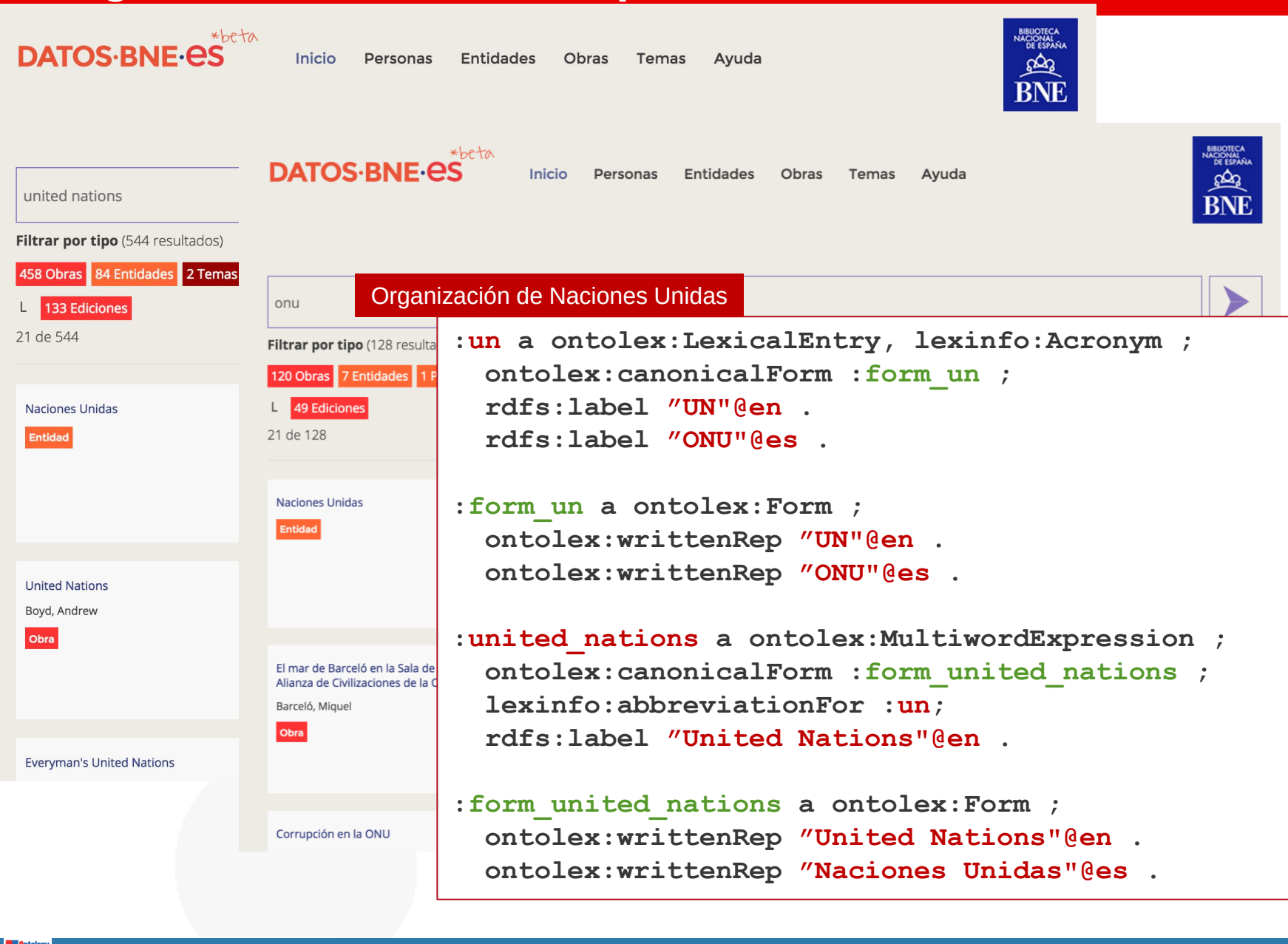
Links: Dbpedia, Wikidata

data.bnf.fr

“La cité antique”

Date: 1864
Digitization available:
http://gallica.bnf.fr/ark:/12148/bpt6k6105986m

Linguistic Linked Data helps with alternative names and titles




```
:un a ontolex:LexicalEntry, lexinfo:Acronym ;
  ontolex:canonicalForm :form_un ;
  rdfs:label "UN"@en .
  rdfs:label "ONU"@es .

:form_un a ontolex:Form ;
  ontolex:writtenRep "UN"@en .
  ontolex:writtenRep "ONU"@es .

:united_nations a ontolex:MultiwordExpression ;
  ontolex:canonicalForm :form_united_nations ;
  lexinfo:abbreviationFor :un;
  rdfs:label "United Nations"@en .

:form_united_nations a ontolex:Form ;
  ontolex:writtenRep "United Nations"@en .
  ontolex:writtenRep "Naciones Unidas"@es .
```

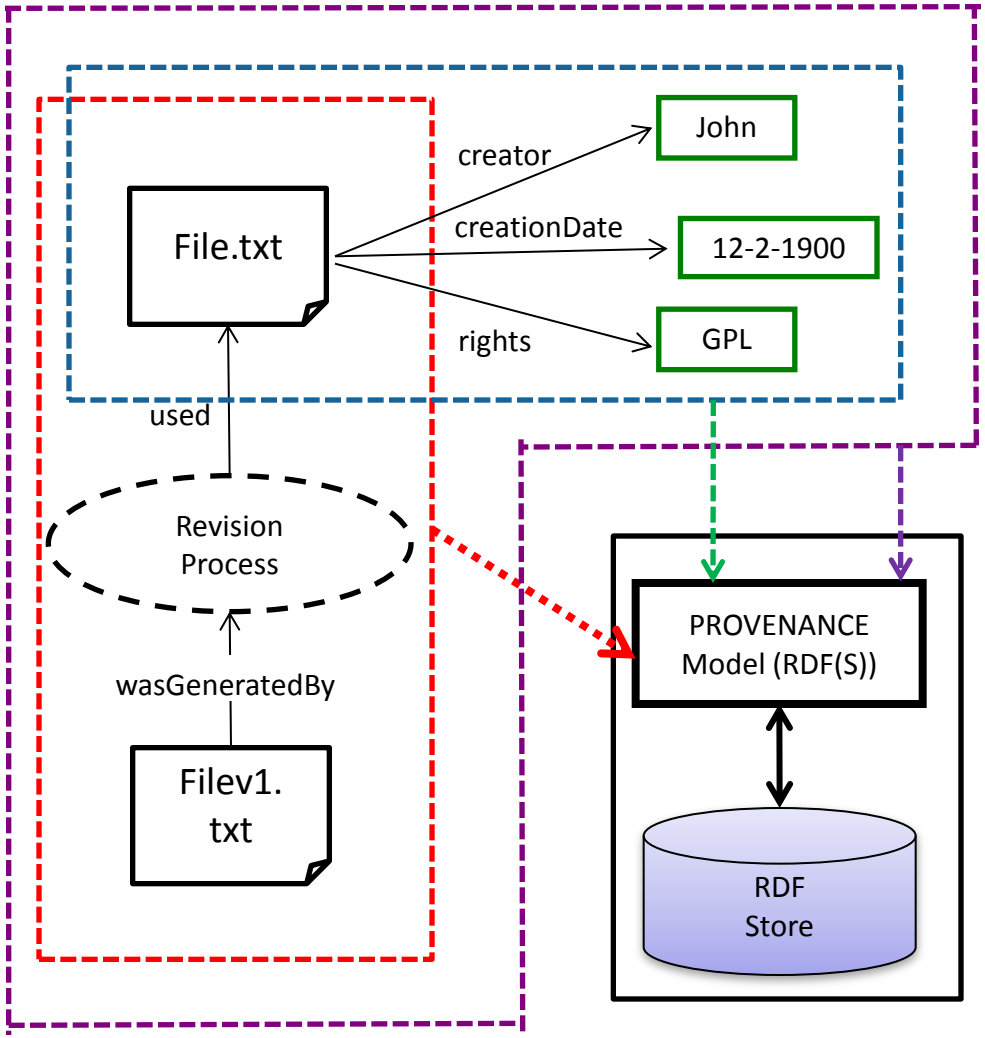
Linked Data have Provenance

Modeling provenance

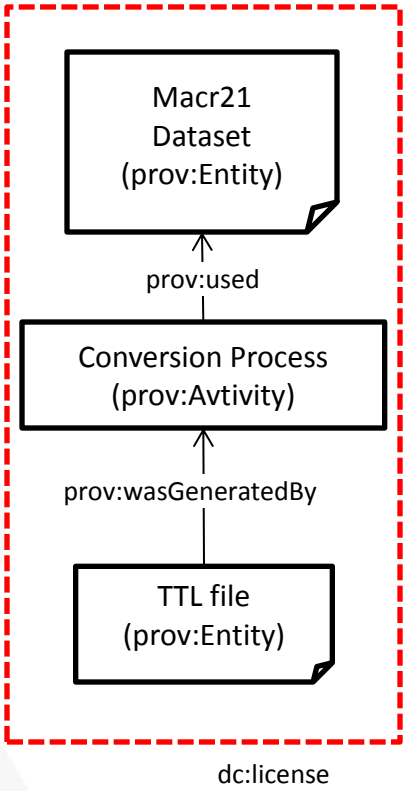
Process centric provenance
- **PROV-O @W3C**

Resource provenance
- **DC, PROV-O, Premis, SWANL**
- **EDM (including agregation)**

