



Lift your data

Introduction to Linked Data

Ghislain Atemezing¹ , Boris Villazón-Terrazas²

1 EURECOM, France

auguste.atemezing@eurecom.fr

2 Ontology Engineering Group, FI, UPM

bvillazon@fi.upm.es

Slides available at: <http://www.slideshare.net/atemezing/>

INSPIRE Conference 2012, Istanbul

Acknowledgements: Oscar Corcho, Asunción Gómez-Pérez, Luis Vilches, Raphaël Troncy, Olaf Hartig, Luis Vilches and many others that we may have omitted.

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

- Linked Data
- Geospatial LD Datasets
- 5-star deployment scheme for Linked Open Data

- Linked Data
- Geospatial LD Datasets
- 5-star deployment scheme for Linked Open Data

Geospatial
Database
(Spain)



Statistical
Database
(Spain)



The screenshot shows the homepage of the Instituto Geográfico Nacional (IGN) website. The header includes the IGN logo, the Spanish Government logo, and the Ministry of Development logo. The main navigation bar contains links for Inicio, Acerca de IGN, Actividades, Herramientas, Servicios del CNIG, and Información de compra. The main content area features a sidebar with links to various services like Servicio de Información Sísmica, Datos Geodésicos, and Datos Geomagnéticos. The central area has large icons for 'Buscar datos', 'Descargas', and 'Comprar'. A large blue box with white text is overlaid on the page, stating 'Data exposed to the Web via HTML, pdf, etc.' Below this, there is a section titled 'Cifras de población referidas al 01/01/2009 Real Decreto 1918/2009, de 11 de diciembre'. This section contains a 'Resumen por Provincias' with links to various population statistics, a 'Resumen por Capitales de provincias', a 'Resumen por Comunidades Autónomas', and a 'Resumen por Islas'. At the bottom, there is a 'Detalle municipal' section with links to download Excel files for various provinces, including Álava, Albacete, Alicante/Alacant, and Almería.



The screenshot shows the homepage of the Instituto Geográfico Nacional (IGN) website. At the top, there are logos for the Instituto Geográfico Nacional, the Government of Spain, and the Ministry of Forestry. Below the logos is a navigation bar with links: Inicio, Acerca de IGN, Actividades, Herramientas, Servicios del CNIG, and Información de compra. A dropdown menu for 'Servicio de Información' is open, showing options: Datos Geodésicos, Datos Geomagnéticos, Datos Gravimétricos, Vigilancia Volcánica, Series Cartográficas, Atlas Nacional de España, and Fotos aéreas e imágenes de satélite. The main content area features four large blue circular buttons: 'Consultar y visualizar', 'Buscar datos', 'Descargas', and 'Comprar'. A black arrow points to the 'Buscar datos' button. On the right side, there is a 'Suscribirse' section with social media icons and a 'Noticias' section with a list of recent news items. At the bottom, there is a yellow banner with the text 'Sede Electrónica del Catastro'.



2.1.1	Listado de PROVINCIAS	7
2.1.2	Listado de MUNICIPIOS de una provincia	8
2.1.3	Listado de VÍAS de un municipio	9
2.1.4	Listado de los NÚMEROS de una vía	10
2.1.5	Consulta de DATOS CATASTRALES NO PROTEGIDOS de un inmueble identificado por su localización.	11
2.1.6	Consulta de DATOS CATASTRALES NO PROTEGIDOS de un inmueble identificado por su Referencia Catastral.	17
2.1.7	Consulta de DATOS CATASTRALES NO PROTEGIDOS de un inmueble identificado por polígono-parcela.	17
2.2	Convertor de coordenadas.	18
2.2.1	Conversión X,Y a RC.	18
2.2.2	Conversión RC a X,Y	19
2.2.3	Localización de referencias por proximidad a coordenadas X,Y.	20

