

Publishing Linked Data – There is no One-Size-Fits-All Formula

Asunción Gómez-Pérez

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

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

Acknowledgements:

O.Corcho, D. Garijo, D. Vila, L.Vilches, B. Villazón
Our partners at: BNE, IGN, ...

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0

LOV SYMPOSIUM: LINKING AND OPENING VOCABULARIES
18th June, 2012

1. The concept
2. Foundations
3. The process
4. Examples

- Libraries: <http://datos.bne.es>
- Geo: <http://geo.linkeddata.es/>
- Metereology:<http://aemet.linkeddata.es/>
- Travelling: <http://webenemasuno.linkeddata.es/>

Complex queries using data from heterogeneous Web pages

Catálogo BNE

Resultados de la búsqueda

Resultados no ordenados porque la búsqueda recuperó más de 500.

Cervantes Saavedra, Miguel de 1547-1616 La búsqueda recuperó 5764 Registros bibliográficos.

Páginas < 1 2 3 4 ... 100 ... 193 >

1997

#1 XX(239923-1) Hipsumuj Idago Don Khot Lamanškij. Toz 2 [Texto impreso] Cervantes Saavedra, Miguel de 1547-1616

Guardar 1 ejemplar disponible en Sede de Recursos en Salón General

#2 12/84771 Pasos y entremeses [Texto impreso] 1ª ed. Cervantes Saavedra, Juan Martín

Guardar 8 ejemplares disponibles en Sede de Alcalá

#3 CERV/CS/34 Homenaje a Cervantes [Texto impreso] Cervantes Saavedra, Miguel de 1547-1616

Guardar No hay ejemplares disponibles actualmente.

#4 12/865448 La cervantina [Texto impreso] Cervantes Saavedra, Miguel de 1547-1616

Guardar 2 ejemplares disponibles en Sede de Alcalá

1905

2011

Resumen

<http://www.bne.es/>

Cervantes enthusiast from Germany visiting Madrid and willing to know more about Cervantes' work and life

VIAF

Fichero de Autoridades Virtual Internacional

Búsqueda

Selección campo Selección índice Términos de búsqueda

Todos los nombres Todo VIAF

Búsqueda

Cervantes Saavedra, Miguel de 1547-1616

Cervantes, Miguel de, 1547-1616

De Cervantes Saavedra, Miguel

Сервантес Сааведра, Маркьон де, 1547-1616

1547-1616, מיגל דה סאוVEDRA

VIAF ID: 17220427 (Personal)

Formas preferidas

100 1 _ za Cervantes Saavedra, Miguel de, id 1547-1616

100 1 _ za Cervantes Saavedra, Miguel de, id 1547-1616

100 1 _ za Cervantes Saavedra, Miguel de, id 1547-1616

100 2 _ za Cervantes Saavedra, Miguel de, id 1547-1616

100 1 _ za Cervantes Saavedra, Miguel de, id 1547-1616

100 1 _ za Cervantes Saavedra, Miguel de, id 1547-1616

<http://www.viaf.org/>



Alcalá de Henares / Ruta de las Tapas

Ruta de las Tapas

Alcalá de Henares

Alcalá de Henares es conocida entre otras muchas cosas por su gastronomía y particularmente por sus tapas. Desde La Calle Mayor.net queremos ponerte fácil y por eso te ofrecemos nuestra "Ruta de las tapas" por la ciudad Cervantina. Una gran forma de acompañar la vista del patrimonio histórico y el gastronómico de la ciudad. Esta ruta se puede comenzar desde cualquier punto de la ciudad, aunque nuestra ciudad, la C/ visita a la casa de

<http://elviajero.elpais.com/>

Locales

Macandé

Categoría: Restaurantes

Descripción: Tapas exquisitas durante todo el año, terraza de temporada, especialidad en tostas, pates, ibéricos, menú diario. Muy centrado, idóneo para tertulias.

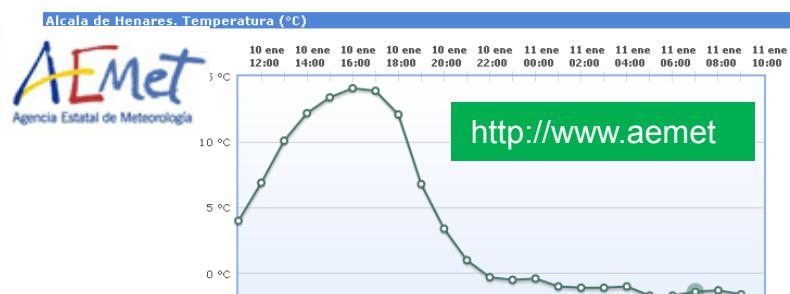
[Ampliar Información](#)

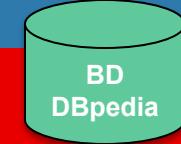
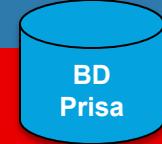
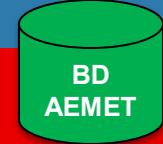
O'Bogavante Restaurante Marisquería

Categoría: Restaurantes

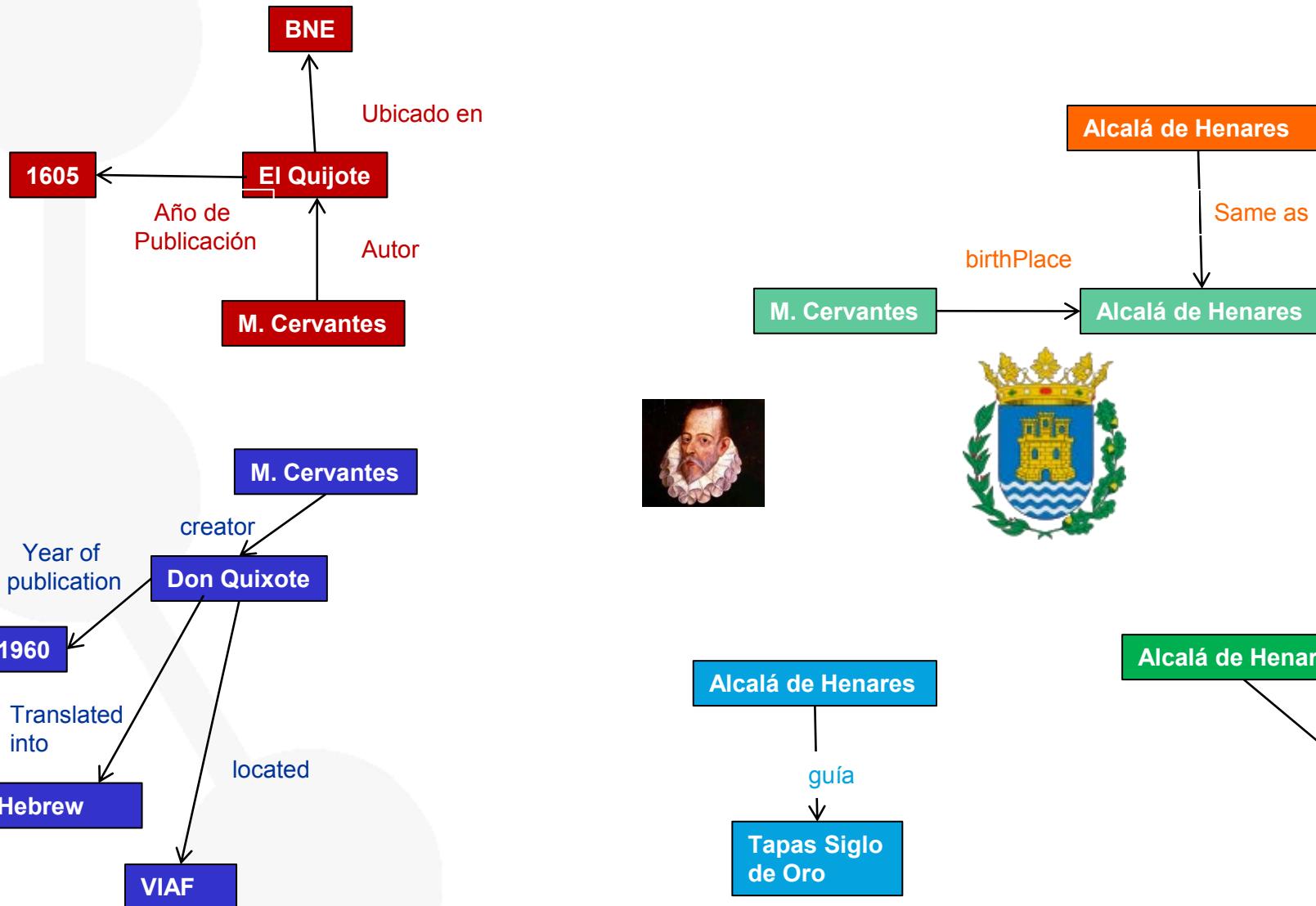
Descripción: Meco también es un lugar imprescindible para tapear en la Ciudad de la Luz. Te recomendamos: Parrillada de marisco, carne a la piedra, arroz con bogavante,... Viernes y Sabados Cenas - Baile musica de los 60 a los 90 Menú degustación 35 € por persona.