Metadata provenance
- **DC, PROV @ W3C**

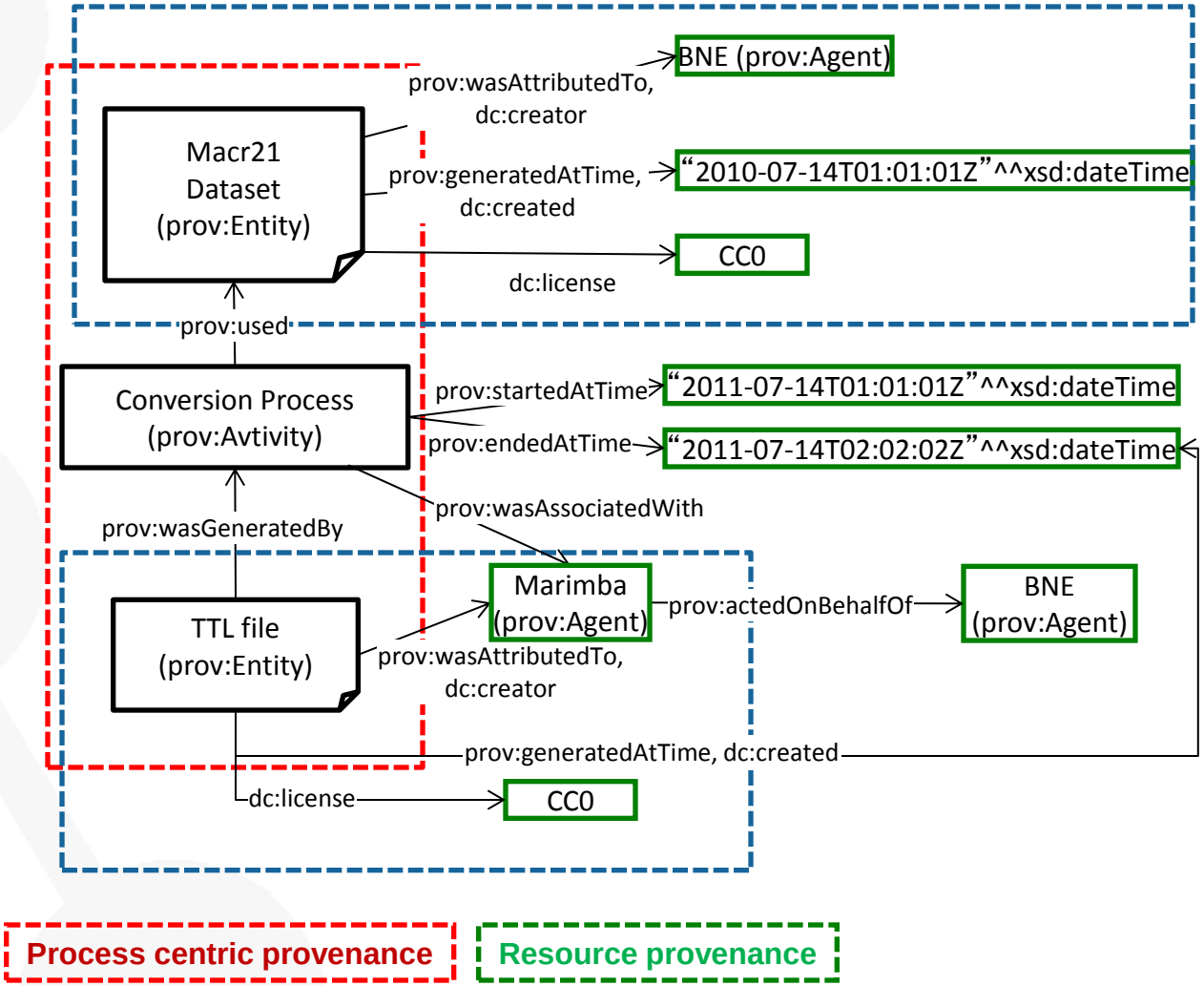


Example of provenance with datos.bne.es



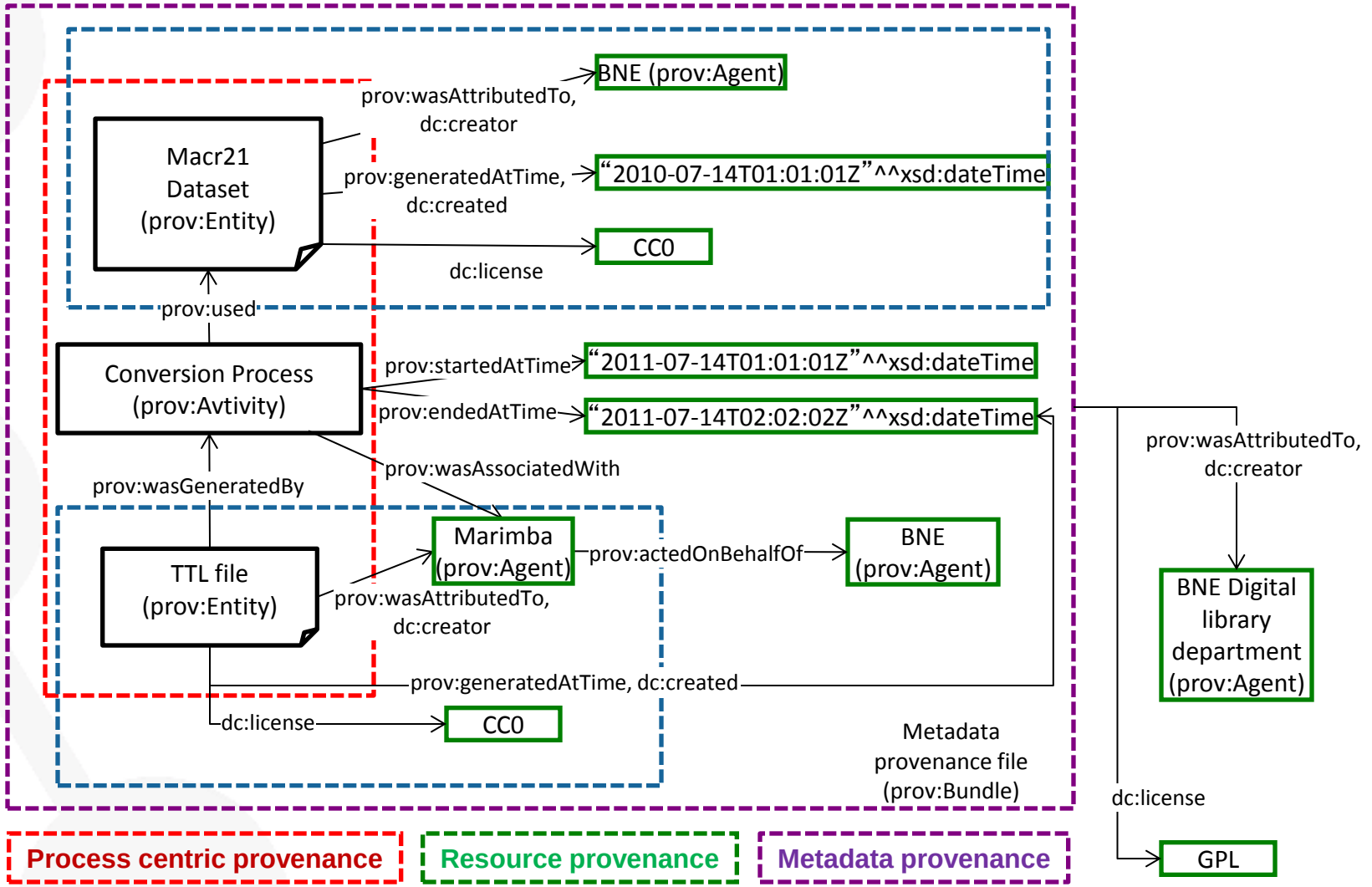
Process centric provenance

Example of provenance with datos.bne.es



M3ilD Hearing Luxembourg, 19 June 2012

Example of provenance with datos.bne.es



Linked Data have Licenses

Research funded by the project
4V: Volumen, Velocidad, Variedad y Validez en la gestión innovadora de datos (TIN2013-46238-C4-2-R)

Always license your data

Data shops

Governments

Individuals

License A

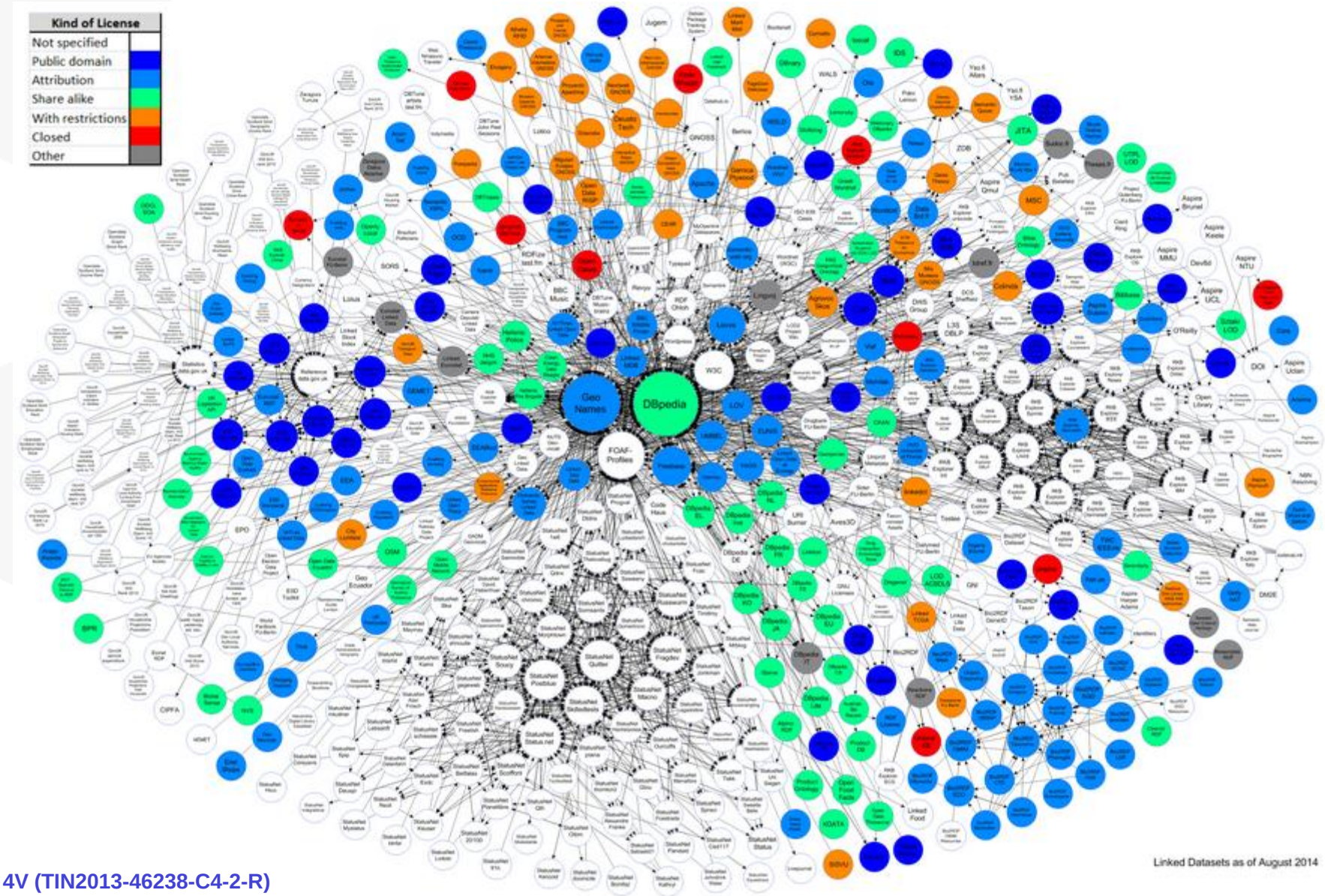
License B

License C

Create, consume, aggregate,
derive and publish Linked Data
in a lawful environment

4V (TIN2013-46238-C4-2-R)

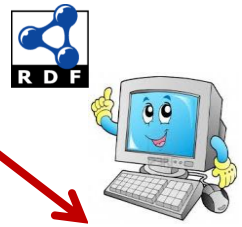
How Open is the Linked Open Data Cloud?

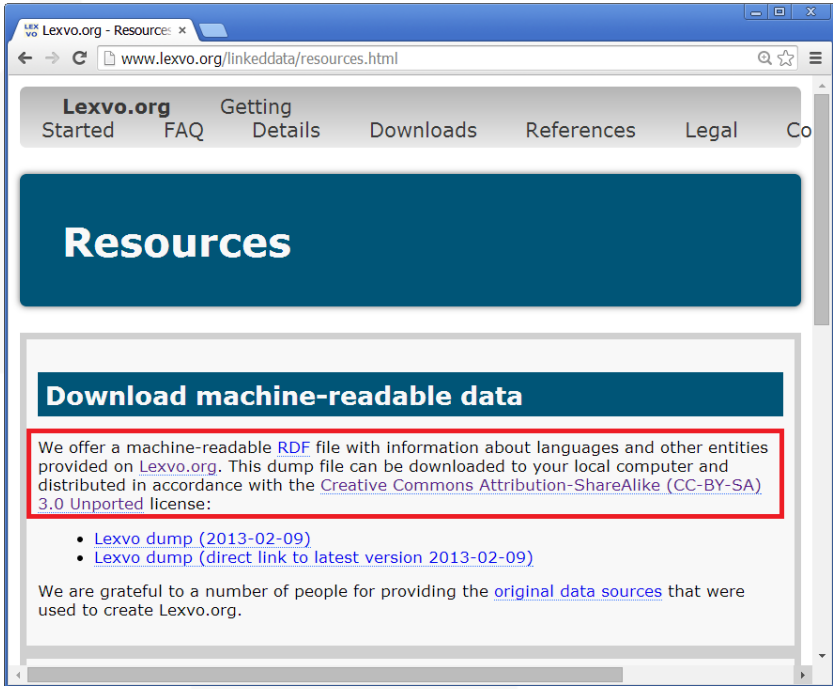


4V (TIN2013-46238-C4-2-R)

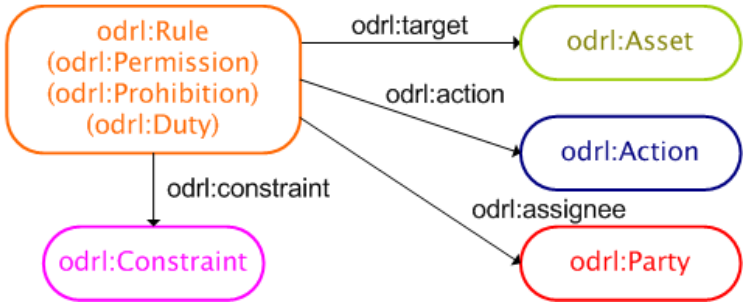
Linguistic Linked Licensed Data

How do we represent license information?



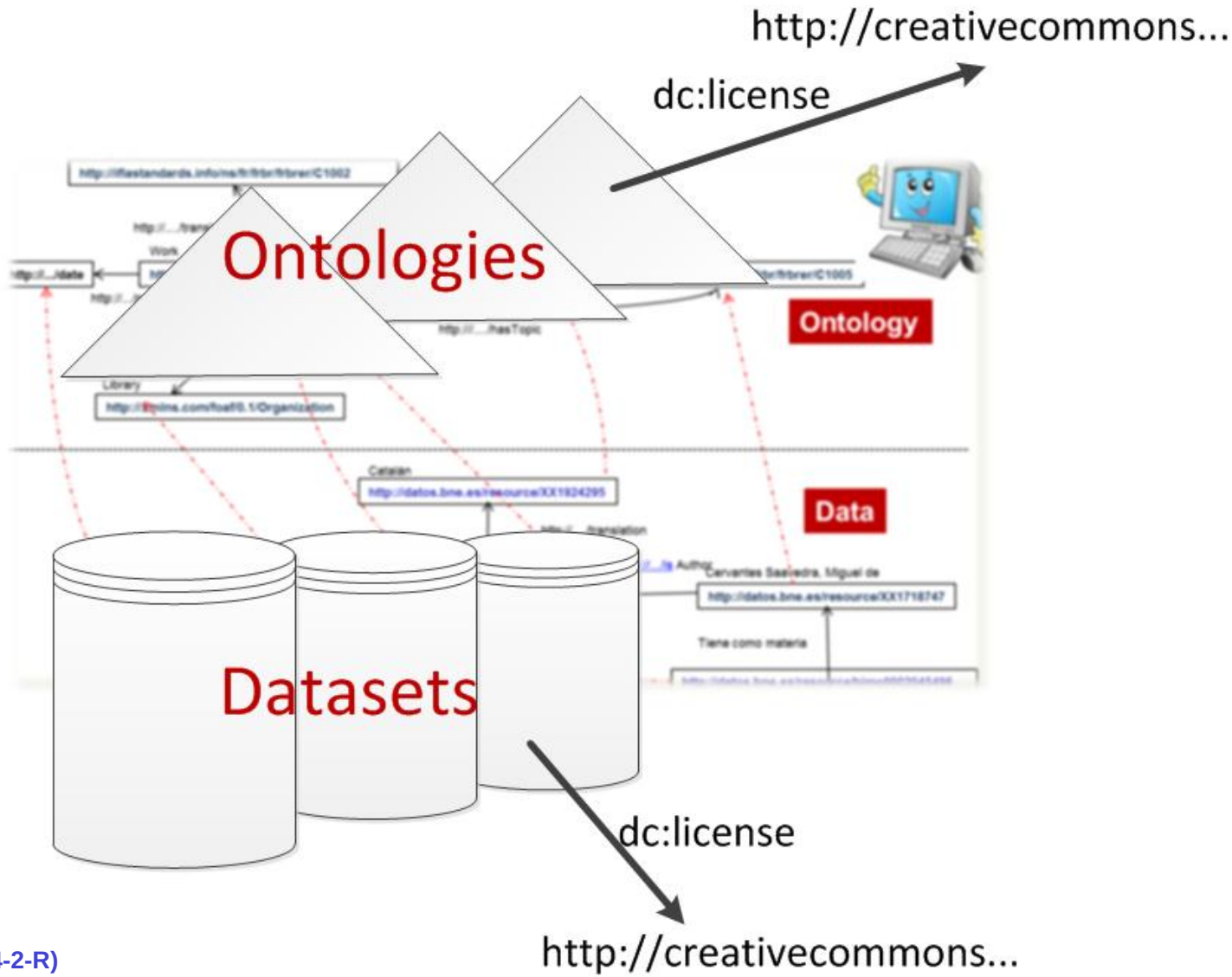


Rights expression in ODRL:
Who [can|cannot|must] act what in which resource how



4V (TIN2013-46238-C4-2-R)

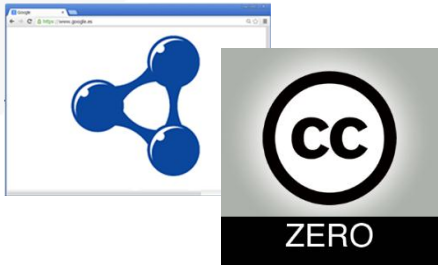
Linked Licensed Data



4V (TIN2013-46238-C4-2-R)