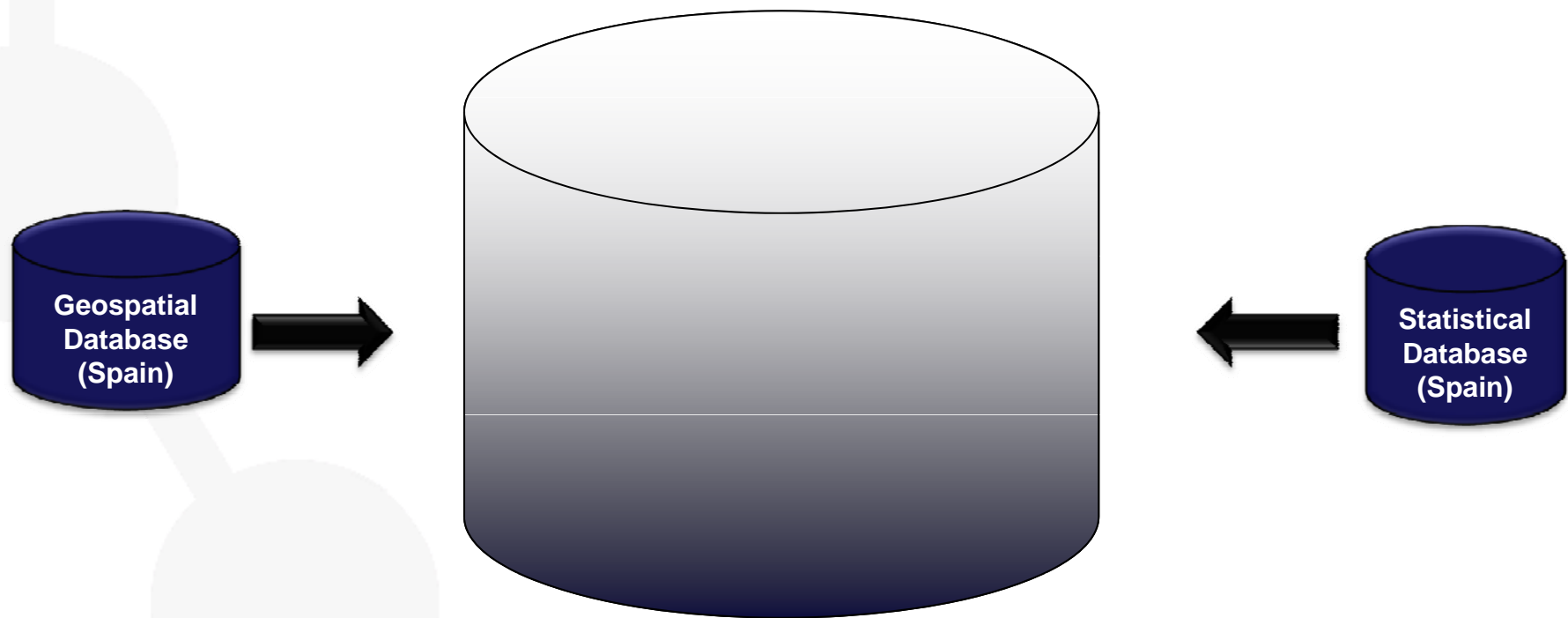
Complex queries over multiple pages/data sources?

© Slide adapted from "5min Introduction to Linked Data"- Olaf Hartig
Tutorial "Lift your data", Istanbul 2012

6

What do we actually want?