[Ampliar Información](#)





Data Integration

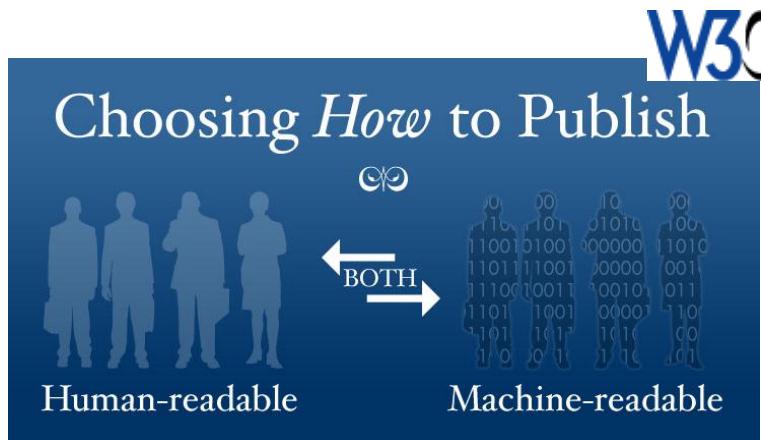
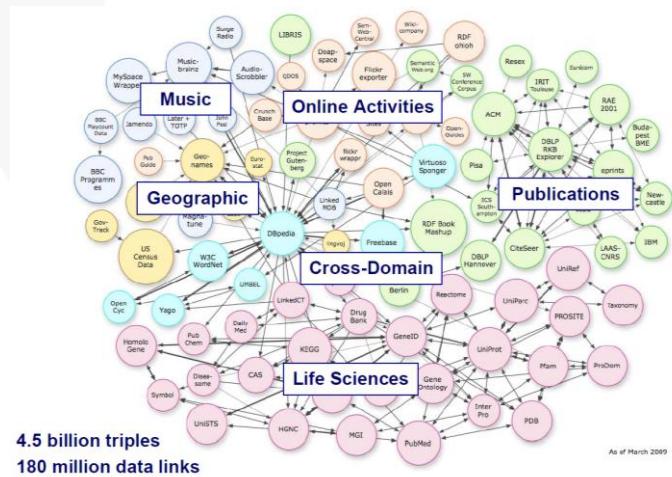


1. The concept
2. Foundations
3. The process
4. Examples

- Libraries: <http://datos.bne.es>
- Geo: <http://geo.linkeddata.es/>
- Metereology:<http://aemet.linkeddata.es/>
- Travelling: <http://webenemasuno.linkeddata.es/>

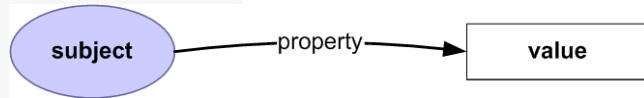
Linked Data: why it is important?

- Facilitate data integration
 - From heterogeneous sources
 - In different formats
 - Different granularity
 - In different languages
 - From different countries



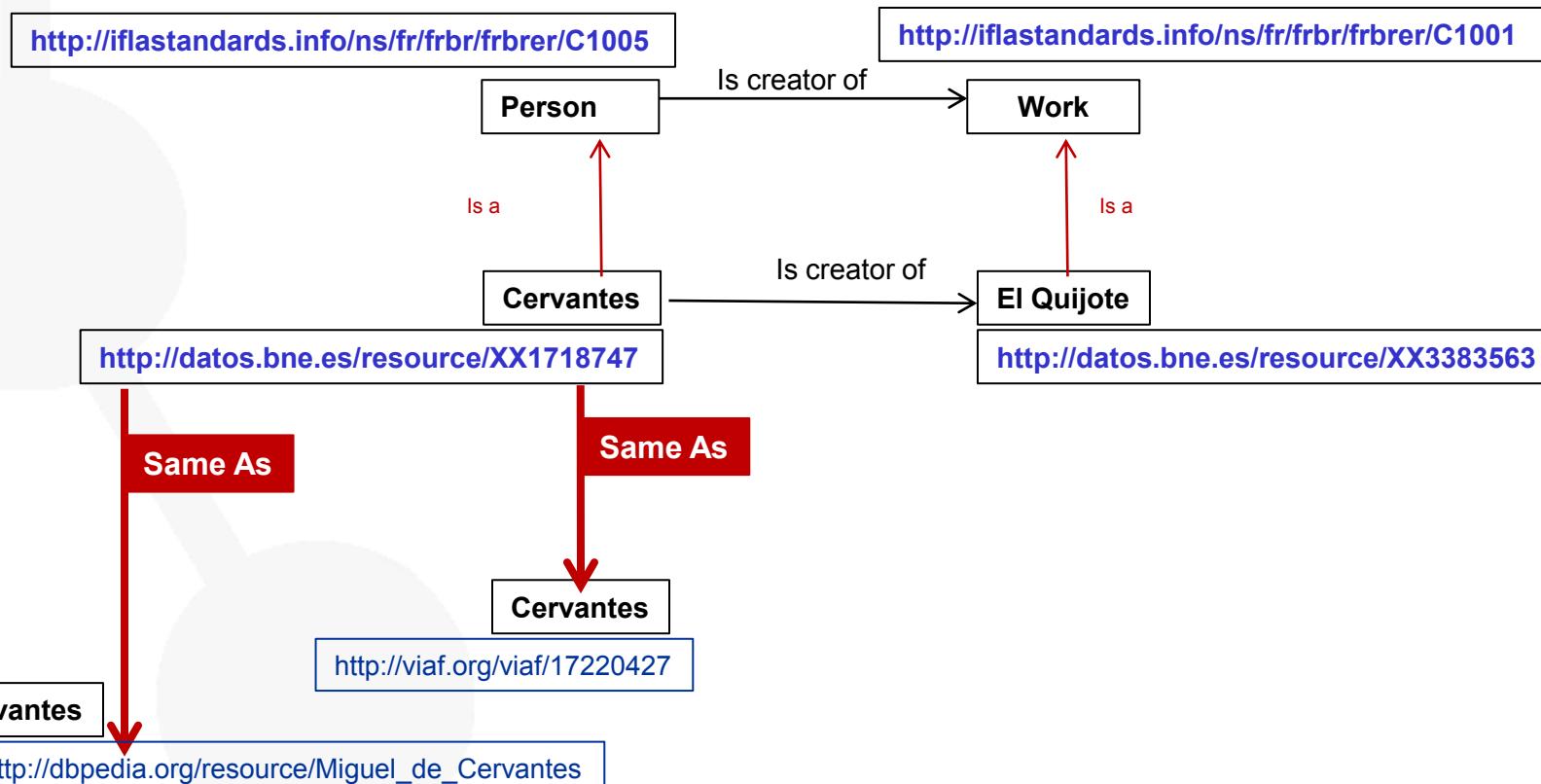
© Slide adapted from “5min Introduction to Linked Data”- Olaf Hartig