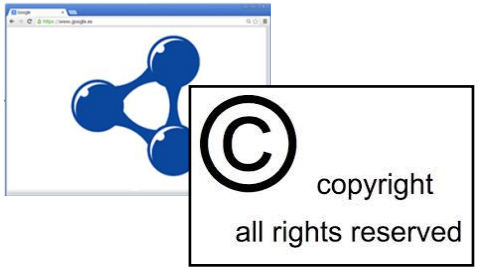
(Linked) Licensed Data in practice

(Linked) Open Data



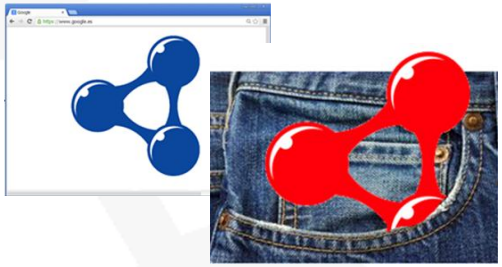
Published
Open License

(Linked) Closed Data



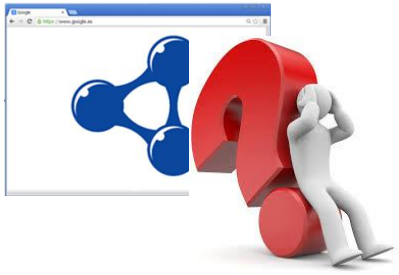
Published
No Open License

(Linked) Private Data



Not Published

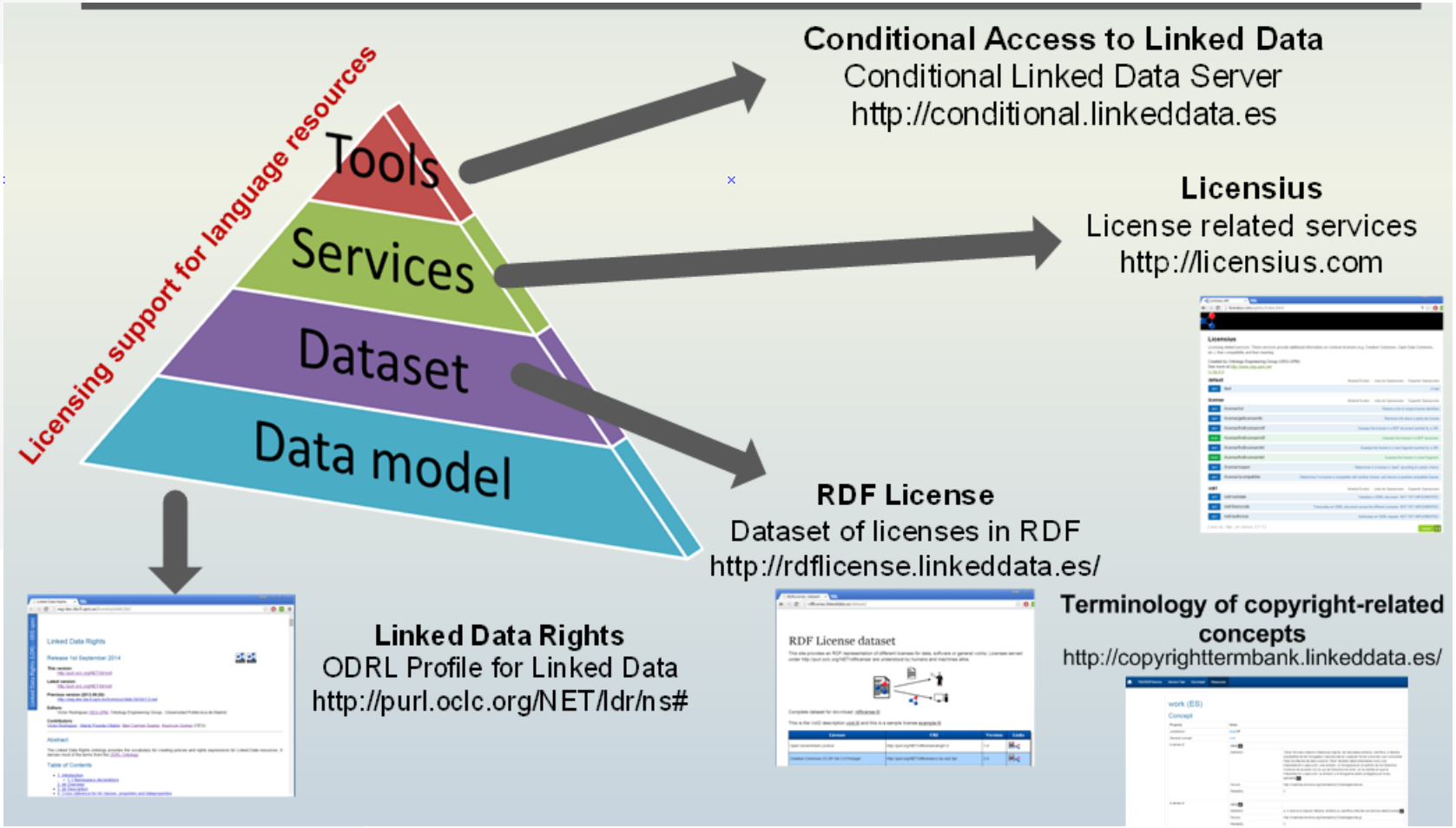
Available Data without explicit license



Published
Without License

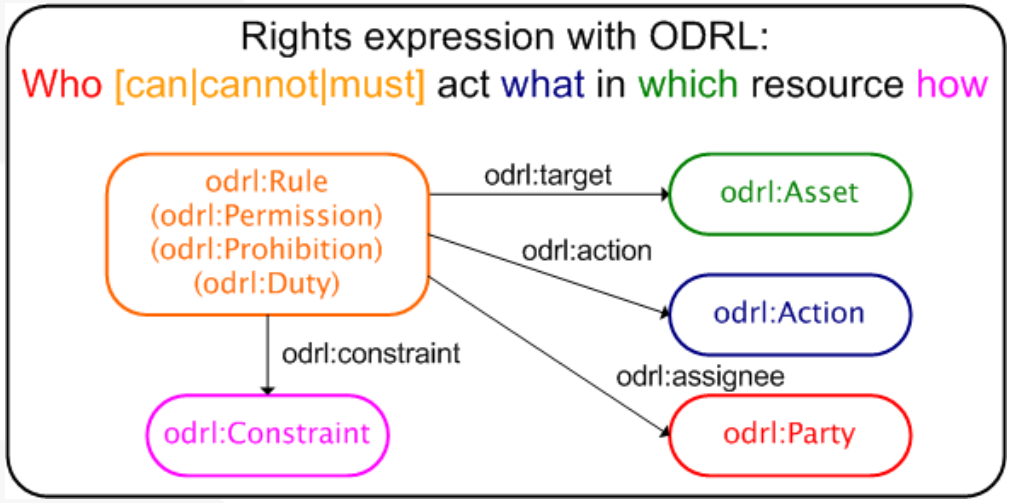
4V (TIN2013-46238-C4-2-R)

RDF Licensing support



4V (TIN2013-46238-C4-2-R)

Data Model: ODRL *(Open Digital Rights Language)*



«An action is (permitted /prohibited / obliged) to be acted by the party over the asset, provided that the constraints hold»

Asset:

Statement (rdf:Statement)

Dataset (void:Dataset)

Ontology (owl:Ontology)

Mapping (void:Linkset)

LDP Container

(ldp:Resource)

Action:

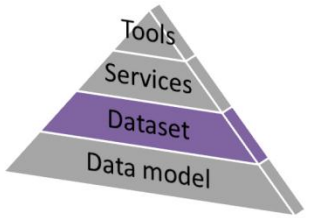
Derive (cc:DerivativeWorks)
Translate (odrl:translate)
Distribute (cc:Distribution)
Reproduce (cc:Reproduce)
Print (odrl:print)
Anonymize (odrl:anonymize)
Index (odrl:index)
… plus ~30 others in ODRL/CC…

Party:

One individual (ej: mailto:timbl@w3.org)
One organization: (http://www.oeg-upm.net)
One key owner using Web of Trust: (using http://xmlns.com/wot/0.1/hasKey)

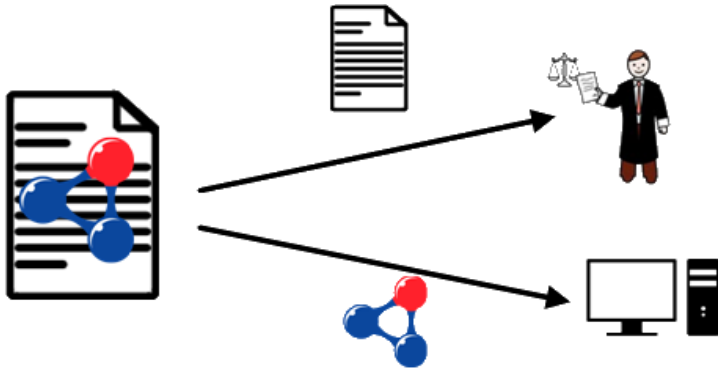
Constraint:

Acknowledgement (cc:Attribution)
A country, city… (odrl:spatial)
A time frame (odrl:timeInterval, odrl:dateTime…)



RDFLicense – Dataset of licenses in RDF

- Content negotiation: machine readable version of common licenses
- Based on ODRL (W3C spec)
- 148 licenses (as of June 2015): Creative Commons, ODC, GNU…
- **Permanent URIs.** Example: http://purl.org/NET/rdflicense/gpl2.0.ttl
- Browse them here: http://rdflicense.appspot.com/
- Contribute here: https://github.com/oeg-upm/rdflicense
- Catalogued here: http://datahub.io/dataset/rdflicense
- Read more here: A Dataset of RDF Licenses, V. Rodriguez-Doncel, S. Villata, A. Gomez-Perez, in Proc. of the 27th Int. Conf. on Legal Knowledge and Information System (JURIX), R. Hoekstra (Ed.), ISBN 978-1-61499-467-1, pp. 187-189, IOS Press, 2014



4V (TIN2013-46238-C4-2-R)

Sample license in ODRL: Creative Commons CC-BY-SA

A generic license (like Creative Commons') **has no party,** as the recipient is anybody accessing the licensed work

No Constraints (spatial, temporal, are not found in Creative Commons licenses), they are universal

```
<http://purl.org/NET/rdflicense/cc-by-sa3.0.ttl>
a odrl:Policy ;
   rdfs:label "Creative Commons CC-BY-SA" ;
   rdfs:seeAlso <http://creativecommons.org/licenses/by-sa/3.0/rdf> ;
   cc:legalcode <http://creativecommons.org/licenses/by-sa/3.0/legalcode> ;
   dct:hasVersion "3.0" ;
   dct:language <http://www.lexvo.org/page/iso639-3/eng> ;
   dct:publisher "Creative Commons" ;
   odrl:permission
   [
     odrl:action cc:Distribution , cc:DerivativeWorks , cc:Reproduction ;
     odrl:duty
     [
       odrl:action cc:Attribution , cc:Notice , cc:ShareAlike
     ]
   ] .
```

URI

How do I use in my data set?

:myDataset dct:license <http://purl.org/NET/rdflicense/cc-by-sa3.0>

Example of a license offering Paid Linked Data

The reproduction of a dataset is limited until the end of this year after paying 15€.

```
@prefix gr:    <http://purl.org/goodrelations/> .
@prefix dcat:  <http://www.w3.org/ns/dcat#> .
<http://samplepolicy/1234>
  a odrl:Offer ;
  rdfs:label          "License Offering Paid Linked Data" ;
  odrl:permission [
    odrl:target  <http://example.org/dataset/ds01> ;
    odrl:action  odrl:reproduce ;
    odrl:duty    [
      rdfs:label              "Pay" ;
      gr:UnitOfMeasurement    dcat:Dataset ;
      gr:amountOfThisGood     "1" ;
      odrl:action             odrl:pay ;
      odrl:target             "15,00 EUR"
    ] ;
    odrl:constraint
    [
      odrl:operator odrl:lt ;
      odrl:dateTime "2015-12-31"^^xsd:date ]
    ] ;
  ]
```

Example of a license for classes/concepts

- Old book offered for free, new book to be sold for 5 EUR

```
@prefix cc:    <http://creativecommons.org/ns#> .
@prefix odrl:  <http://www.w3.org/ns/odrl/2/> .

:example a odrl:Policy ;
  odrl:permission [
    odrl:action odrl:reproduce ;
    odrl:target :oldBook
  ] ;
  odrl:permission [
    odrl:action odrl:reproduce ;
    odrl:target :newBook ;
    odrl:duty [
      odrl:action odrl:pay ;
      odrl:target  "5,00 EUR"
    ]
] .
```

Using Policies to govern conditional access to Linked Data

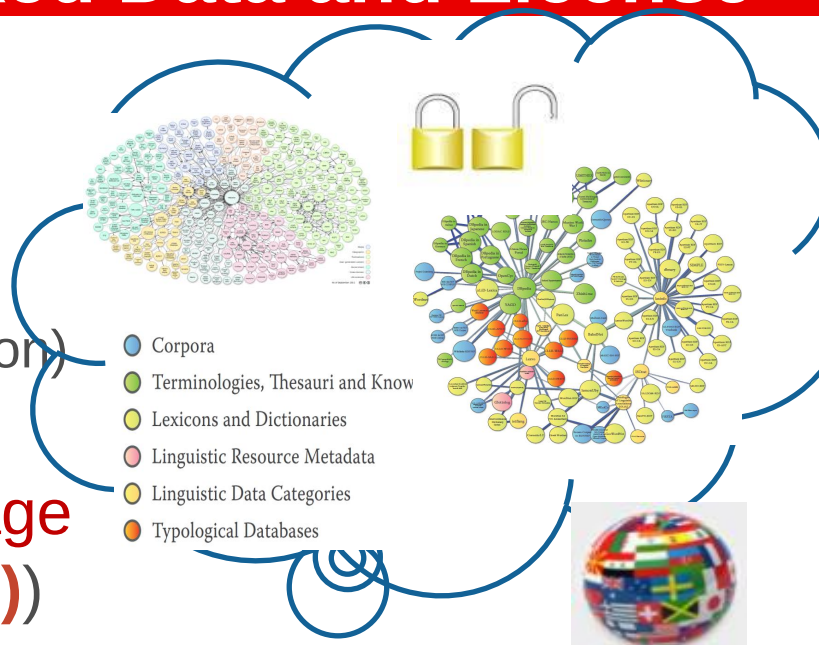
*Example of access to Linked Data for a price (**15EUR for the dataset** or 0.01EUR for a triple thereof)*