- Use the Web like a single global database
 - web of documents -> web of data



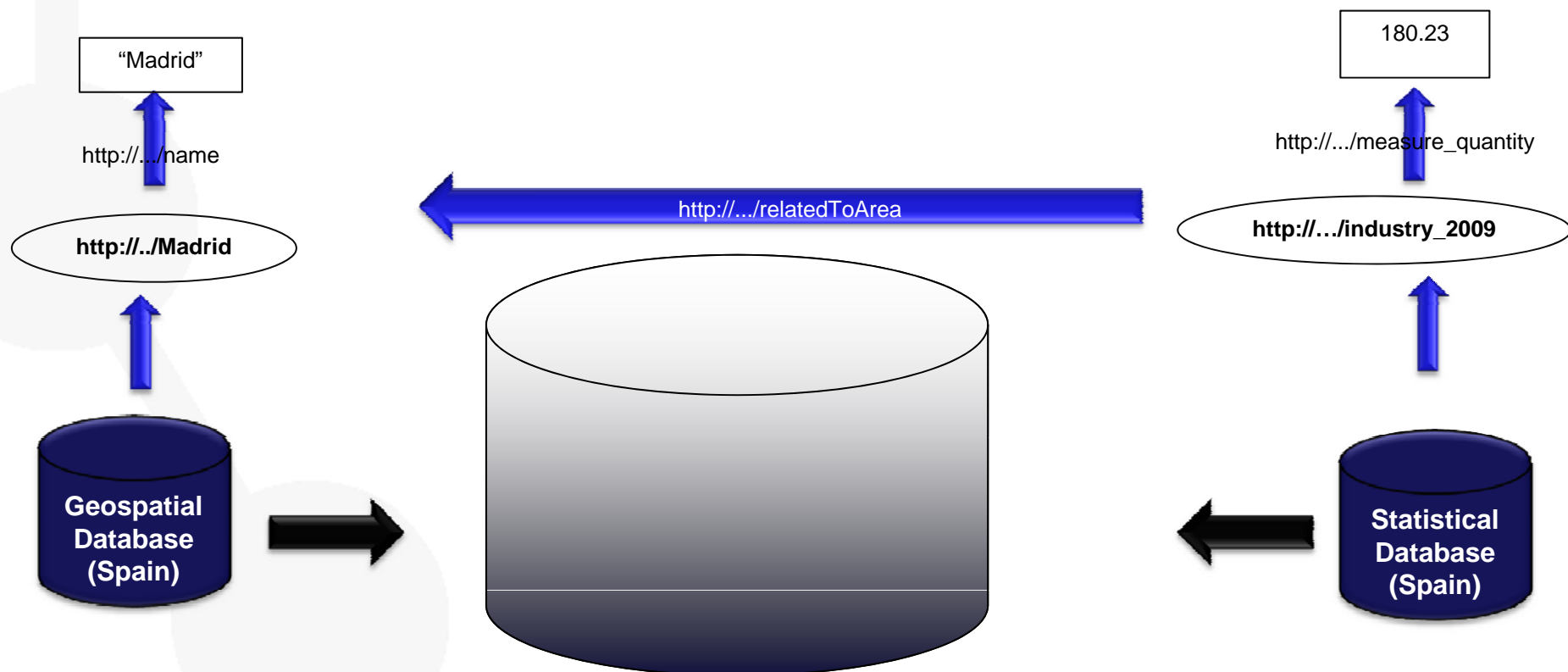
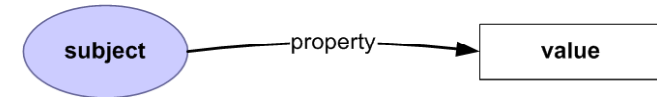
Linked Data enables such Web of Data

Global Identifier: URI (Uniform Resource Identifier), which is a string of characters used to identify a name or a resource on the Internet.

Data Model: RDF (Resource Description Framework), which is a standard model for data interchange on the Web