RDF(S) models

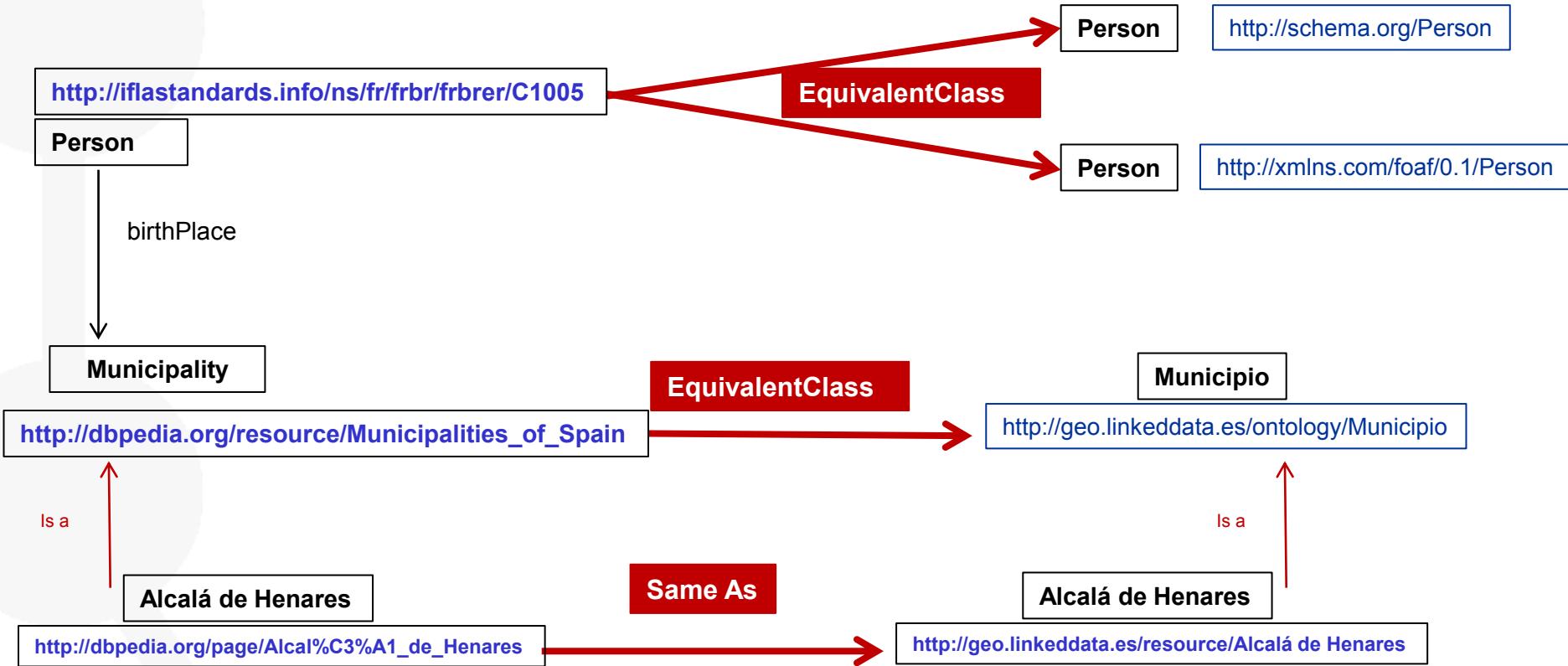


Unique identifiers: URI
identify or name a resource

Equivalence links to other datasets
Same As
Data navigation



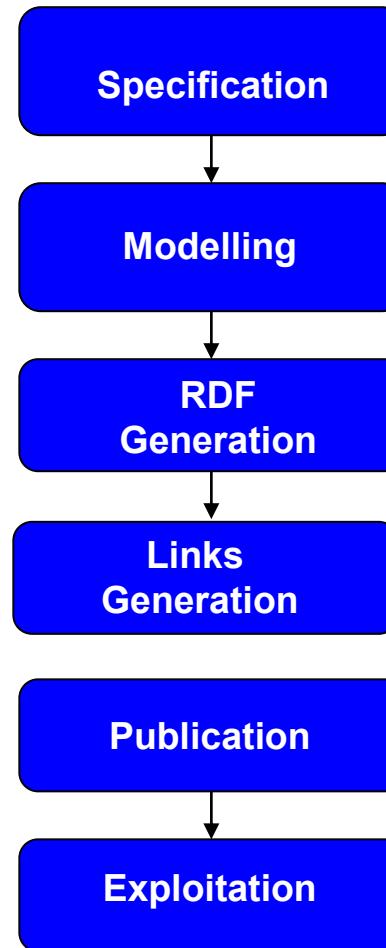
Aligning Models with Owl EquivalentClass

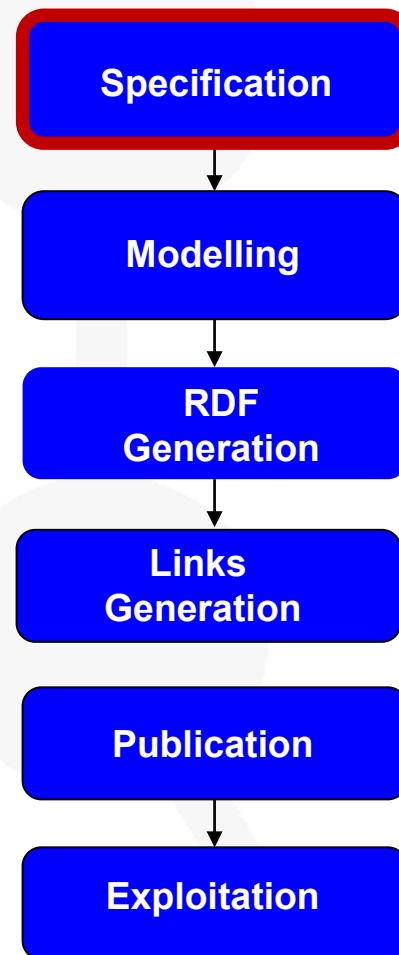


Lessons learnt

1. Reuse existing models
2. Align the data and the concepts.

1. The concept
2. Foundations
3. The process
4. Examples





- **Data sources analysis**
- **URI Design**
- **License definition**

Identification and selection of data sources

Specification

Modelling

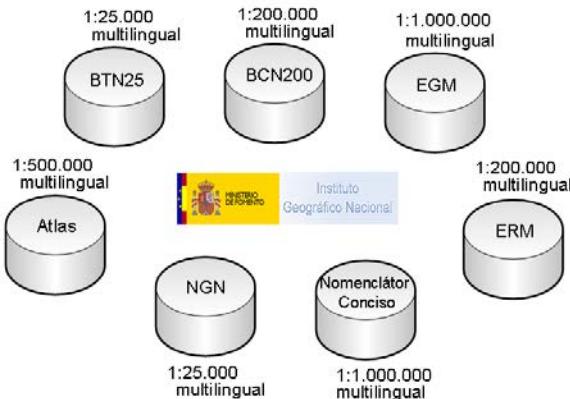
RDF Generation

Links Generation

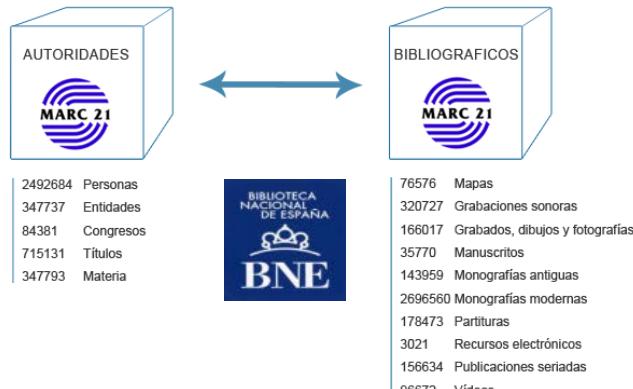
Publication

Exploitation

Geographical Spanish Institute



Spanish National Libraries



Statistical Spanish Institute



Industry



Population

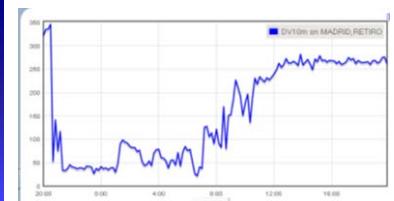


Building trade



Unemployment

Metereological Office (AEMET)



1. Identification and selection of the data sources

Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation

- Geographic Spanish Institute
 - Multilingual (Spanish, Vasc, Galician, Catalan)
 - Conceptualization mismatches
 - Granularity (scale concept)
 - Domain vocabulary
 - Inform. hidrográfica.** Embalse, albufera, río, etc.
 - Transportes.** Vía desdoblada, Ferrocarril, ...
 - Unidades Administrativas.** Municipio.
 - Particularities
 - Longitude and latitude
- Statistic Spanish Institute
 - Monolingual
 - Numerical information
 - Particularities
 - Geo (textual level) and Temporal