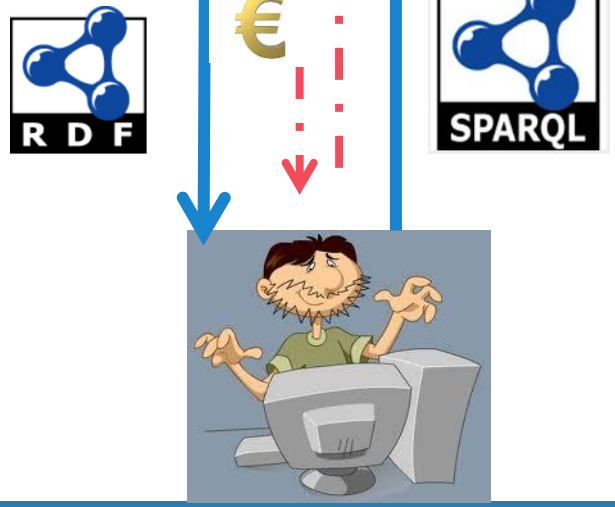
```
@prefix gr:    <http://purl.org/goodrelations/> .
@prefix dcat:  <http://www.w3.org/ns/dcat#> .
<http://salonica.dia.fi.upm.es/ldr/policy/cdaddba4-fc2e-4ee0-a784-e62f1db259bf>
        a                   odrl:Set ;
        rdfs:label          "License Offering Paid Linked Data" ;
        odrl:permission     [ a             odrl:Permission ;
                              odrl:target   <http://example.org/dataset/ds01> ;
                              odrl:action   odrl:reproduce ;
                              odrl:duty     [ a                       odrl:Duty ;
                                              rdfs:label              "Pay" ;
                                              gr:UnitOfMeasurement    dcat:Dataset ;
                                              gr:amountOfThisGood     "1" ;
                                              odrl:action             odrl:pay ;
                                              odrl:target             "15,00 EUR"
                                            ]
                            ] , [ a             odrl:Permission ;
                              odrl:action   odrl:reproduce ;
                              odrl:target   <http://example.org/dataset/ds01> ;
                              odrl:duty     [ a                       odrl:Duty ;
                                              rdfs:label              "Pay" ;
                                              gr:UnitOfMeasurement    rdf:Statement ;
                                              gr:amountOfThisGood     "1" ;
                                              odrl:action             odrl:pay ;
                                              odrl:target             "0,01 EUR”

                                            ]
                            ] ..
```

4V (TIN2013-46238-C4-2-R)

Linked Data, Linguistic Linked Data and License

1. **Agree on vocabularies** for describing
   - Domain data
   - Language related information (Lemon)
   - LR license metadata (ODRL)
2. Unified and standardized language for describing resources (**RDF(S)**)
3. Unified and standardized **query language** (SPARQL)
4. Standardized non-proprietary APIs
5. Links to other resources





Linked Data Process

Examples from datos.bne.es

datos.bne.es and MARiMba: an insight into library linked data D. Vila-Suero, A. Gómez-Pérez
Library Hi-tech (ISSN: 0737-8831); Emerald Group Publishing Limited . **Vol.**: 31(4).**Pages:** 575-601 Noviembre 2013

datos.bne.es: the team

Cataloguing services

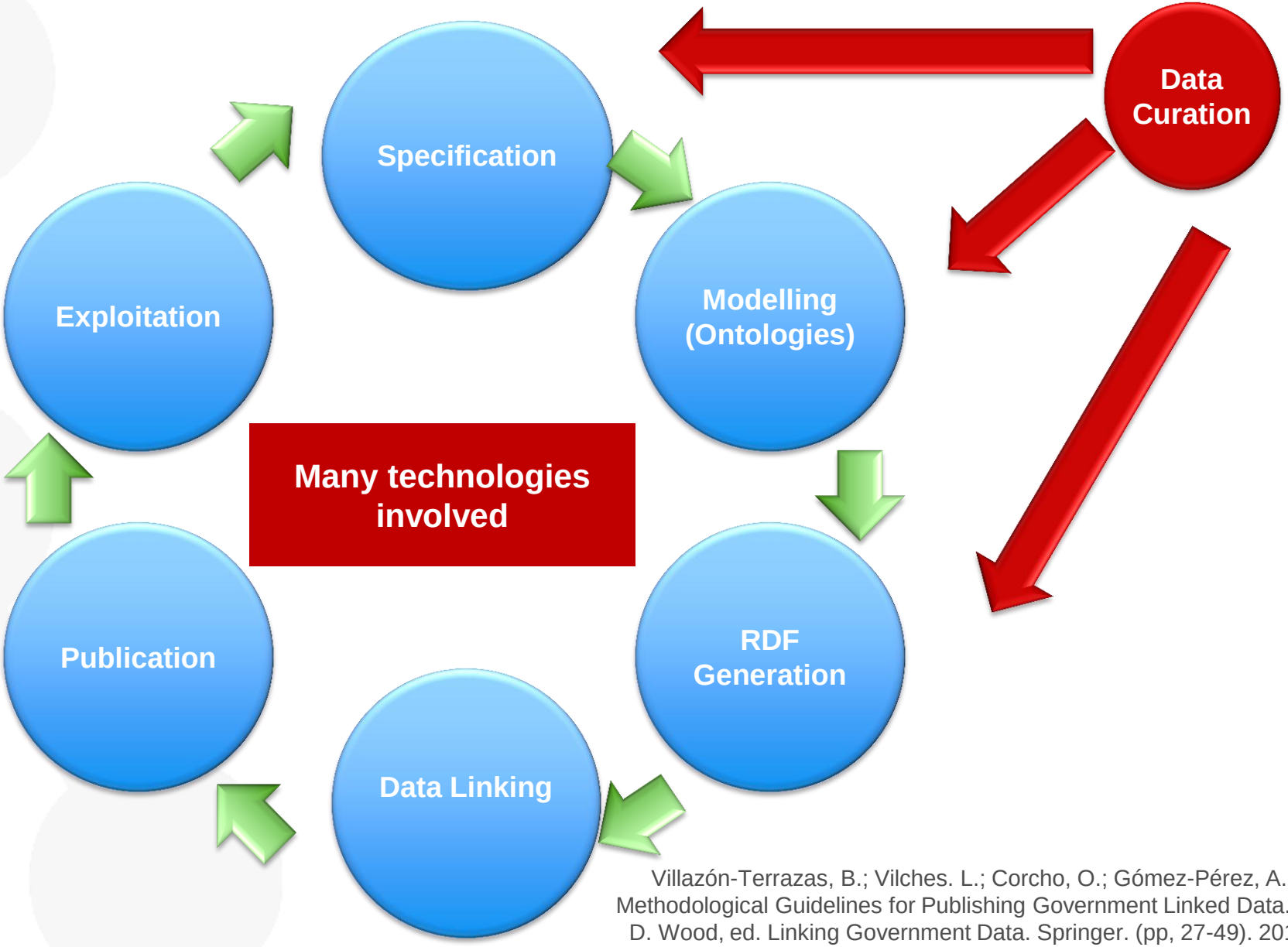
Digital Library and automatization

Focus group

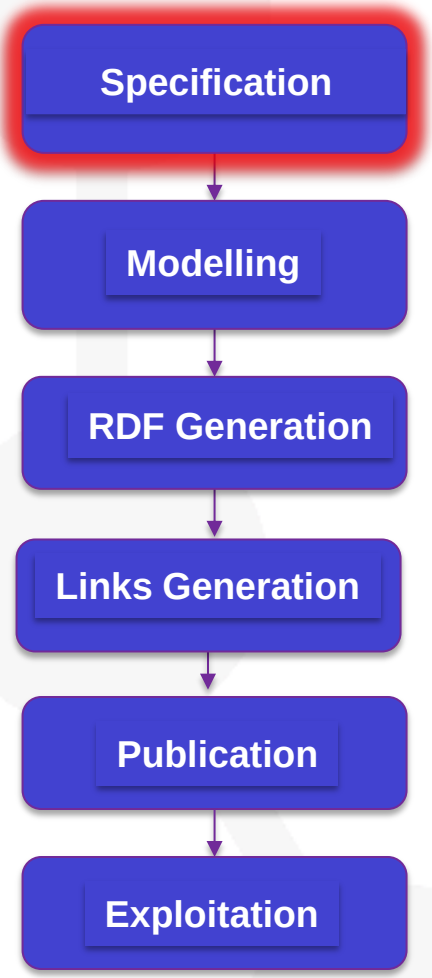
R&D in Linked Data

Ontology engineering

Methodology



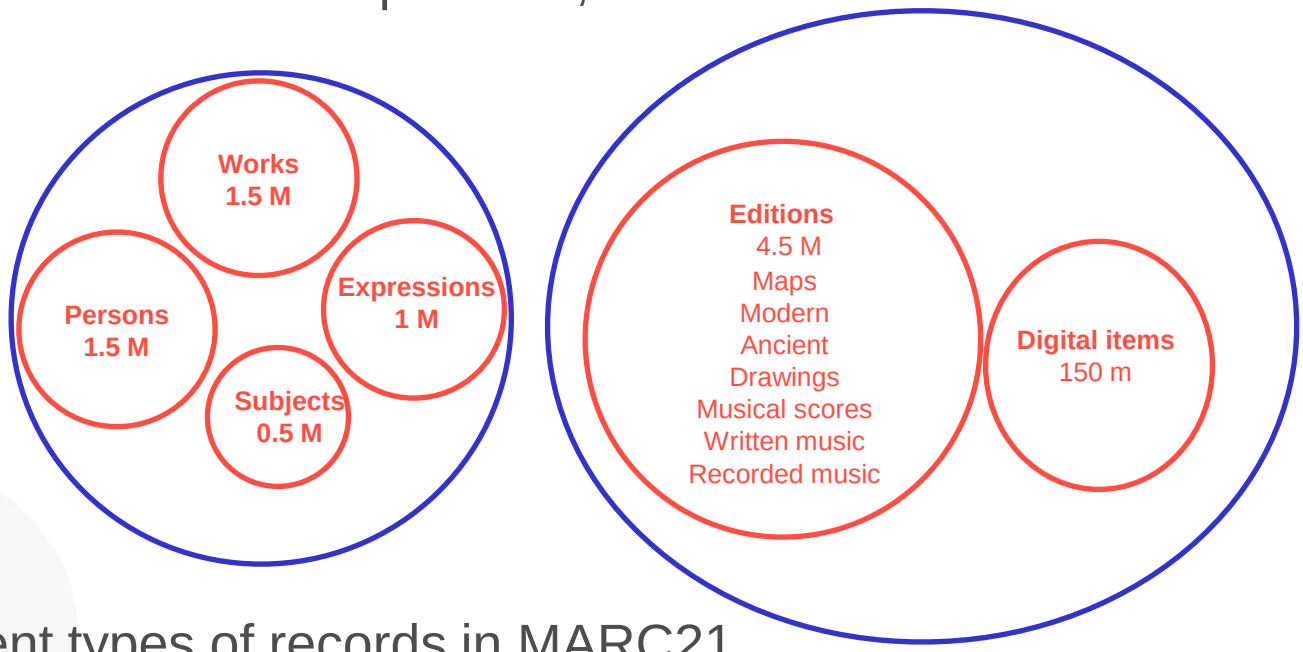
Villazón-Terrazas, B.; Vilches. L.; Corcho, O.; Gómez-Pérez, A. Methodological Guidelines for Publishing Government Linked Data. In D. Wood, ed. Linking Government Data. Springer. (pp, 27-49). 2011



Goal

Linked Data generation of the Spanish National Library Metadata

- Records in the **MARC 21 format**
- **4.5 million** bibliographical records
- **4.5 million** authority records
- Version: September, 2015

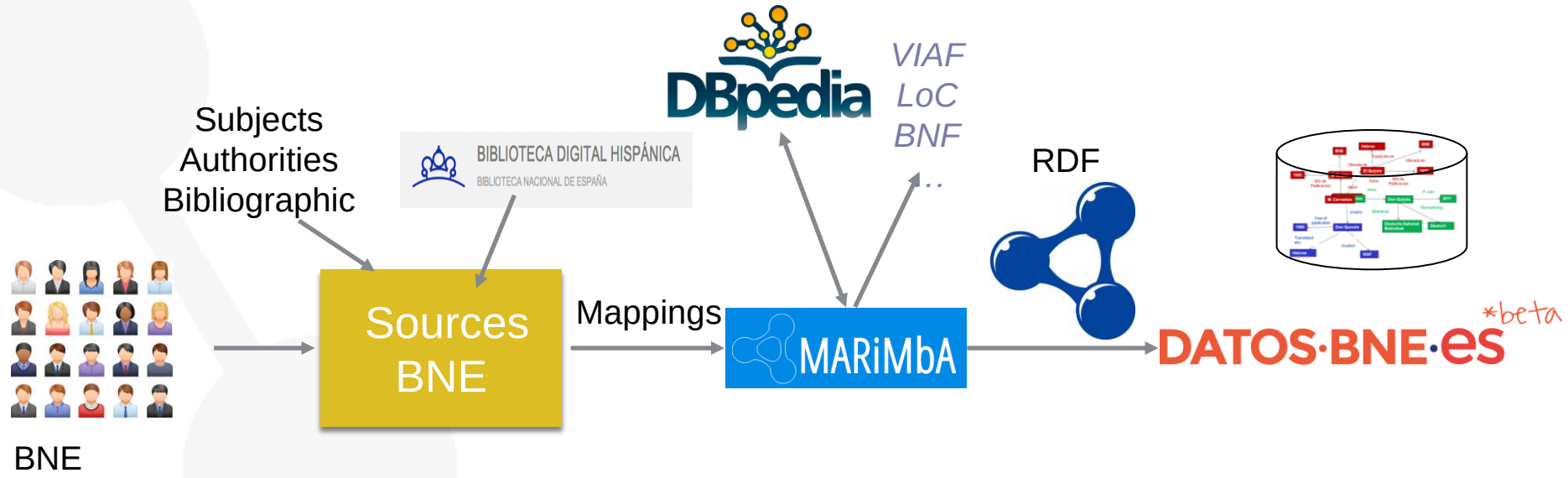


Different types of records in MARC21

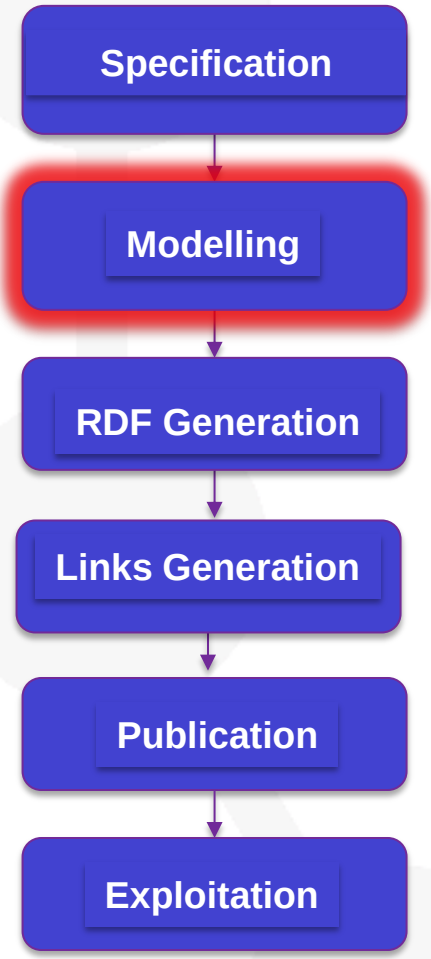
Specification

Domain experts from BNE (catalogers) part of the mapping process.