Access Mechanism: HTTP

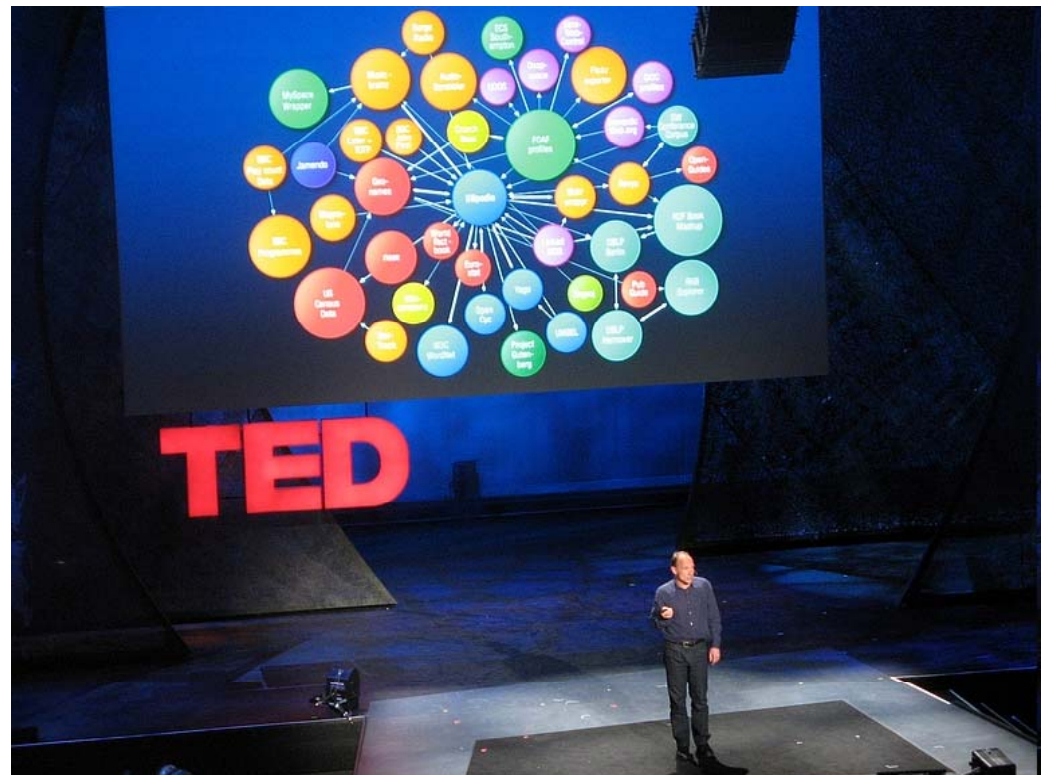
Connection: Typed Links



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

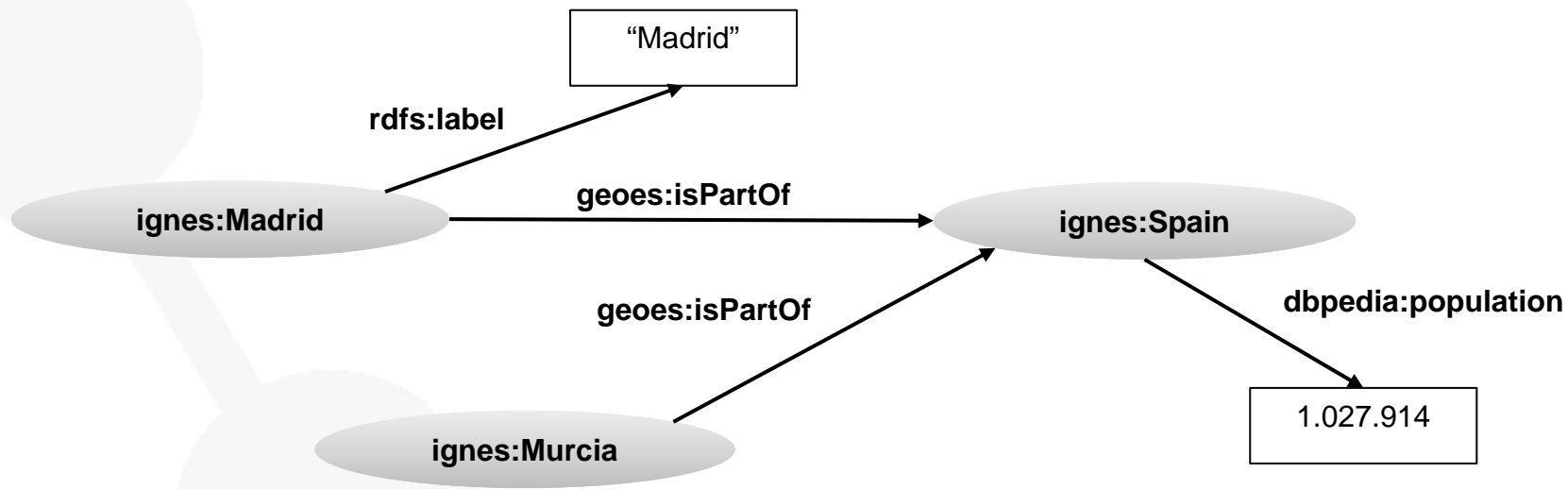
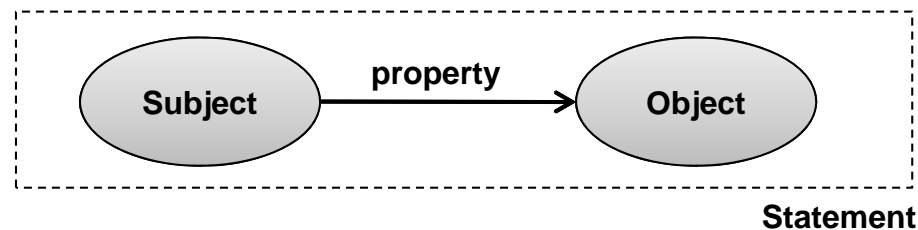
The four principles (Tim Berners Lee, 2006)