1. Identification and selection of the data sources: Geographical information

NOMBRE	ENTIDAD	COMPROV	CODINE	XUTM	YUTM	HUSO	HOJA25	LongitudG	LatitudG	LongitudGMS	LatitudGMS
Abejuela, Olivar de / Oli	Lugar/Paraje	0 02	02042	584300	246500	30	0867-3	-2,03512445814873	38,3620693572292	-02°02'06.44	38°21'43
Abejuela, Rambla de	Corriente fluvial	0 04	04053	597560	152000	30	0974-4	-1,89618811907142	37,509177056304	-01°53'46.27	37°30'33
Abejuela, Rambla de la	Corriente fluvial	1 46	46262	579190	123000	30	0638-2	-0,90285531710738	39,9374839997029	-00°54'10.27	39°56'14
Abejuela, Rambla de la	Corriente fluvial	1 44	44002	578800	124000	30	0638-2	-0,907142388290964	39,9465700679236	-00°54'25.71	39°56'47
Abelá, A	Población	0 27	27025	521500	336500	29	0003-3	-7,49293473672413	43,670782491816	-07°29'34.56	43°40'14
Abeladoira	Población	0 27	27039	522000	374030	29	0073-1	-7,50066276897865	43,1084031553505	-07°30'02.38	43°06'30
Abeladoira	Lugar/Paraje	0 27	27039	521550	374200	29	0073-1	-7,50615404547912	43,1100057447368	-07°30'22.15	43°06'36
Abelaedo	Lugar/Paraje	0 27	27005	548100	321300	29	0010-3	-7,16730269277051	43,5291689758955	-07°10'02.28	43°31'45
Abelaedo, Chao do	Llanura/Raso	0 27	27013	525500	330200	29	0009-1	-7,4448008623332	43,613414717808	-07°26'41.28	43°36'48
Abelaedo, Monte do	Lugar/Paraje	0 27	27064	509000	341000	29	0002-4	-7,6470271512085	43,7132267731492	-07°38'49.29	43°42'47
Abelaedo, O	Población	0 27	27064	508500	340500	29	0002-4	-7,65333254552948	43,708799463333	-07°39'11.99	43°42'31
Abelaedo, Rego do	Corriente fluvial	0 27	27064	509000	340540	29	0002-4	-7,64712030731384	43,7090862344245	-07°38'49.63	43°42'32

NOMBRE	GMLGEOMETRY
Abengibre	(HUGECLOB)
Alatoz	(HUGECLOB)
Albatana	<gml:Polygon srsName="SDO:8223" xmlns:gml="http://www.opengis.net/gml">
Balsa de Ves	<gml:outerBoundaryIs><gml:LinearRing><gml:coordinates decimal="." cs="," ts=" ">-1.48374108,39.23127677,0 -1.48404
Ballesteros, El	-1.48480765,39.22531695,0 -1.48544493,39.2231538,0 -1.4858725,39.22212319,0 -1.48608465,39.22141415,0 -1.4860584,
Alborea	-1.48421128,39.21319056,0 -1.48390698,39.21064575,0 -1.48429166,39.2094524,0 -1.48529268,39.20490629,0 -1.492506
Alcadozo	-1.49590016,39.20383639,0 -1.49905691,39.20405732,0 -1.50256876,39.20446292,0 -1.50338941,39.20453654,0 -1.50483
Alcalá del Júcar	-1.51378498,39.2047147,0 -1.51398161,39.20472622,0 -1.5171434,39.20471246,0 -1.51894734,39.20373528,0 -1.5196271,
	-1.52344981,39.20108038,0 -1.52699007,39.19901681,0 -1.5284903,39.19815278,0 -1.53151867,39.19619073,0 -1.532086
	→ -1.5329951,39.19479473,0 -1.53398595,39.19503246,0 -1.53537021,39.19528414,0 -1.53607317,39.19544615,0 -1.536914
	-1.53953498,39.19602121,0 -1.54099994,39.19629185,0 -1.54110394,39.19630216,0 -1.54231647,39.19646151,0 -1.54461
	-1.54957924,39.19633692,0 -1.55131402,39.1964386,0 -1.55382202,39.19666799,0 -1.55798564,39.19690882,0 -1.559800
	-1.56633376,39.19750743,0 -1.5703604,39.19765603,0 -1.57231502,39.19777011,0 -1.57278711,39.19790204,0 -1.573636
	-1.57633787,39.19982863,0 -1.57807687,39.20087703,0 -1.57814542,39.20092292,0 -1.57871557,39.20136237,0 -1.57915
	-1.5805024,39.20233924,0 -1.58088715,39.20278543,0 -1.58107708,39.20313014,0 -1.58122865,39.20365457,0 -1.581447
	-1.5819531,39.20563664,0 -1.58191475,39.20581637,0 -1.58194513,39.20603299,0 -1.58285235,39.20756676,0 -1.584270

1. Identification and selection of the data sources

Statistical information

Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation

POBLA_01	Población año 2001
POBLA_08	Población año 2008
PBL01_08	Evolución de la población desde el 2001 al 2008
TS_CREC_EF	Tasa de Crecimiento efectivo
DENS_01	Densidad de Población año 2001
DENS_08	Densidad de Población año 2008
DENS01_08	Evolución de la densidad desde el año 2001 al 2008
TS_CREC_VE	
TS_CREC_VE	
TS_CREC_VE	
TS_NAT_01	
TS_NAT_08	
TS_NAT01_0	
TS_MORT_01	
TS_MORT_08	
TS_MORTAL_	
TS_FECUN_0	
TS_FECUN_0	
TS_FECUN01	
MUJ_FERTIL	
MUJ_FERTIL	
MUJ_FETIL_	
POBL_EXTR_	
POBL_EXTR_	
EVOL_POBL_	
IND_DEP_01	
IND_DEP_08	
IND_DEPEN_	
ESP_IND_EN	
ESP_¬NDI_E	
ESP_IND_EN	
PORT_IND_E	
PORT_¬NDI_	

	2009	2008	2007	2006	2005	2004	2003
Total Nacional	3355830	3422239	3336657	3174393	3064129	2942583	2813159
Alava	21988	22318	20676	20349	19838	19779	19638
Albacete	27380	27647	27068	25531	24685	23550	22547
Alicante	136239	142307	140145	133016	123333	113852	111805
Almería	43501	45130	43970	40871	38766	36260	33947
Asturias	71853	73124	72276	70115	68175	67039	65062
Ávila	11455	11708	11434	10900	10611	10319	10211
Badajoz	40874	41358	40168	38045	37052	34972	34866
Illes Balears	91826	93335	91254	88027	87024	85425	75951
Barcelona	467385	477942	469432	444410	436294	417425	397693
Burgos	25567	25891	25372	24504	23733	22882	22159
Cáceres	26307	26494	26064	25039	24846	20596	23440
Cádiz	62817	64505	63338	61691	58986	57138	54462
Cantabria	39611	40393	39560	37690	36561	35649	34017
Castellón	42122	43855	42476	39749	37865	37214	34213
Ciudad Real	32046	33011	31881	30446	29521	29011	26778
Córdoba	48979	50057	49302	47155	45405	43394	41964
Coruña, A	83748	84220	82873	79170	77023	74809	71748
Cuenca	14747	14928	14741	13822	13336	12829	12546
Girona	58404	51467	50108	47169	46827	45145	52482
Granada	60016	62269	61055	57223	54341	50508	49662
Guadalajara	13507	13735	12874	11825	10438	10120	9422
Guipúzcoa	62034	63569	59546	58486	57193	56498	55983
Huelva	26783	27463	27063	25487	24777	24270	22547
Huesca	16837	17109	16694	16025	15390	15078	14283
Jaén	36557	37368	36962	35383	34675	33157	32444
León	33564	34012	33563	32359	31664	30992	30256
Lleida	36920	37638	36065	33956	32739	31515	29605
Lugo	24861	25035	24609	23780	23122	22479	22396
Madrid	511804	519307	503000	478202	456175	436074	407655
Málaga	113362	116683	114547	108713	102382	96587	88257
Murcia	95636	100075	97374	90698	85110	82484	75973
Navarra	43282	43847	43142	41083	40730	39679	38936
Ourense	23304	23711	23520	22843	22452	22118	21560
Palencia	10964	11111	11060	10694	10575	10399	10297

Specification

Modelling

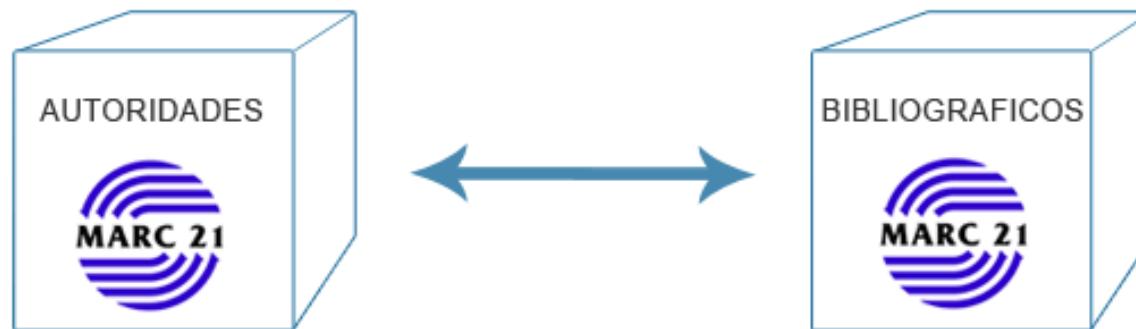
RDF Generation

Links Generation

Publication

Exploitation

- Records in the **MARC 21 format**
- **3.9 million** bibliographical records
- **4.2 million** authority records
- Version: November, 2011