Multilinguality, collaboration with IFLA



Modeling in datos.bne.es



Two phases:

1. **Re-using URIs:** from standard vocabs, mainly IFLA FRBR, ISBD and FRAD
2. **Integrating ontology: BNE ontology** that re-uses and links to external vocabs.

Motivation for the creation of BNE ontology:

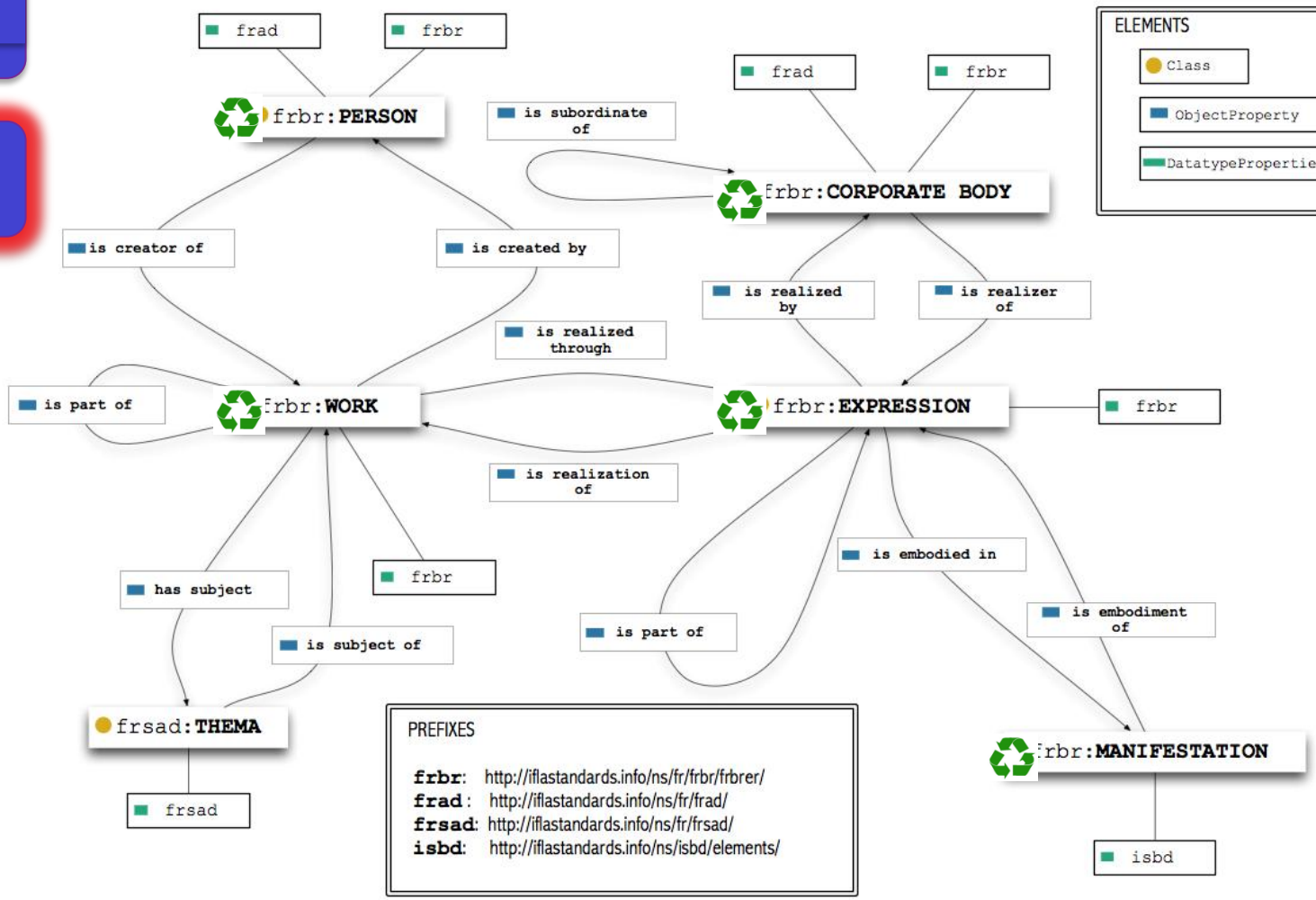
- To provide stability under the datos.bne.es domain
- Documented in a central document
- Provide new properties and relationships (e.g., spanish legal deposit)
- Control over descriptions and labels (e.g., labels for visualization (bne:label))
- Multilingualism: BNE ontology is described in English and Spanish
- Content-negotiation enables using programming APIs (Jena, Sesame, etc.) and Ontology editors (Protégé, TopBraid, etc.)

Available and documented at http://datos.bne.es/def/

IFLA Vocabulary-based Ontology - Reusing URIs

Specification

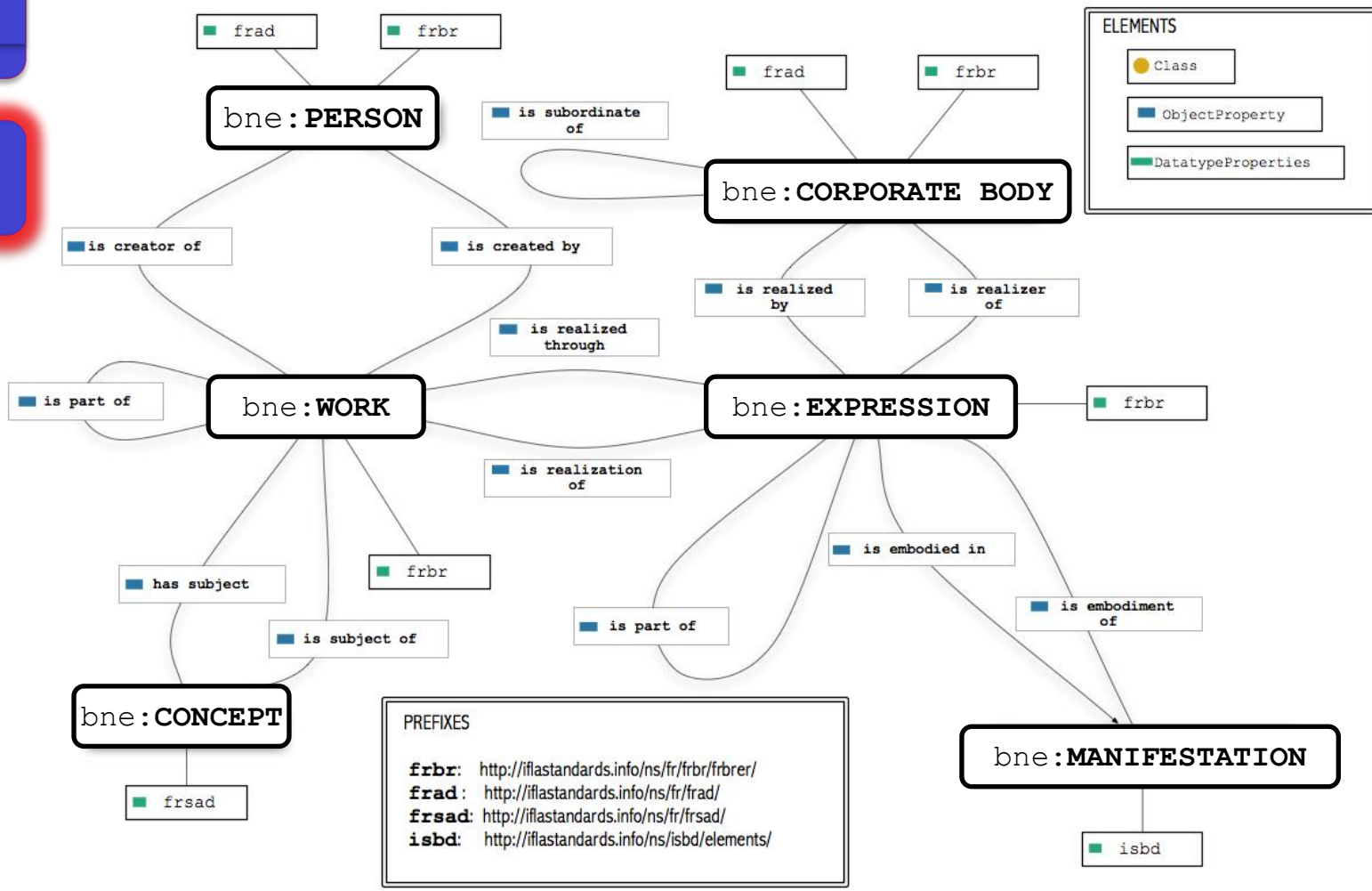
Modelling



- Available and documented at http://datos.bne.es/def/

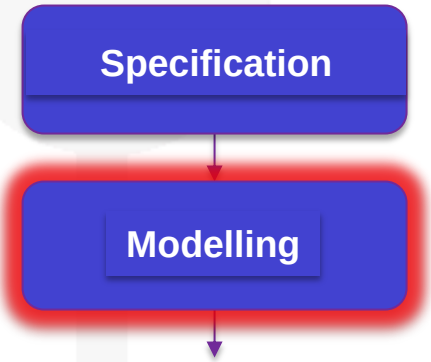
Specification

Modelling



Exploiting alternative names and titles

- Indexing alternatives (acronyms, translations,etc.) increases recall of search results



Person → **Authority Name** *Cervantes Saavedra* → **Variant names** *Servantes*, Сервантес Сааведра

Work → **Authority title** *Rome and Juliet* → **Variant titles** *The tragedy of Romeo and Juliet*, *Romeo y Julieta*, *Romeu i Julieta*

Org → **Authority Name** *Naciones Unidas* → **Variant names** *United Nations*, ONU (Spanish acronym), UN (English acronym)

Who will be the mapping generator?

Specification

Modelling

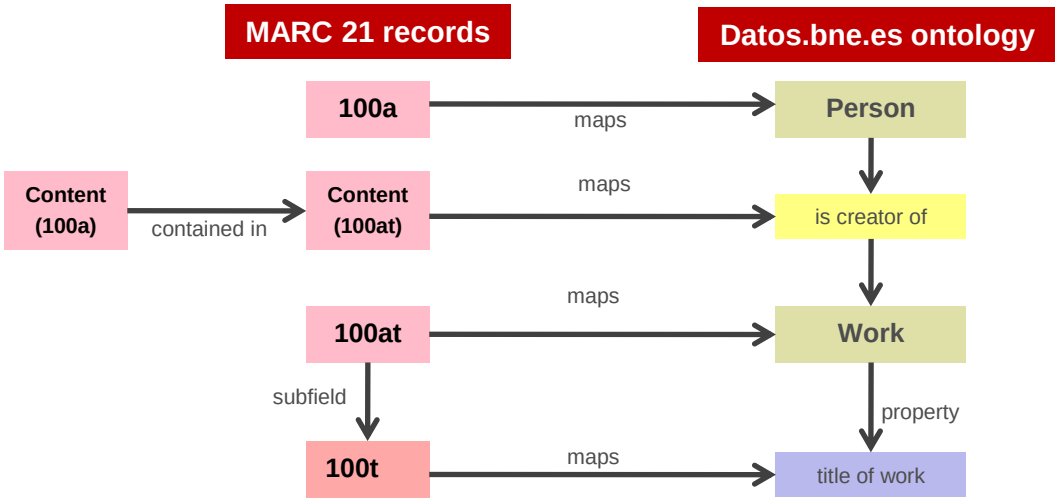
RDF Generation

Links Generation

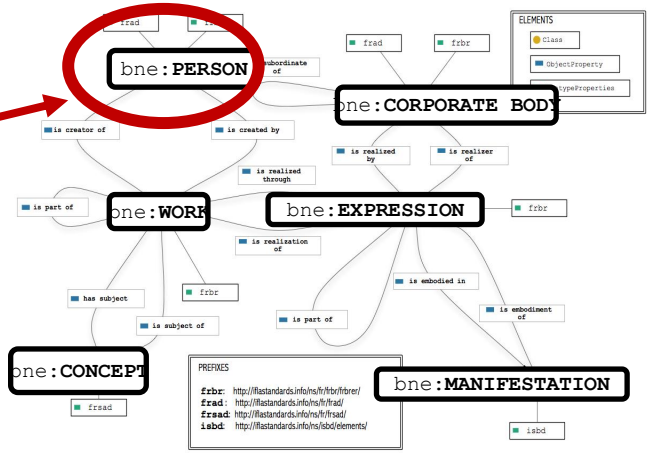
Publication

Exploitation

- Librarians built mappings



001 XX1721208
005 200012181124
008 901120nn aijnnaabn n aaa
016 $a BNE19900178994
040 $a SpMaBN $b spa $c SpMaBN $e rdc $f embne
100 10 $a Camus, Albert
$d 1913-1960
670 $a El mite de Sísif, 1987 $b port. (Albert Camus)
670 $a Dic. de filosofía, de J. Ferrater Mora, 1980$b(Camus., Albert (1913-1960); n. Mondovi, Argel)
670 $a Aut. BN-OPALE, 1995 $b (Camus, Albert)