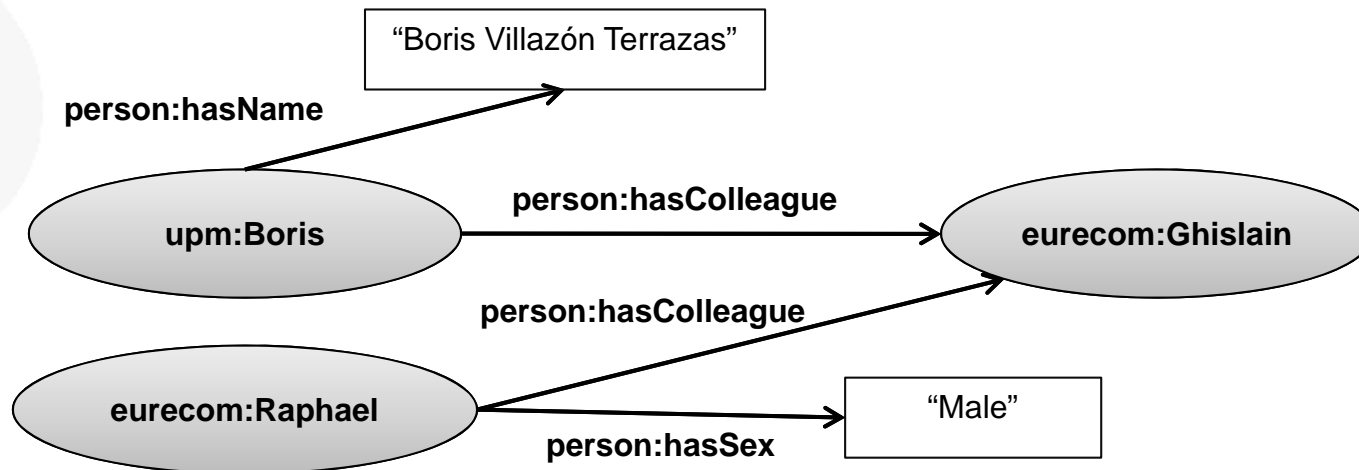
1. Use URIs as names for things
 2. Use HTTP URIs so that people can look up those names.
 3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
 4. Include links to other URIs, so that they can discover more things.
- <http://www.w3.org/DesignIssues/LinkedData.html>
 - http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html



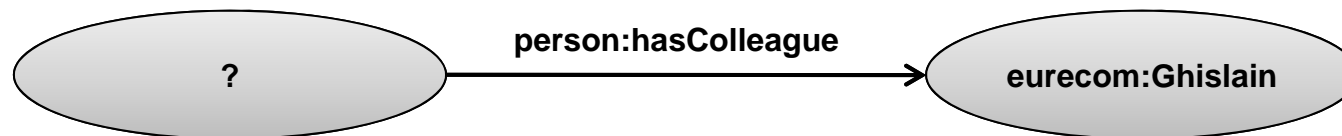
Resource Description Framework (RDF)

- RDF is a basic KR language based on semantic networks
 - Useful to represent metadata and describe any type of information in a machine-accessible way.



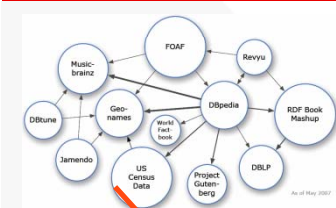


- Query: "Tell me who are the persons who have Ghislain as colleague"

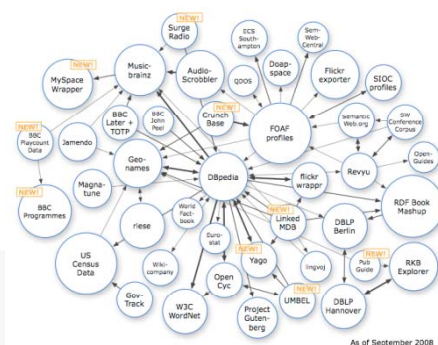


- Result: **upm:Boris** and **eurecom:Raphael**
- SPARQL query language for RDF. W3C recommendation
SELECT ?s
WHERE { ?s person:hasColleague eurecom:Ghislain . }