2492684 Personas
347737 Entidades
84381 Congresos
715131 Títulos
347793 Materia

76576 Mapas
320727 Grabaciones sonoras
166017 Grabados, dibujos y fotografías
35770 Manuscritos
143959 Monografías antiguas
2696560 Monografías modernas
178473 Partituras
3021 Recursos electrónicos
156634 Publicaciones seriadas
96672 Vídeos

Specification



Modelling



RDF Generation



Links Generation



Publication



Exploitation

- Meaningful URIs versus Opaque URIs
- Separate TBox (ontology model) from ABox
- **Base URI**
<http://linkeddata.es/>
<http://datos.bne.es/>
<http://geo.linkeddata.es/>
<http://otalex.linkeddata.es/>
- **OntologyTBox URLs)**
<http://iflastandards.info/ns/fr/frbr/frbrer/C1005>
<http://phenomenontology.linkeddata.es/ontology/{concept|property}>
<http://phenomenontology.linkeddata.es/ontology/Municipio>
We use the RDF Data Cube Vocabulary and/or other vocabularies
- **Data (ABox URLs)**
<http://datos.bne.es/resource/XX1718747>
<http://geo.linkeddata.es/resource/{resource type}/{resource name}>
<http://geo.linkeddata.es/resource/Municipio/Badajoz>

Specification

- **Ontologies:**

- A set of terms
- A set of explicit assumptions regarding the intended meaning of the terms.
 - Almost always including concepts and their classification
 - Almost always including properties between concepts

Modelling

RDF
Generation

Links
Generation

Publication

Exploitation

- Shared understanding of a domain of interest
- Ontologies expressed in OWL or **RDF(S)**, both based on RDF
- The **NeOn methodology** helps to build ontologies

2. Vocabulary development

Specification

Modelling

RDF Generation

Links Generation

Publication

Exploitation

- Features
 - Lightweight :
 - Taxonomies and a few properties
 - Consensuated vocabularies
 - To avoid the mapping problems
 - Multilingual
 - Linked data are multilingual
 - The NeOn methodology can help to
 - Re-engineer Non ontological resources into ontologies
 - Pros: use domain terminology already consensuated by domain experts
 - Withdraw in heavyweight ontologies those features that you don't need
 - Reuse existing vocabularies

The Ontology for BNE: based on IFLA vocabularies

Specification

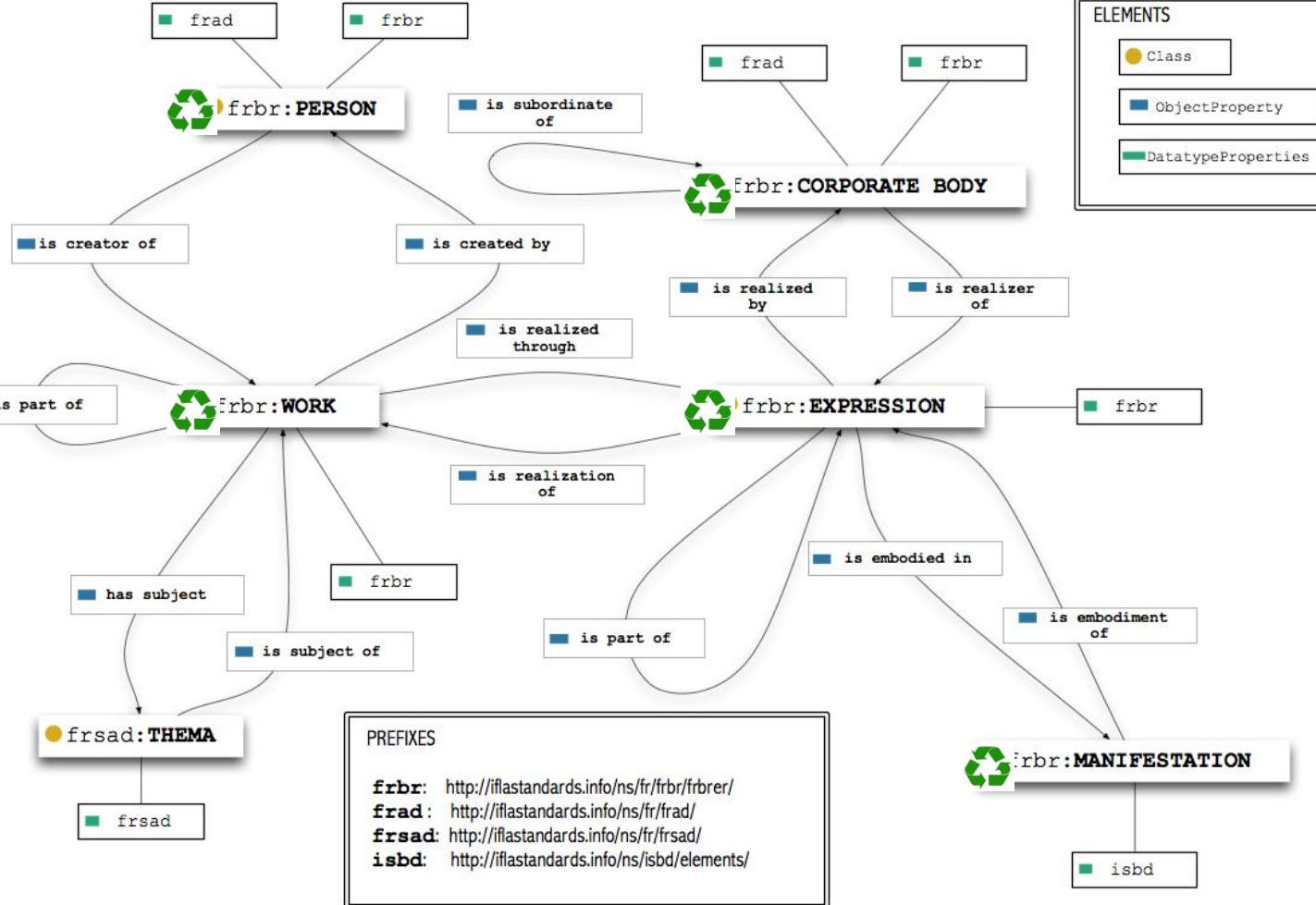
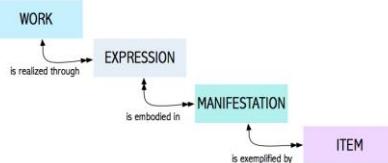
Modelling