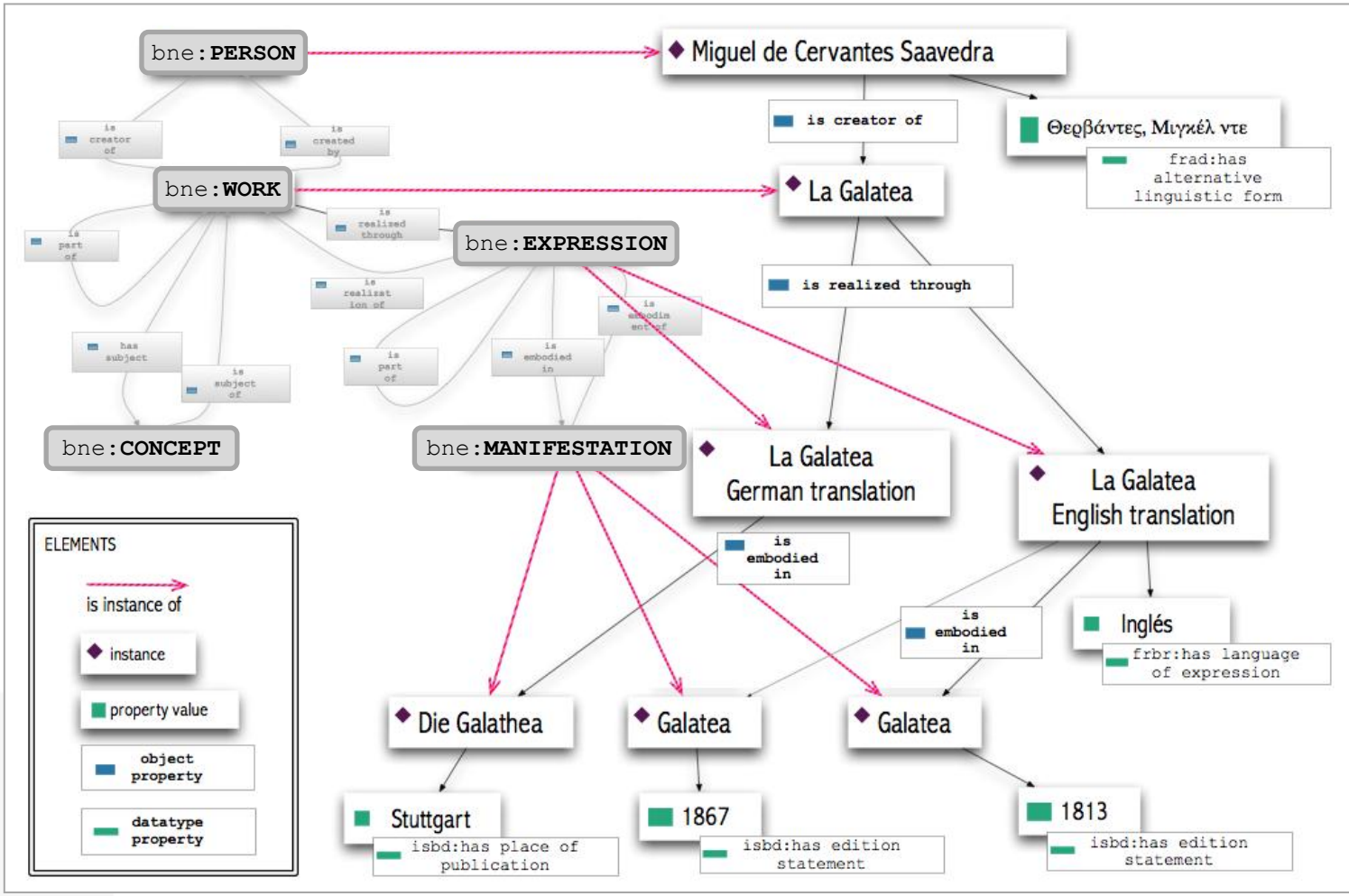


MARiMbA generates RDF using RDFS/OWL ontologies

Specification

Modelling

RDF Generation



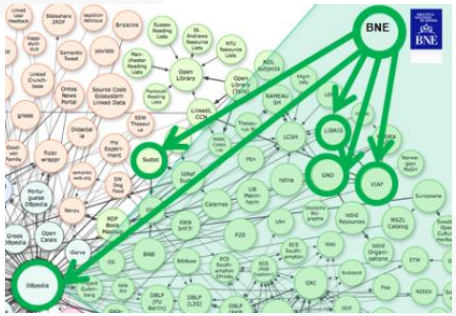
Marimba links with other resources: VIAF, DNB, SUDOC, LIBRIS, DBpedia

Specification

Modelling

RDF Generation

Links Generation



MARiMbA

http://d-nb.info/gnd/11851993X

DNB

http://viaf.org/viaf/17220427

VIAF

Same As

Same As

http://dbpedia.org/resource/Miguel_de_Cervantes

DBpedia

Same As

http://datos.bne.es/resource/XX1718747

BNE

Same As

Same As

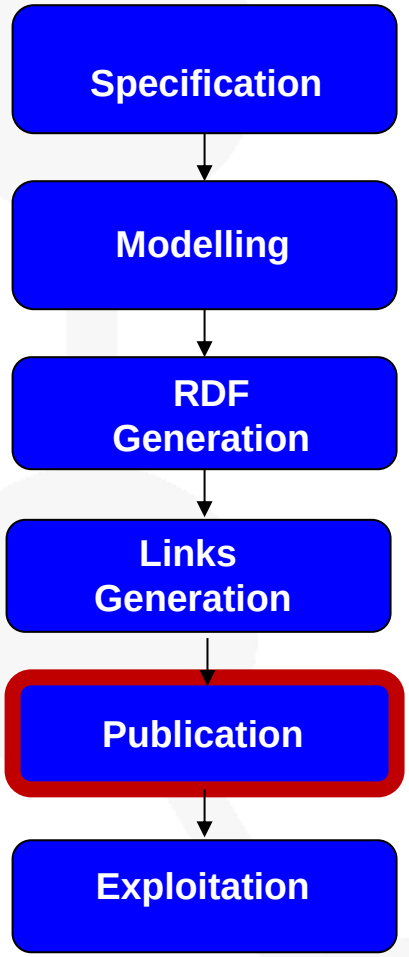
http://www.idref.fr/026774771/id

SUDOC

http://libris.kb.se/resource/auth/45369

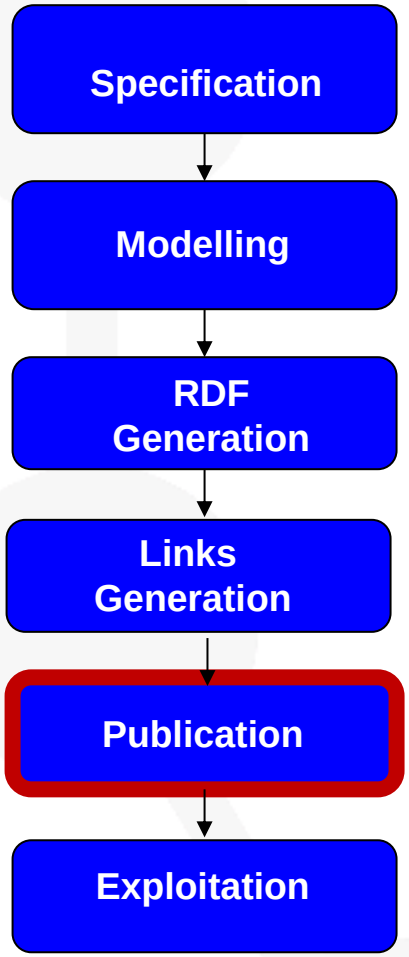
LIBRIS

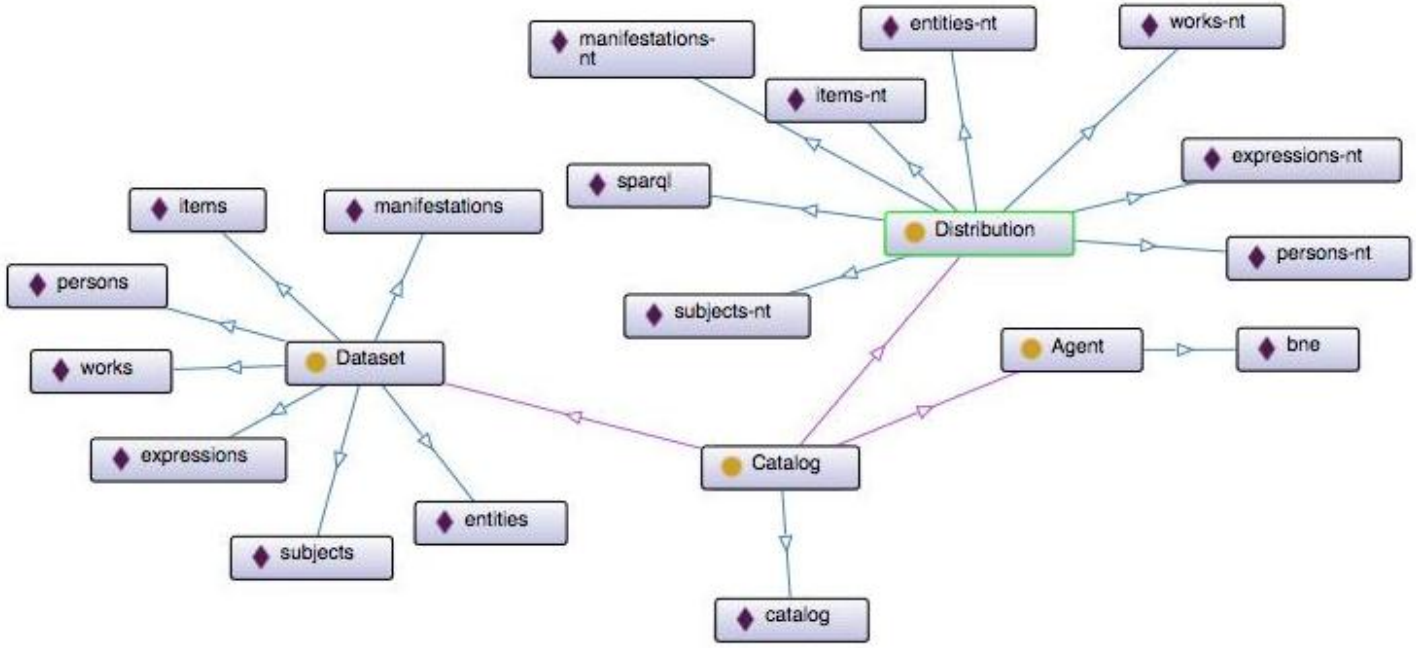
Publication



- **Combines DCAT and VoID, to include:**
  - **Licenses:** Pointing to rdflicense dataset
  - **Provenance:** Using PROV-O
  - **Language information:** Using lexvo.org
  - **Access mechanisms:** SPARQL endpoint, data dumps
- datos.bne.es is a **Catalog**
  - composed of 7 (interlinked) **Datasets**
  - with different **Distributions**
    - **Data dumps**
    - **SPARQL endpoint**
    - **Search API**
- Available at http://datos.bne.es/inicio.ttl

Full DCAT Graph





Datos.bne.es: DCAT Catalog

```
:catalog
    a dcat:Catalog ;
    dct:title "The RDF catalog of the National Library of Spain"@en ,
              "El catálogo RDF de la Biblioteca Nacional de España"@es ;
    rdfs:label "datos.bne.es RDF catalog"@en ,
               "Catálogo RDF datos.bne.es RDF"@es ;
    foaf:homepage <http://datos.bne.es/inicio> ;
    dct:publisher :bne ;
    dct:license rdflicense:cc-zero1.0 ;
    dct:language <http://lexvo.org/id/iso639-3/spa>
                 ,<http://lexvo.org/id/iso639-3/eus>
                 ,<http://lexvo.org/id/iso639-3/por>
                 ,<http://lexvo.org/id/iso639-3/fra>
                 ,<http://lexvo.org/id/iso639-3/lat>
                 ,<http://lexvo.org/id/iso639-3/ita>
                 ,<http://lexvo.org/id/iso639-3/cat>
                 # Up to 195 languages
```

RDFLicense dataset

Publication

Datos.bne.es: The publisher as agent

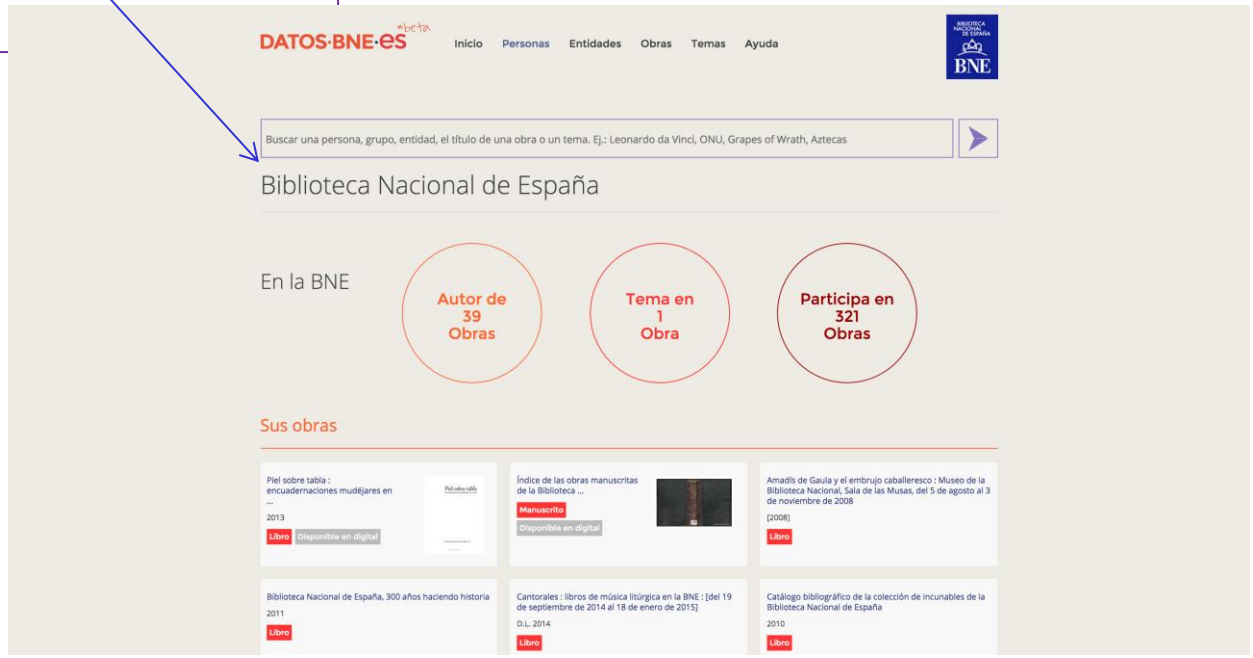
```
:catalog
    dct:publisher :bne ;
```

BNE as publisher of the catalog

```
:bne
    a foaf:Agent ;
    owl:sameAs datosbne:XX4891886 .
```