Evolution of the *Linked Open Data*

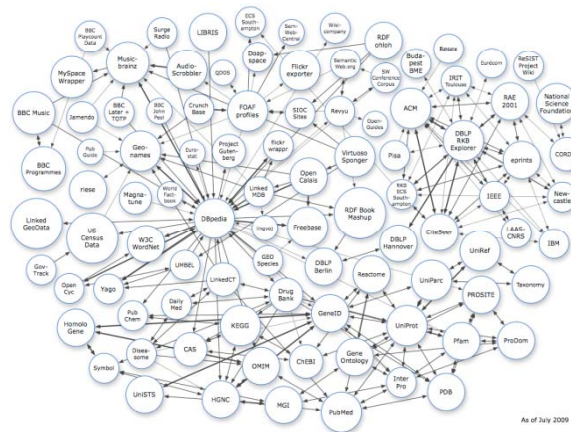


■ 2007

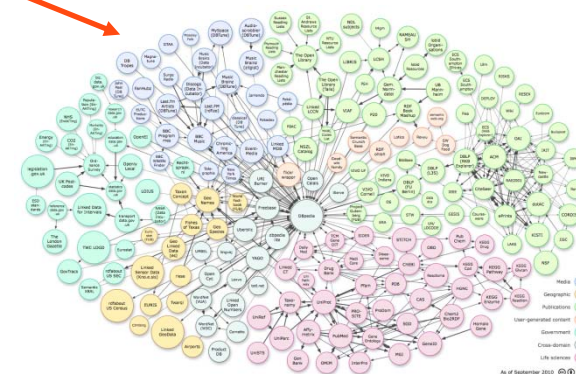


2008

2009

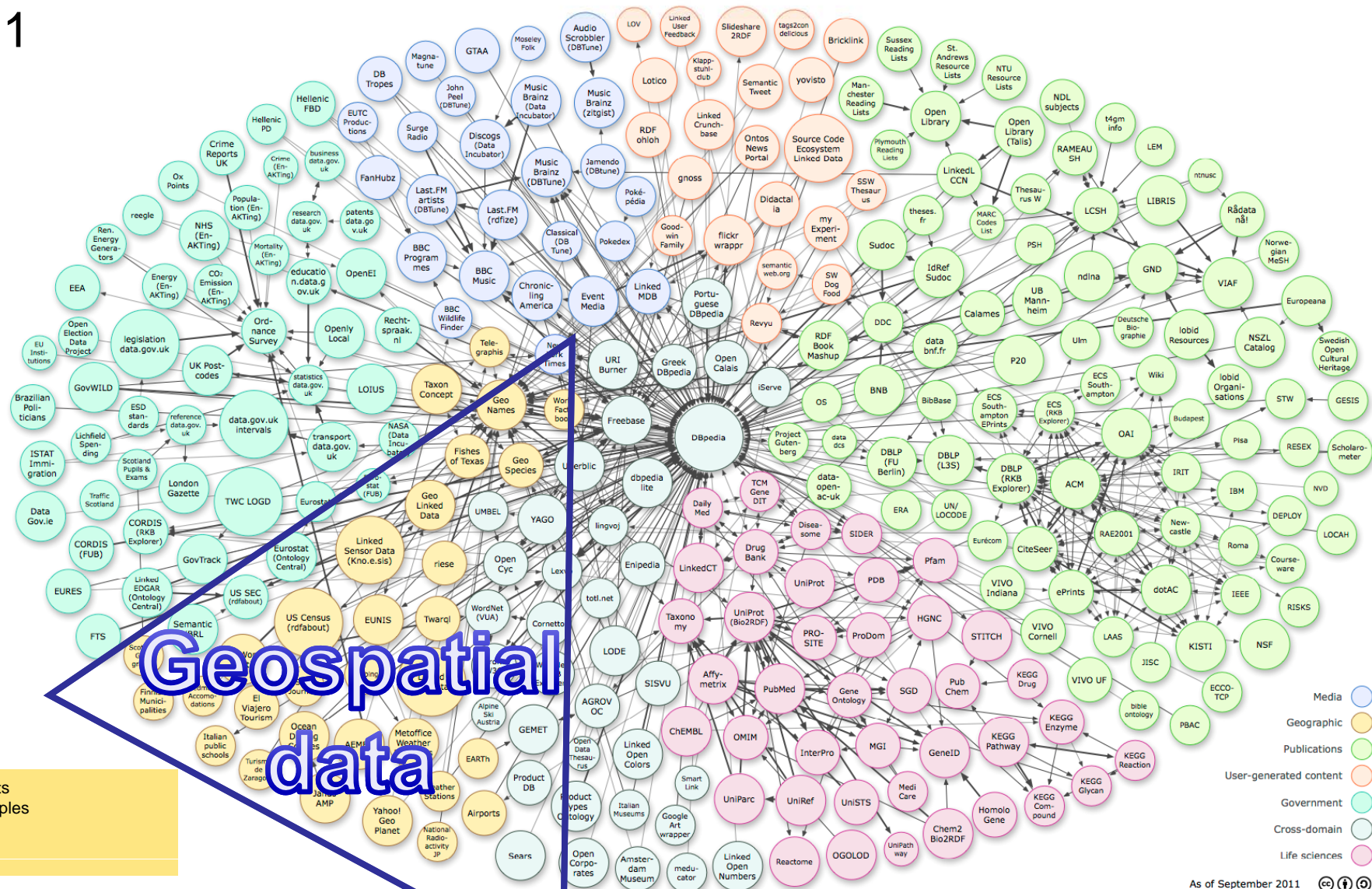


■ 2010



Linked Open Data

2011



<http://lod-cloud.net/state>

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

- Linked Data
- Geospatial LD Datasets
- 5-star deployment scheme for Linked Open Data