RDF Generation

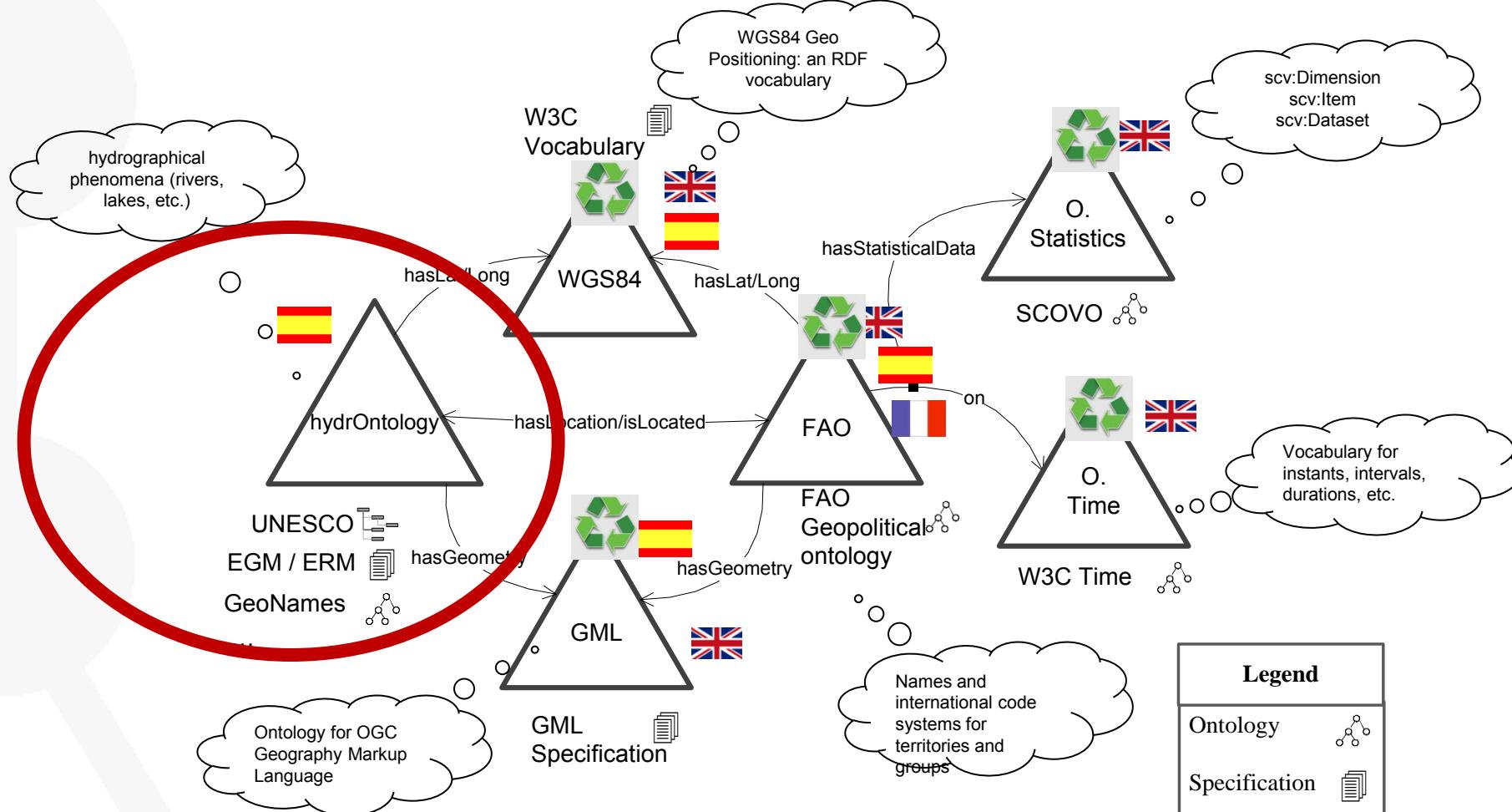
Links Generation

Publication

Exploitation



Geolinkeddata ontology



Legend	
Ontology	
Specification	
Thesaurus	

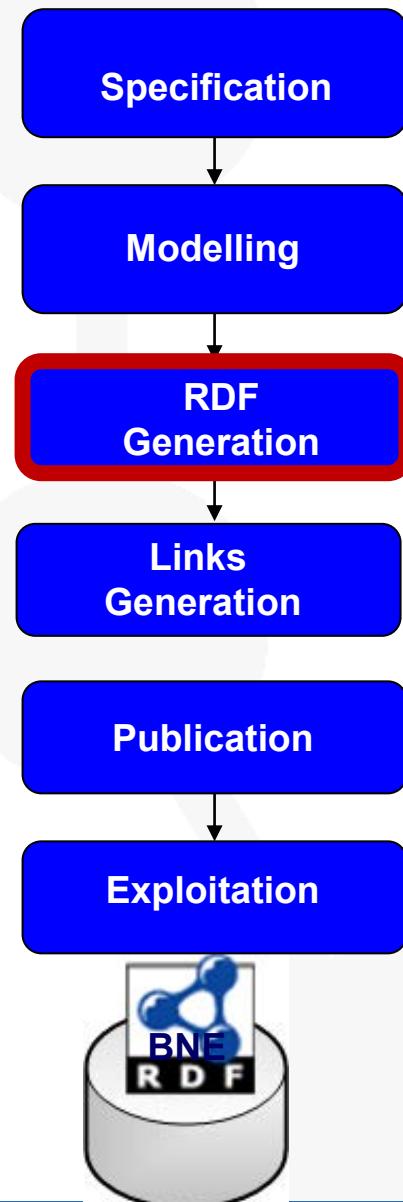
reused

Classes	33	33
Object Properties	44	44
Data Properties	318	318



Following the INSPIRE
(INfrastructure for SPatial InfoRmation in Europe) recommendation.
hydrOntology, SCOVO, FAO Geopolitical, WGS84, GML, and Time

3. Generation of RDF



- From the Data sources
 - Geographic information (Databases)
 - Statistic information (.xsl)
 - Geospatial information
 - Biobibliographic information (MARC 21)
- Different technologies for RDF generation
 - NOR20 (from excell, XML, text files, ...)
 - R20 and ODEMapster (from Databases)
 - Geometry2RDF and SPh2RDF (for Geo data)
 - Marimba for Libraries