BNE as agent
Linked to the RDF resource in datosbne

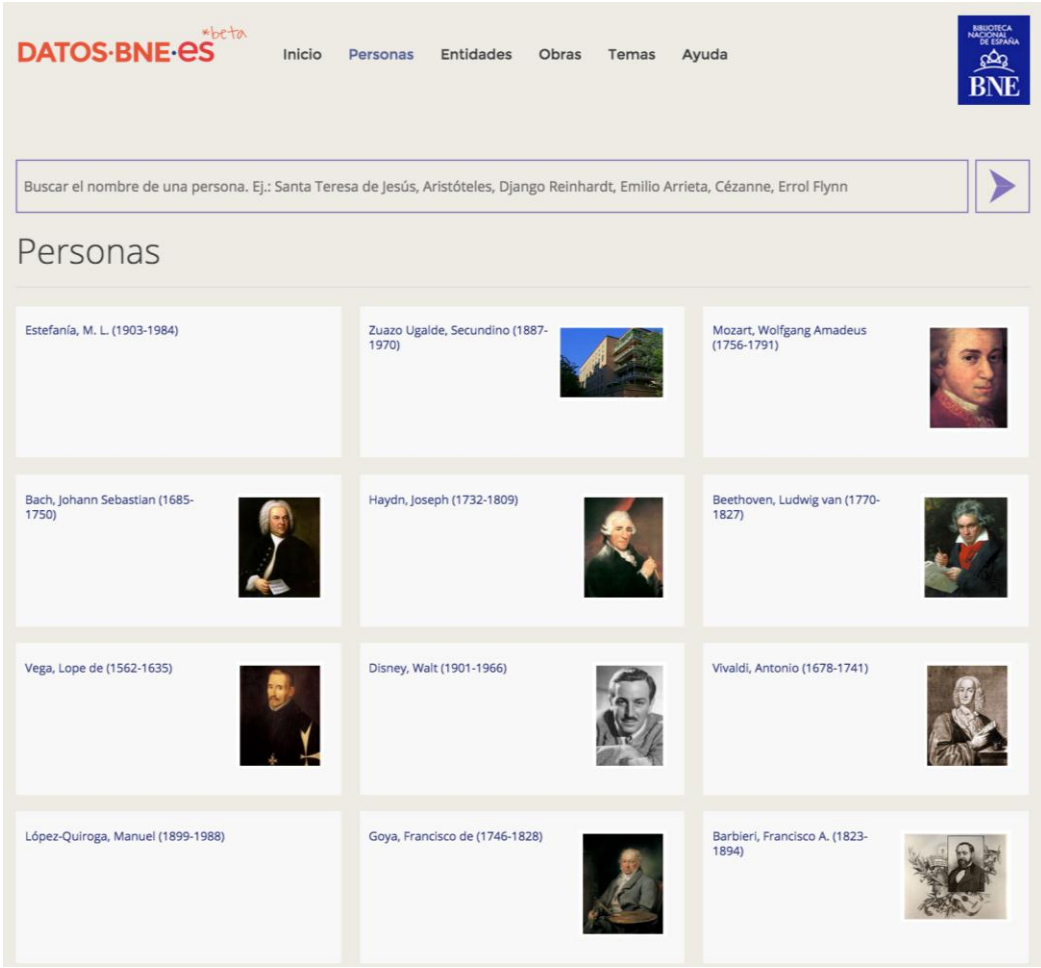
Publication



Datos.bne.es: The catalog contains 7 datasets

```
:catalog
   dcat:dataset :entities,
                :works,
                :expressions,
                :manifestations,
                :items,
                :subjects,
                 :persons;
```

One dataset per type of entity



Publication

Datos.bne.es: Persons dataset

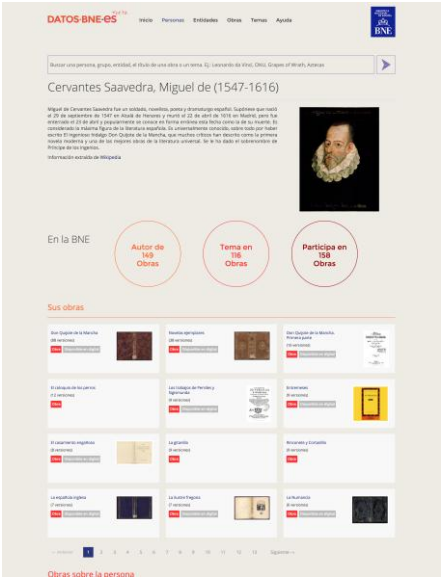
```
:persons
    a dcat:Dataset ;
    dct:title "Persons in datos.bne.es"@en,
              "Personas en datos.bne.es"@es  ;
    dcat:keyword "persons"@en, "personas"@es,
                 "authors"@en, "autores"@es ;
    dct:issued "2011-12-05"^^xsd:date ;
    dct:modified "2011-12-05"^^xsd:date ;
    dcat:contactPoint
    <mailto:info.datosenlazados@bne.es?subject='Persons%20dataset'> ;
    dcat:distribution :persons-nt, :sparql ;
    .
:persons-nt
    a dcat:Distribution ;
    dcat:downloadURL <http://datos.bne.es/datadumps/persons101115.nt.bz2>;
    dct:title
   "Distribution of persons description in RDF (N-triples)"@en ,
   "Distribución de descripciones de personas en RDF (N-triples)"@es ;
    dcat:mediaType "application/x-bzip2" ;
    dct:license rdflicense:cc-zero1.0 ;
```

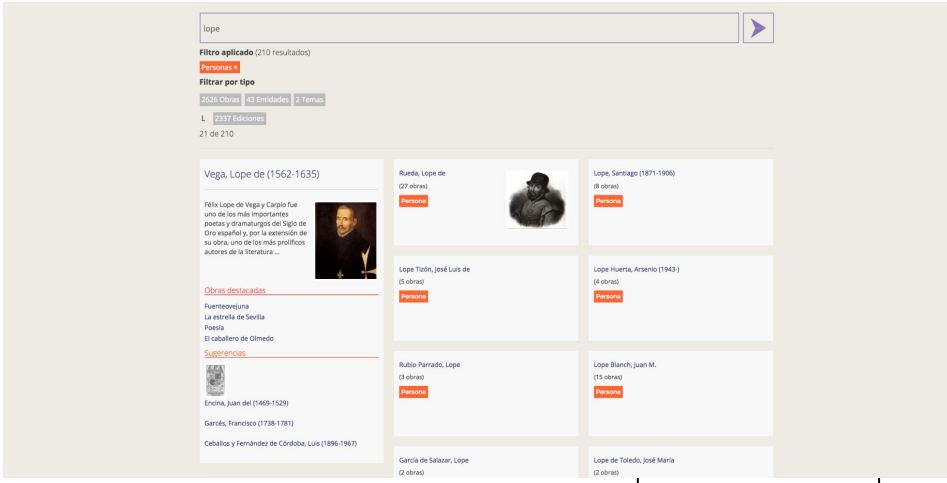
 # Persons dataset is distributed in two ways: SPARQL and N-triples dump

 # Each distribution has its own license

- Specification
- Modelling
- RDF Generation
- Links Generation
- Publication
- Exploitation

Web portal and service





SPARQL queries

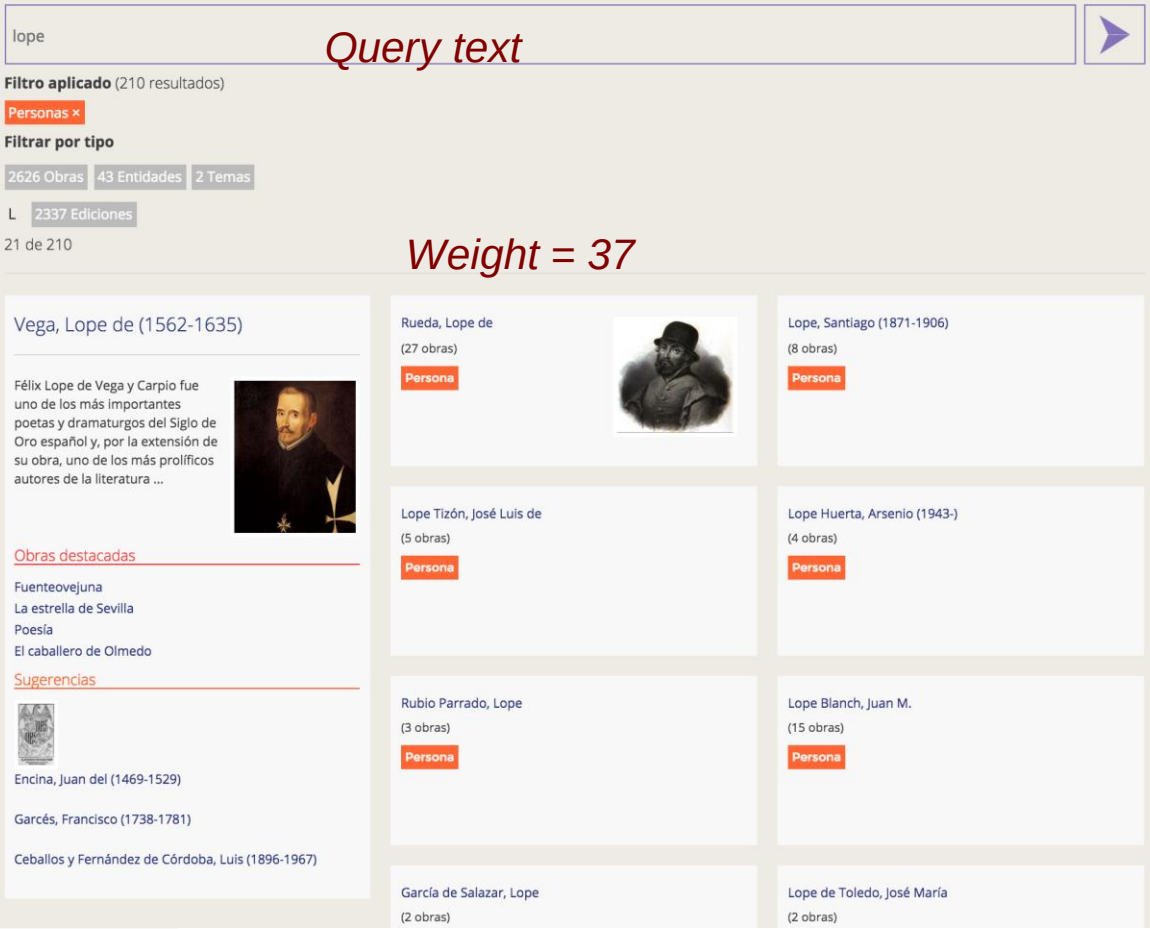
URI Cervantes

```
select distinct COUNT(?Obras) where {

http://datos.bne.es/resource/XX1718747

<http://datos.bne.es/def/OP501>
?Obras
}
```

Is creator

Exploitation: Ranking and recommedantion in datos.bne.es

Weighting function based on: the ontology, incoming and outgoing links, string similarity



Weight = 613

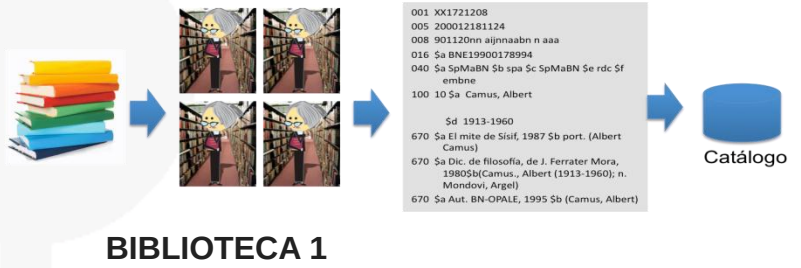
Ranked recommendations: Works and similar persons

Results: datos.bne.es

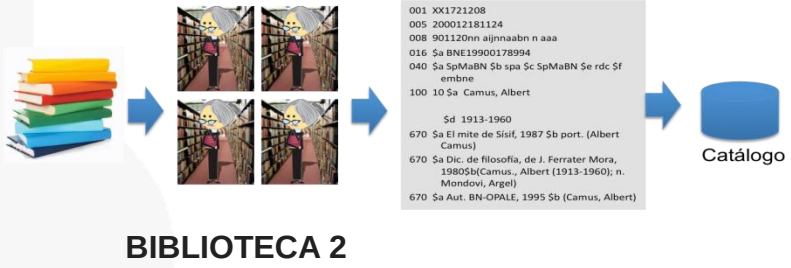
- Total number of authority records: **4.784.303**
- Total number of bibliographic records: **4.083.671**
- Total number of RDF triples: **143.153.218**
- Total number of owl:sameAs links: **1.395.108**
- Linked sources:
  - VIAF
  - SUDOC (French collective university catalogue
  - GND (German National Library of authorities)
  - LIBRIS Sweden
  - DBPedia
  - BNF, France
  - geo.linkeddata.es

Uses of Library Linked Metadata

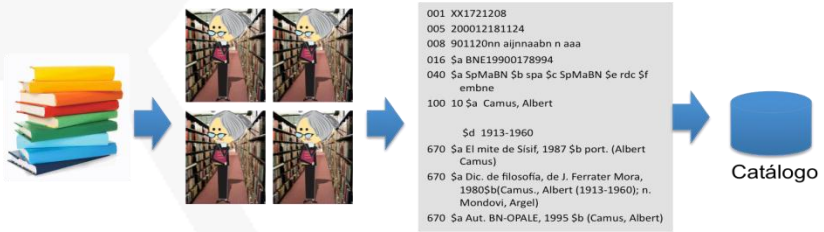
Validación y enriquecimiento de registros MARC21



BIBLIOTECA 1



BIBLIOTECA 2



BIBLIOTECA 3

...