GeoData is becoming increasingly relevant

- Datasets in the geospatial domain contribute more than one fifth of the RDF triples in the Web of Data.
- Well-know Geo LD datasets
 - Ordnance Survey
 - GeoLinkedData
 - LinkedGeoData
 - GeoNames
 - DBPedia
 - GDAM-RDF

- Topic: Administrative units
- Datasets: The administrative gazeteer for Great Britain.
- URI pattern: <http://data.ordnancesurvey.co.uk/id/{id}>
- Vocabulary: Spatial Relations Ontology, Administrative Geography Ontology, WGS84, FOAF, and Gazeeter Ontology

- Topic: Hydrography, Administrative units, statistical information, meteorological features
- Datasets: Statistical, geospatial, and meteorological data.
- URI pattern:
 - <http://geo.linkeddata.es/ontology/{concept|property}>
 - <http://geo.linkeddata.es/resource/{type}/{name}>
- Vocabulary: SCOVO, FAO Geopolitical, hydrOntology, WSG84, GeoLinkedData Geometry Model, and Time Ontology.

- Topic: Points of interest
- Datasets: OpenStreetMap database
- URI pattern: <http://linkedgeodata.org/triplify/{id}>
- Vocabulary: LGD Ontology, WGS84, NeoGeo

- Topic: Toponyms
- Datasets: Datasets used by geonames
- URI pattern: <http://sws.geonames.org/{id}>
- Vocabulary: GeoNames Ontology, WGS84

- Topic: General knowledge
- Datasets: Wikipedia
- URI pattern: <http://dbpedia.org/resource/{name}>
- Vocabulary: DBPedia ontology, WGS84

- Topic: Global administrative areas
- Datasets: countries and lower level subdivisions.
- URI pattern: <http://gadm.geovocab.org/id/{id}>
- Vocabulary: gadm ontology, NeoGeo Ontology, DBPedia ontology, WGS84

- Linked Data
- Geospatial LD Datasets
- 5-star deployment scheme for Linked Open Data

The Five Stars deployment scheme

Is your data five stars?

- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ use non-proprietary formats (e.g., CSV instead of Excel)³
- ★★★★ use URIs to identify things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵



<http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

One Star: Open Data & Open License

data.gouv.fr^{BETA}
INNOVATION TRANSPARENCE . OUVERTURE

data.gov.uk^{BETA}
Opening up government

Proyecto
aporta
Reutilización de la Información
del Sector Público



nature.com linked data

License	Domain	By	SA
Creative Commons Attribution	Content	Y	N
Creative Commons Attribution Share-Alike	Content	Y	Y
Creative Commons CCZero	Content, Data	N	N
GNU Free Documentation License	Content	Y	Y
UK PSI Public Sector Information	Content, Data	Y	N
Free Art License	Content	Y	Y
MirOS License	Code, Content	Y	N

<http://opendefinition.org/licenses/>

Two stars: Reusability

- Data are reusable
- Some formats are helpful
 - Excel, CSV, JSON, XML, GML
- Others not really
 - PDF, HTML, MS Word



Credit: Bernadette Hyland : <http://www.slideshare.net/3roundstones>

Three stars: Specialist formats

- Specialist tools often have specialist formats
 - Few people have the tools
 - Expensive
 - Difficult to re-use
 - Geospatial tools, statistical packages, etc..
- Use Open standards
 - CSV, JSON, XML, GML, OGC Web services, TSV, RDF

Credit: Richard Ciganiak: <http://www.slideshare.net/cygri/edf2012-the-web-of-data-and-its-five-stars>

Towards INSPIRE (5-Stars) Dataset?

Welcome to the INSPIRE geoportal

The INSPIRE Directive requires the Commission to establish a community geo-portal and the Member States shall provide access to their infrastructures through the geo-portal as well as through any access points they themselves decide to operate.

[More...](#)

Discovery / Viewer

Search, discover and access geographic information provided by European governmental, commercial, and non-commercial organizations.

[More ...](#)

Validator


The purpose of the INSPIRE Metadata Validator is to test the compliancy of INSPIRE metadata with the INSPIRE Metadata Requirement.

[More ...](#)

Metadata Editor

Create metadata according to the INSPIRE implementing rules.