Libraries: Marimba uses the ontology to generate RDF



Specification

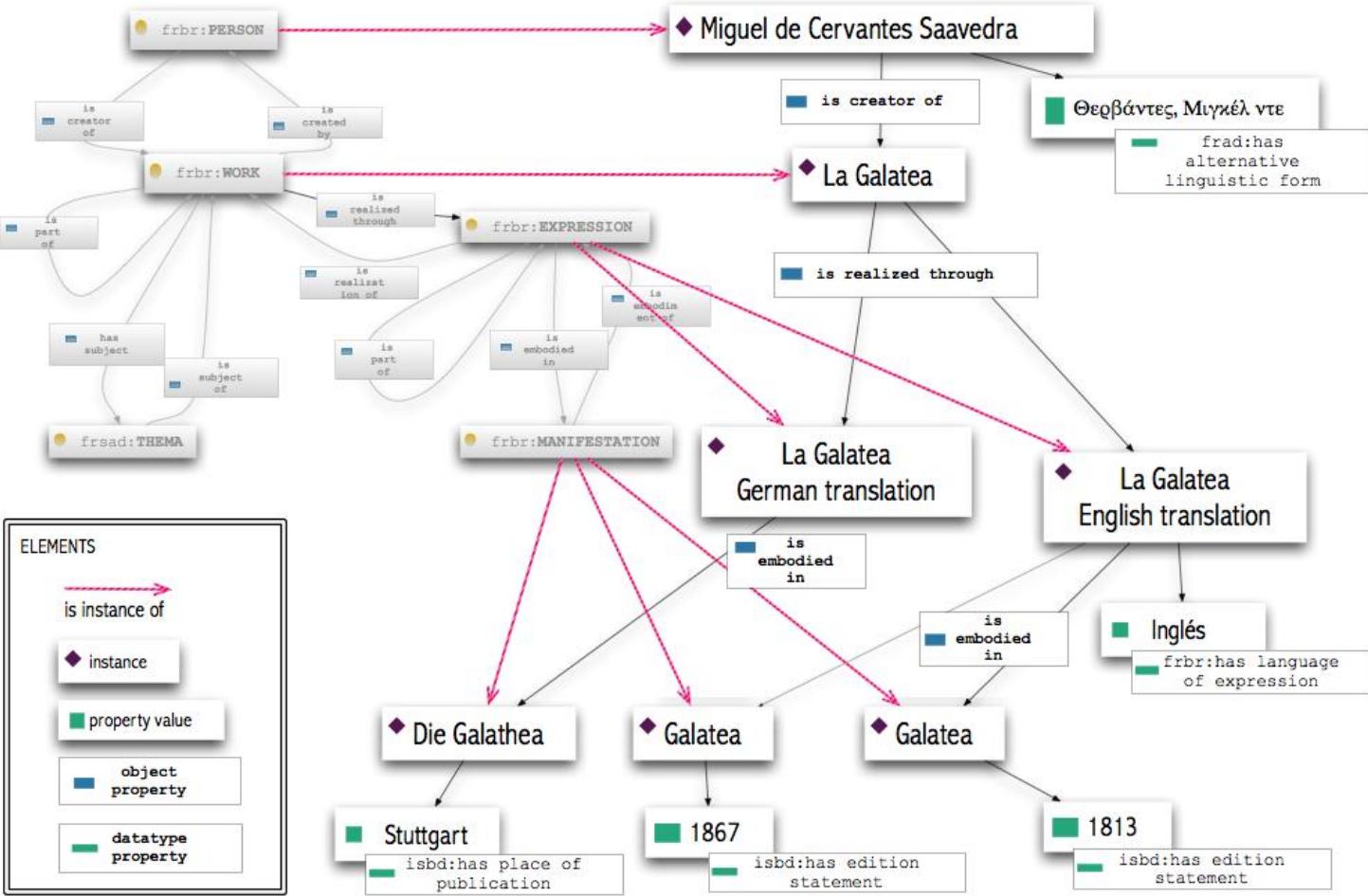
Modelling

RDF Generation

Links Generation

Publication

Exploitation



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

Specification

Modelling

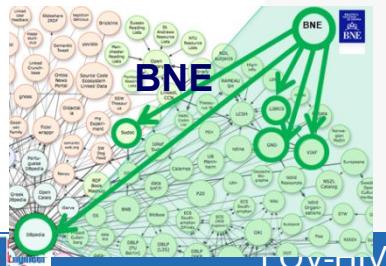
RDF
Generation

Links
Generation

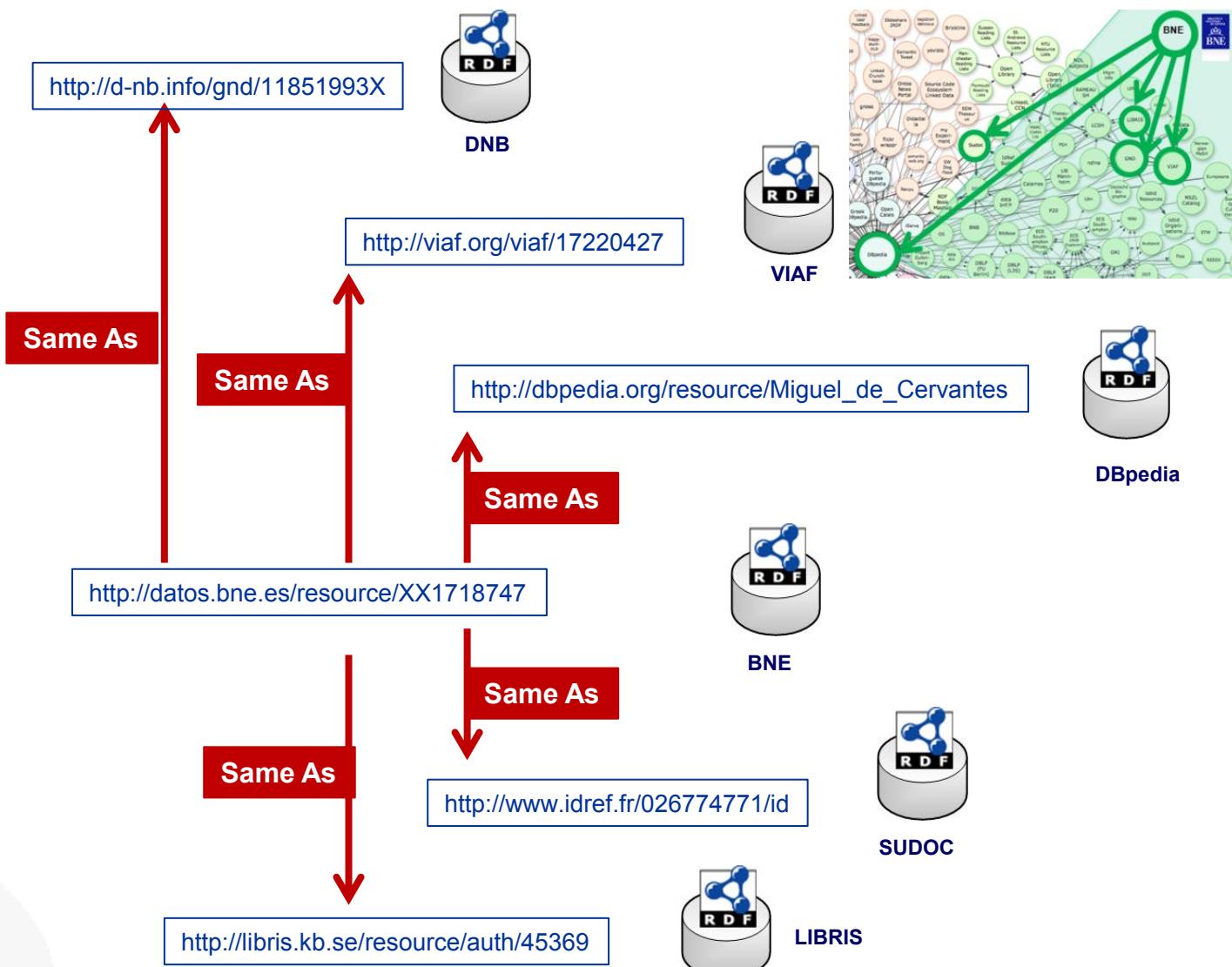
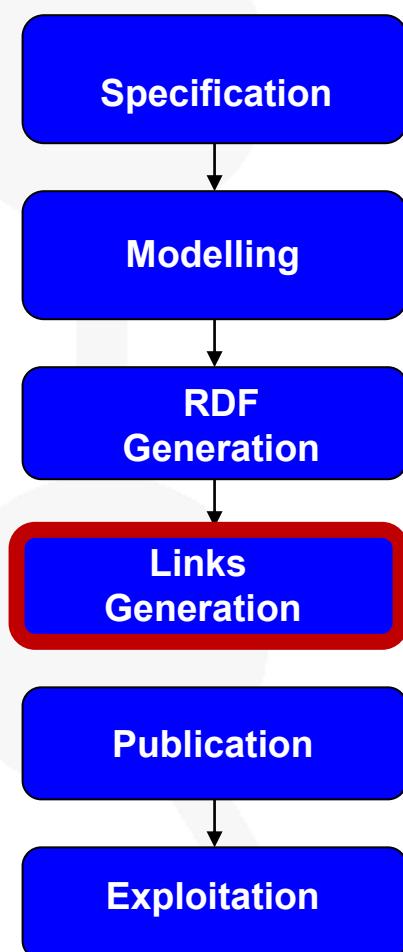
Publication

Exploitation

ifla-frbr:P3039	■ bne:resource/XX5027917
ifla-frbr:P3040	■ bne:resource/XX5035214
ifla-frad:P4031	■ Cervantes Saavedra, Miguel de ■ 1547-1616
	■ Cervantes ■ Cervantes di Saavedra, Michele ■ Cervantes, Miguel de ■ Servantes Saavedra, Migel ■ Sirfantis Saafedrä, Mīgīl dī ■ Therbantes, Minkel nte ■ Zerbantes eta Saabedra, Mikel ■ Θερβάντες, Μιγκέλ ντε ■ Сервантес Сааведра, Мигель де ■ سرفانتس سافرا، میگل دی
locmads:citationNote	■ (Cervantes Saavedra, Miguel de (Alcalá de Henares, Madrid, 1547-Madrid, 1616)) ■ port. (Migel de Servantes Saavedra) ■ port. (Miguel de Cervantes Saavedra) ■ port. (Miguel de Cervantes) ■ port. (Μιγκέλ ντε Θερβάντες = Minkel nte Therbantes) ■ port. (میگل دی سرفانتس سافرا) = Mīgīl dī Sirfantis Saafedrä) ■ vol. 1, port. (Miguel' de Servantes Saavedra = Мигель де Сервантес Сааведра)
ifla-frbr:C1005	■ http://d-nb.info/gnd/11851993X ■ http://libris.kb.se/resource/auth/45369 ■ http://viaf.org/viaf/17220427 ■ http://www.idref.fr/026774771/id ■ ifla-frbr:C1005



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



- vol. 1, port. (Miguel' de Servantes Saavedra = Мигель де Сервантес Сааведра)
- http://dbpedia.org/resource/Miguel_de_Cervantes
- <<http://d-nb.info/gnd/11851993X>>
- <<http://libris.kb.se/resource/auth/45369>>
- <<http://viaf.org/viaf/17220427>>
- <<http://www.idref.fr/026774771/id>>

Specification

Modelling

RDF Generation

Links Generation

Publication

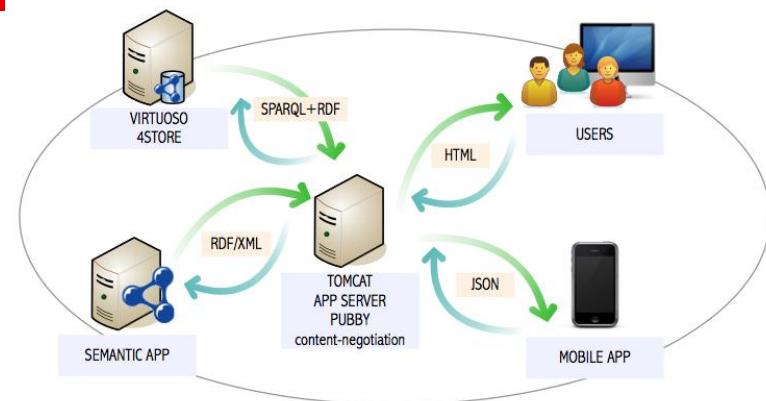
Exploitation

Data publication

Metadata publication using VOID

To facilitate the discovery

- Register in CKAN your dataset
- Use **sitemap4rdf** to generate the site map
- Upload the site map to **Google** and **Sindice**