Catálogos mejorados
Reducción de costes

The long tale: datos.bne.es used in national press articles

¡MADRID! veranos de la villa madrid 2015

MÚSICA / EDICIÓN ILIMITADA

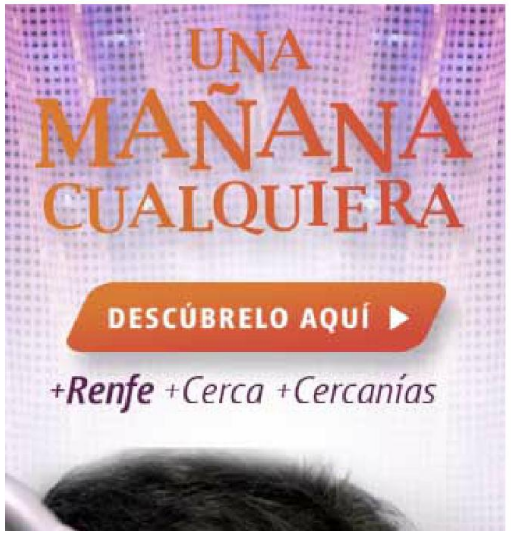
El chotis de un alcalde a 45 revoluciones

JESÚS LILLO / MADRID | Día 11/05/2015 - 14.23h

- Antes que Aguirre y Carmona, José Luis Álvarez hizo campaña en 1979 con un chotis, editado en vinilo

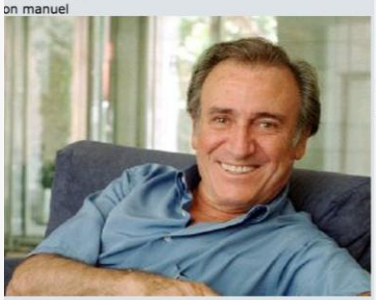
o COMENTARIOS
IMPRIMIR
COMPARTIR 7 85 0

Publicidad



Long Tale: Specialize forums for rare editions

osemanuel
on manuel



egistrado: 12 Ago 2009
ensajes: 917
bicación: Ciudad Real

olver arriba

Publicado: Vie Oct 03, 2014 7:43 pm **Asunto**:

http://www.amazon.es/Conversaciones-hombre-bueno-Fuera-Colecci%C3%B3n/dp/8427041330/ref=sr_1_3?ie=UTF8&qid=1412361736&sr=8-3&keywords=manolo+escobar

Acce4so directo a la tienda online de amazon

_____

Soy escobarista por encima de todo.

profile pm

upermanolo
on manuel

egistrado: 07 Nov 2007
ensajes: 252
bicación: guadalajara

olver arriba

Publicado: Mar Mar 10, 2015 6:29 pm **Asunto**:

Curioseando por ahi e descubierto otro libro sobre Manolo Escobar esta escrito por Juan Gabriel Garcia Escobar en 1976 y salio en 6 fasciculos de 56 paginas.

http://datos.bne.es/edicion/bimo0001848380.html

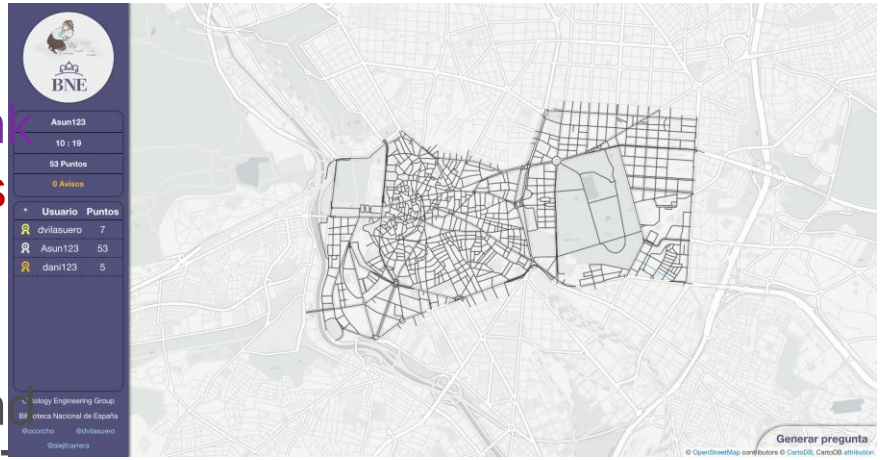
Es la pagina de la biblioteca nacional y se puede obtener muchos datos de otros artistas y escritores.

Saludos

profile pm

Street detective: citizen science

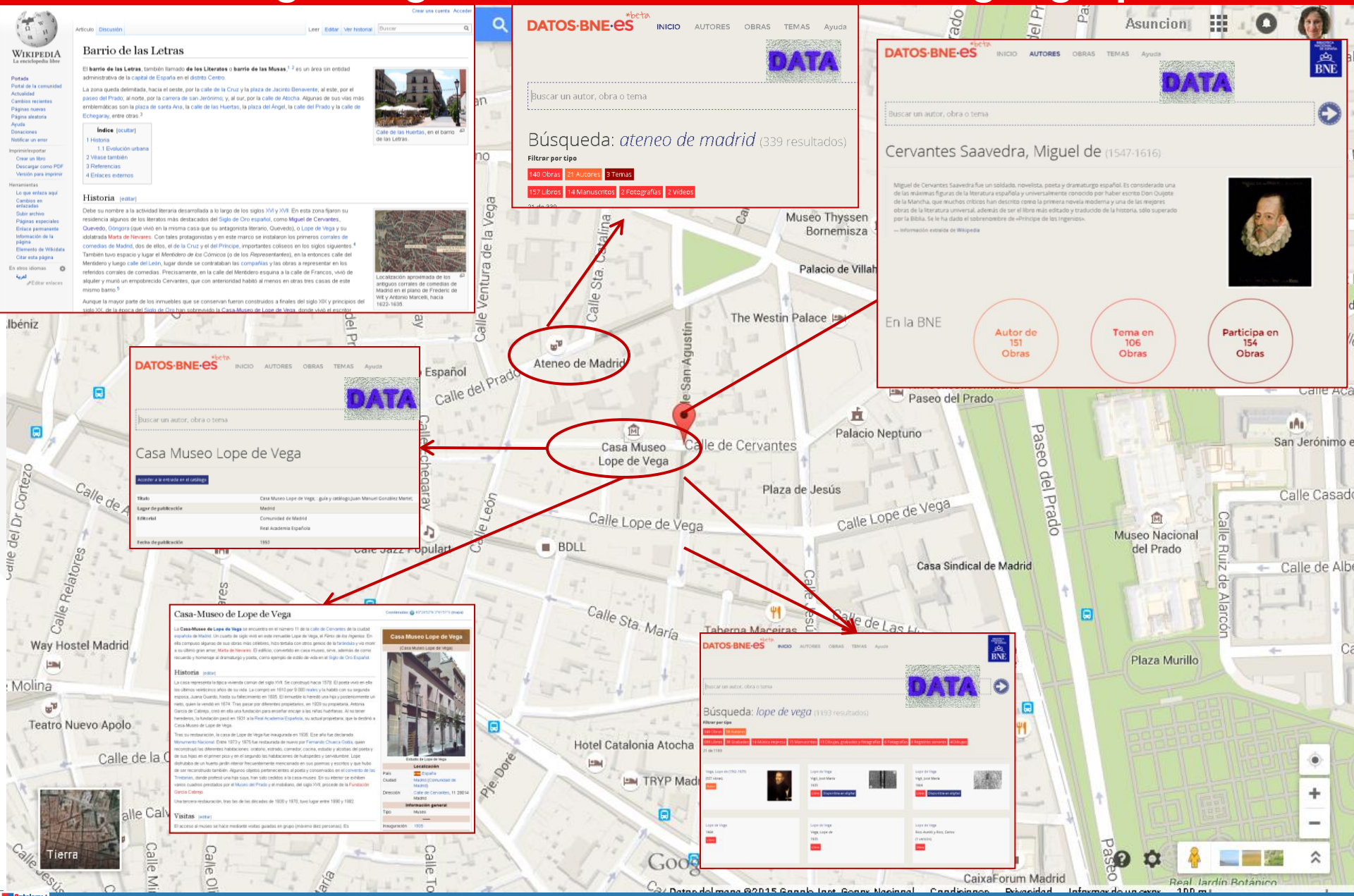
- To improve data quality and access to cultural data in a particular city
- Gaming exercise: Ask questions that link que asocien personajes históricos a las calles de su ciudad.
- Link them with datos.bne.es
- Performed by childrens (between 10 and 15 years old) in **Zaragoza** and **Madrid in collaborartion with the** city hall of Zaragoza and BNE



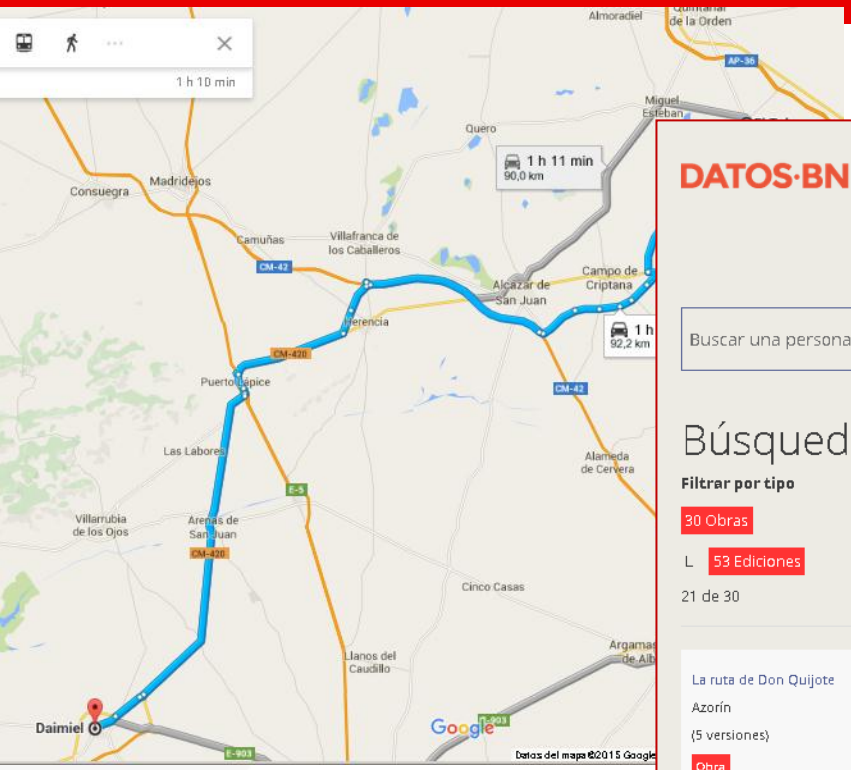
FUJITSU FUJITSU LABORATORIES

"Fujitsu Laboratories of Europe Innovation Award 2015"

Linking: Integration of cultural data and geographical data



“El Quijote” route



DATOS·BNE·es *beta

INICIO PERSONAS ENTIDADES OBRAS TEMAS Ayuda

BIBLIOTECA NACIONAL DE ESPAÑA BNE

Buscar una persona, grupo, entidad, el título de una obra o un tema. **Ej.:** Leonardo da Vinci, ONU, Grapes of Wrath, Aztecas

Búsqueda: *ruta de Don Quijote* (30 resultados)

Filtrar por tipo

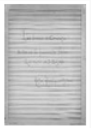
30 Obras

L 53 Ediciones

21 de 30

La ruta de Don Quijote
Azorín
(5 versiones)
Obra

La ruta de Don Quijote. Las bodas de Camacho
Rodríguez Albert, Rafael
(1 versión)
Obra Disponible en digital



Ruta de Don Quijote de la Mancha
Rodríguez, Leandro
(1 versión)
Obra

Ruta de Don Quijote
González García, Pilar
(1 versión)
Obra

Ruta de Don Quijote en BTT
Juaristi Sesma, Luis
(1 versión)
Obra

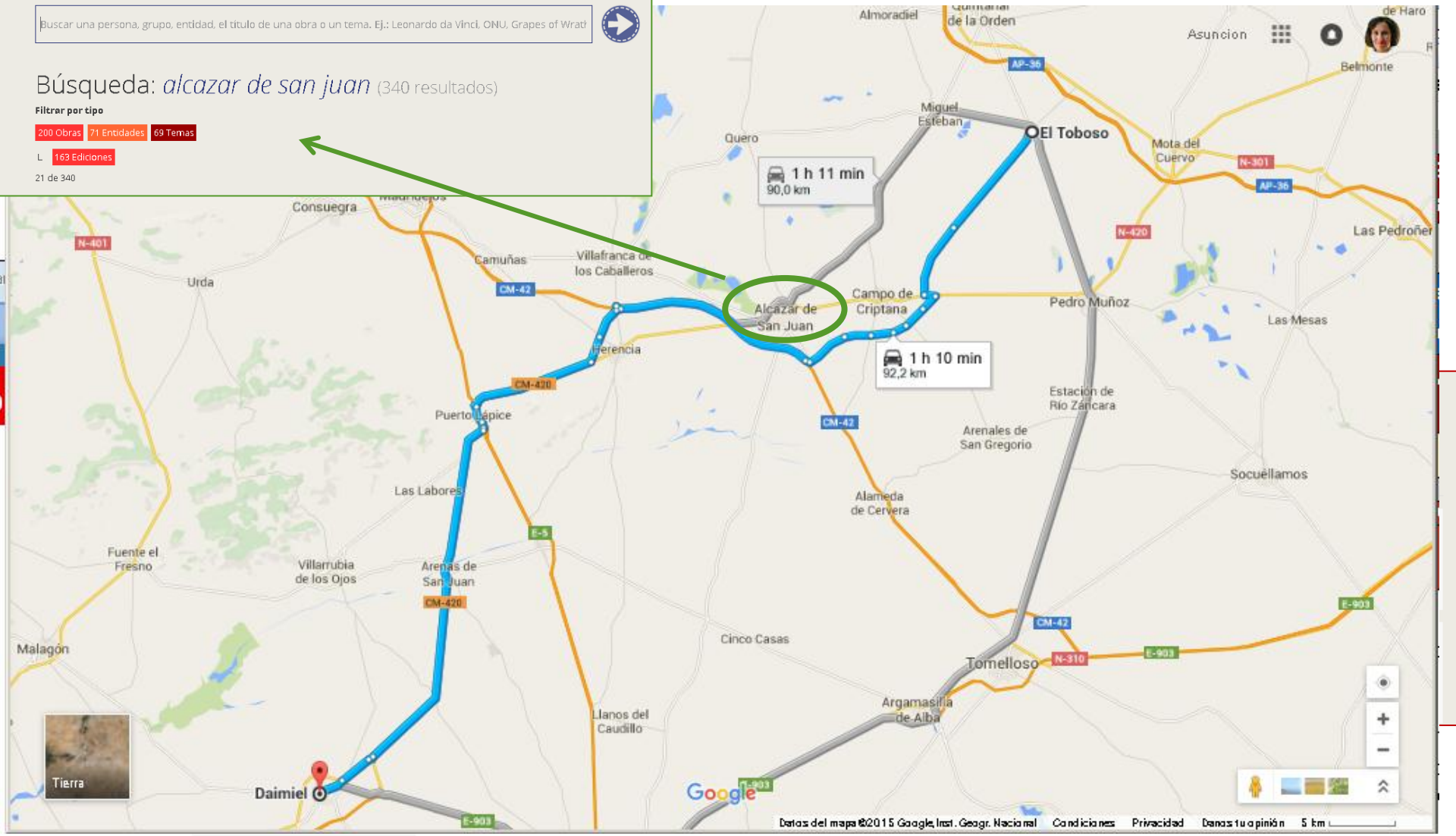
Ruta de Don Quijote
Unceta, María
(1 versión)
Obra

Cuenca en la ruta de Don Quijote
López Martínez, José
(1 versión)
Obra

La ruta de Don Quijote-- y Azorín
Velasco Lizcano, Mariano
(1 versión)
Obra

En la ruta de Don Quijote
Ortiz Alfau, Ángel María
(1 versión)
Obra

Locations related with “El Quijote”



From metadata in images to full text

Messages to take home

1. **Library Linked Data can be easily integrated in other data (e.g. geographical, education, etc.)**
2. **Data providers** should include **language, provenance and license metadata** in their datasets
   - in the original data sources (e.g., MARC21 records)
   - tags into RDF (e.g., @es, @ fr at least)
   - language URIs in the VOID or DCAT descriptions
   - License and provenance information in RDF
3. **Benefits of adding language, provenance and license metadata in datasets**
   - Reduce the time and cost of identifying language in resources and terminology
   - Foster the aggregation and enrichment of data across complementary resources
   - Enhances data curation
   - Improves precision and recall in information retrieval and search