[More ...](#)



About the SemanticLab

The SemanticLab is a server managed by the [Spatial Data Infrastructures Unit of JRC-IES](#) for the purpose of creating, aggregating and deploying thesauri related to geographic information that are made available in the [SKOS format](#), an ontology-based formalism for the encoding of thesauri and knowledge organisation systems in general.

These facilities have been set up in the context of the [GENESIS FP7](#) project for the purpose of grounding semantics-aware annotation and retrieval of geospatial resources.

Sesame RDF triple store

At the core of the services that are provided is an instance of the [Sesame](#) RDF triple store, hosting several repositories organising thesauri in different configurations, depending on the requirements of the applications that are leveraging on them. The repositories can be accessed as SPARQL endpoints over HTTPS.

[SPARQL endpoint](#)

[Sesame Workbench](#)

SKOS Matcher

This prototype application interacts with SPARQL endpoints for the purpose of navigating and interrelating independent SKOS thesauri. The former deployment form is provided on the right and allows to browse the reference thesauri for SDIs that are hosted by the system. It is also available as the matching application.

Browse sample SKOS repository

Vocabulary:

[Back to the Vocabulary](#)

URI:
http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/items/185
PREFERRED LABEL:
en: Addressable object
DEFINITION:
en: a spatial object to which it is meaningful to associate addresses.
DESCRIPTION:
en: Note: Most common addressable objects are real properties, cadastral parcels, buildings, entrances to buildings, dwellings, flats, condominiums/common holds etc., inside a building. Addressable objects can also be other types of sites or constructions like mooring places, points of interest, sports fields, parks, traffic terminals, technical constructions, points of service delivery e.g. utilities, post etc.
TYPE:
http://inspire-registry.jrc.ec.europa.eu/rdfs/schema/inspire-schema.rdf#GeneralTerm http://www.w3.org/2000/01/rdf-schema#Resource http://www.w3.org/2004/02/skos/core#Concept
IN SCHEME:
http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/items

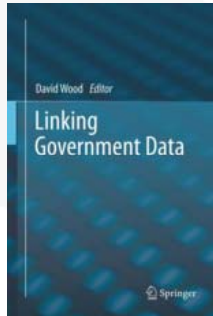
* INSPIRE GeoPortal (to find datasets)
<http://inspire-geoportal.ec.europa.eu/discovery/>

SKOS Glossary for Spatial
Data Infrastructure:
<http://semanticlab.jrc.ec.europa.eu/>



<http://www.flickr.com/photos/wwwworks/4759535950/>

Some references



Wood, David (Ed) Linking Government Data - 2011

Methodological Guidelines for Publishing Government Linked Data

Boris Villazón-Terrazas, Luis M. Vilches, Oscar Corcho, Asunción Gómez-Pérez



Best Practices for Publishing Linked Data

W3C Editor's Draft 31 August 2011

This version:

<http://dvcs.w3.org/hg/gld/bp/>

Latest published version:

<http://www.w3.org/TR/gld/bp/>

Latest editor's draft:

<http://dvcs.w3.org/hg/gld/bp/>

Previous version:

none

Editors:

Michael Hausenblas, DEFI

Bernadette Hyland, 3 Round Stones

Boris Villazón-Terrazas, QEG-UPM

Best Practices for Publishing Linked Data

W3C Editor's Draft – Government Linked Data Working Group

Bernadette Hyland, Boris Villazón-Terrazas, Michael Hausenblas,

<https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

Cookbook for Open Government Linked Data

Editors:
Bernadette Hyland, (3 Round Stones)
Boris Villazón-Terrazas (Universidad Politécnica de Madrid)
Sarven Capadislil (Digital Enterprise Research Institute)

Cookbook for Open Government Linked Data

W3C Editor's Draft – Government Linked Data Working Group

Bernadette Hyland, Boris Villazón-Terrazas, Sarven Capadislil

http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook



Lift your data

Introduction to Linked Data

Ghislain Atemezing¹ , Boris Villazón-Terrazas²

1 EURECOM, France

auguste.atemezing@eurecom.fr

2 Ontology Engineering Group, FI, UPM

bvillazon@fi.upm.es

Slides available at: <http://www.slideshare.net/boricles/>

INSPIRE Conference 2012, Istanbul

Acknowledgements: Oscar Corcho, Asunción Gómez-Pérez, Luis Vilches, Raphaël Troncy, Olaf Hartig, Luis Vilches and many others that we may have omitted.

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