Specification

Modelling

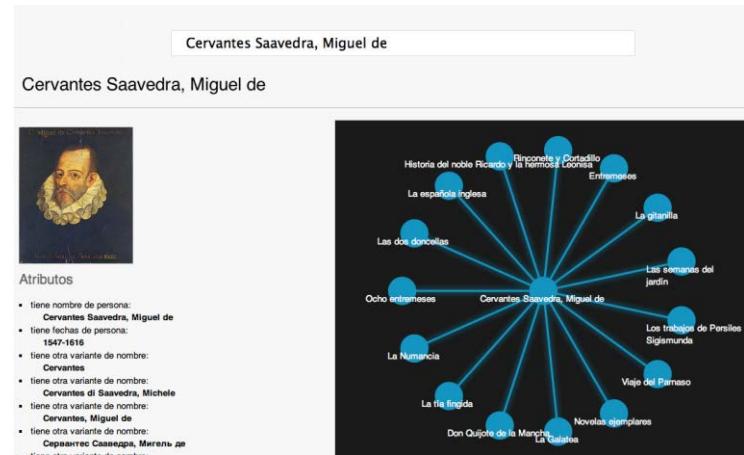
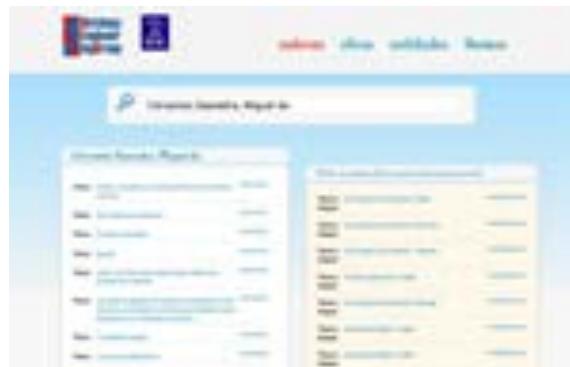
RDF Generation

Links Generation

Publication

Exploitation

Web Interface



SPARQL queries

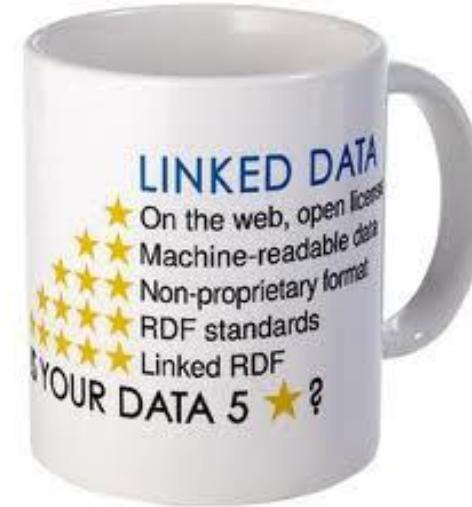
```

select distinct COUNT(?Obras) where {
  http://datos.bne.es/resource/XX1718747
  <http://iflastandards.info/ns/fr/frbr/frbrer/P2010>
  ?Obras
}
  
```

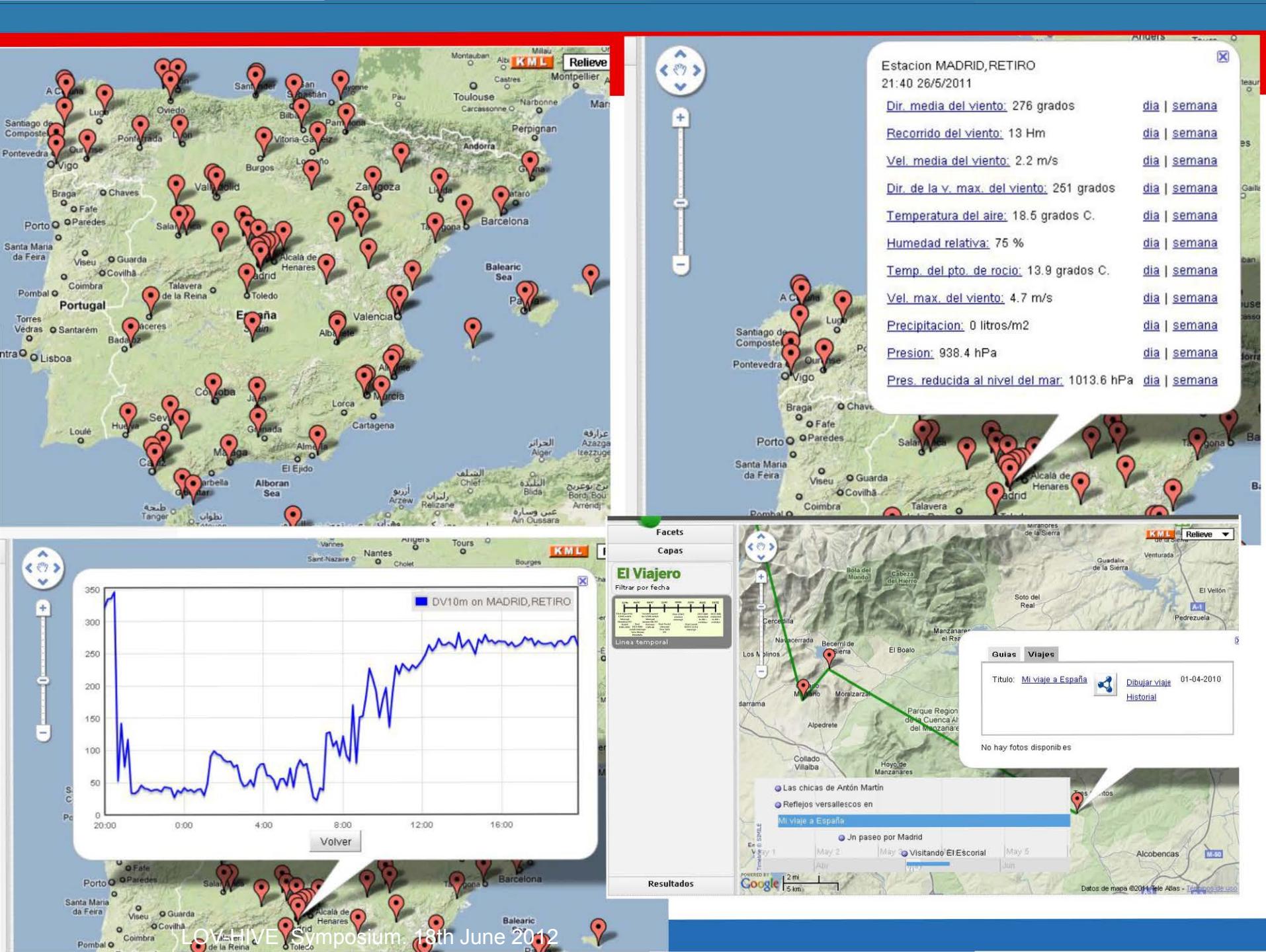
URI Cervantes

Is author

1. The concept
2. Foundations
3. The process
4. Examples



- Libraries: <http://datos.bne.es>
- <http://linkeddata3.dia.fi.upm.es/bne-demo>
- Geo: <http://geo.linkeddata.es/>
- Metereology: <http://aemet.linkeddata.es/>
- Travelling: <http://webenemasuno.linkeddata.es/>

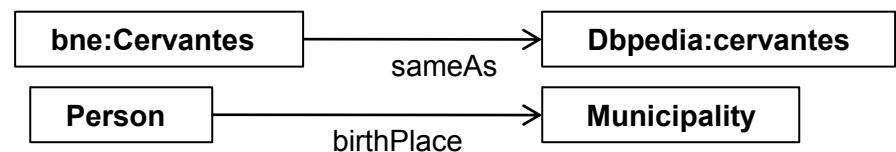


There is no One-Size-Fits-All Formula

Phase	BNE	IGN	AEMET	PRISA	INE
Modeling	  hydronontology Wgs84 time W3C® 	 W3C® SSN ontology		 Open Provenance Model  SIOC	Scovo Data cube
RDF generation	MARiMbA 	R2O RODEMapster geometry2rdf NOR2O	Ontology Engineering Group CSV parser	R2O RODEMapster CSV parser	Ontology Engineering Group NOR2O
Links generation	 DNB VIAF LIBRIS DBPEDIA	Silk DBPEDIA Geonames	Silk Geolinkeddata.es	Silk DBPEDIA Geolinkeddata.es	NOR2O Geolinkeddata.es
Publication					
Exploitation			SPARQL		

- **URI**
 - Follow existing design guidelines for new URIs
 - **Reuse existing URIs from authoritative sources**
- **Models**
 - **Reuse existing models when available**
 - Create new models from authoritative sources
 - Do not forget to align your model with existing models
- **Generation**
 - Vertical domains usually require specific tools for generation
- **Link**
 - Generic link discovery tools performs well in vertical domains
 - **Link to other data sets using**
 - Equivalence links (`sameAs`)
 - Typed links
- **Discovery**
 - Use `sitemap4rdf` to allow search engines to find your data
 - Use an **iterative-incremental life cycle** in your development

Learn about Linked Data with UPM official courses in one week



Publishing Linked Data – There is no One-Size-Fits-All Formula

Asunción Gómez-Pérez

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

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

Acknowledgements:

O.Corcho, D. Garijo, D. Vila, L.Vilches, B. Villazón
Our partners at: BNE, IGN, ...

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0

LOV SYMPOSIUM: LINKING AND OPENING VOCABULARIES
18th June, 2012