GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
UNIVERSIDAD DE MURCIA
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Unión Europea
Una manera de hacer Europa

Introducción a Linked Open Data (espacios enlazados y enlazables)

Diego López-de-Ipiña
MORElab research group, Universidad de Deusto
dipina@deusto.es

HĒRCULES

MORE lab
ICT for good

gnoss

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Temas a tratar

- ❑ What is linked data
- ❑ What is open data
- ❑ What is the difference between linked and open data
- ❑ How to publish linked data (5-star schema)
- ❑ The economic and social aspects of linked data.

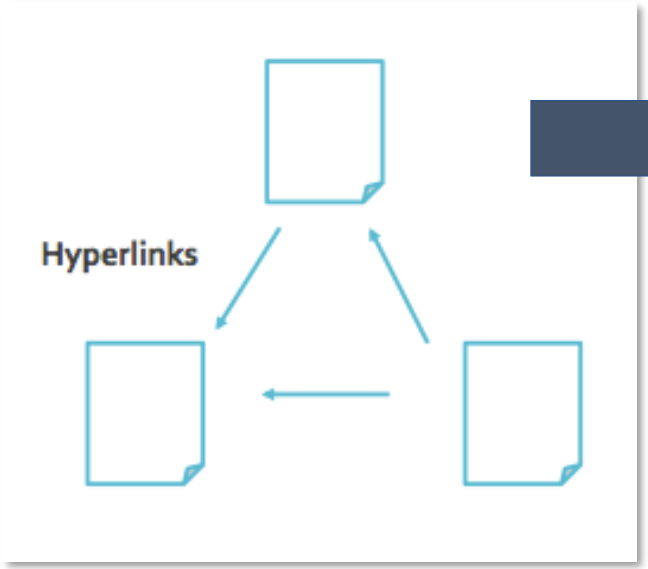
Web of Data: Limitaciones de la Web de Documentos

- **Demasiada información con muy poca estructura** y hecha además **para consumo humano**
  - Es una web sintáctica no semántica
  - La búsqueda de contenidos es muy simplista
    - Se requieren mejores métodos
- Los **contenidos web son heterogéneos**
  - En términos de contenido
  - En términos de estructura
  - En términos de codificación de caracteres
- El futuro requiere integración de **información inteligente**

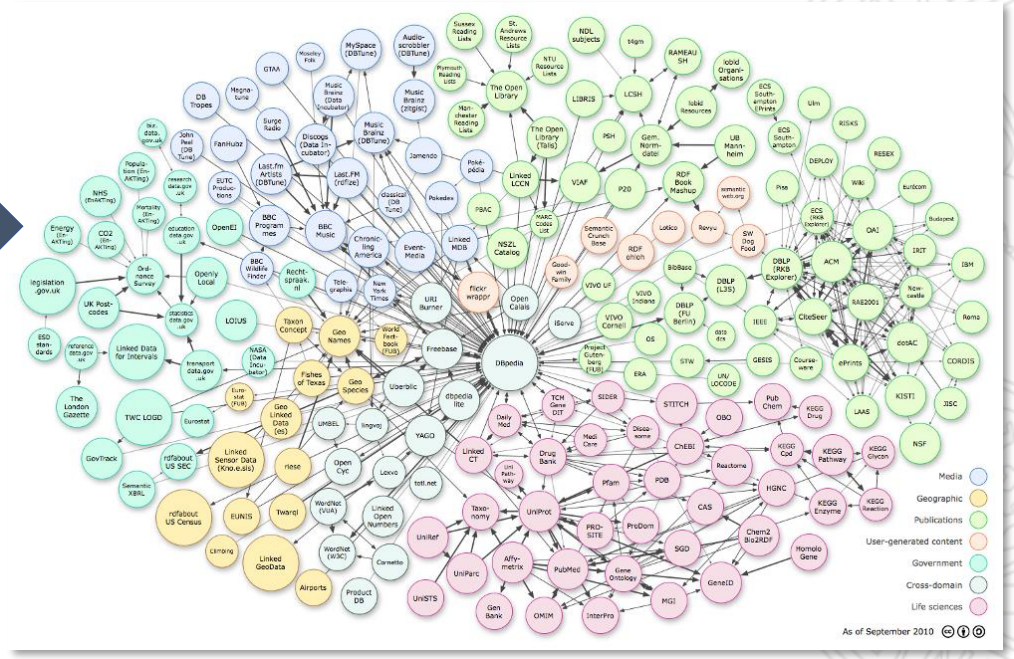
What is Linked Data?

- Evolution from a document-based Web to a Web of interlinked data
- The Web is evolving from a “Web of linked documents” into a “Web of linked data”…

Web of documents…



Web of linked data…



gnoss
Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

The Web is evolving from a "Web of linked documents" into a "Web of linked data"...

- The **Web started as a collection of documents published online** – accessible at Web location identified by a URL.
  - These documents often contain data about real-world resources which is **mainly human-readable and cannot be understood by machines**.
- The **Web of Data** is about in machine-readable **enabling the access to this data**, by making it available formats and connecting it using Uniform Resource Identifiers (URIs), thus **enabling people and machines to collect the data**, and put it together to do all kinds of things with it (permitted by the licence).

- **Machine-readable data (or metadata)** is data in a format that can be interpreted by a computer.
- 2 types of machine-readable data:
  - **human-readable data** that is marked up so that it can also be understood by computers, e.g. microformats, RDFa;
  - **data formats intended principally for computers**, e.g. RDF, XML and JSON.

Democratiando la web semántica: Metadatos empotrados

- Necesitamos que nuestros datos estén preparados para responder adecuadamente a las preguntas de los navegadores y agentes software
  - "Embedded metadata" son datos sobre datos empotrados en una página web que pueden ser extraídos por buscadores y agentes de búsqueda
- Tres opciones principales:
  - RDFa – sistema complejo conectado a XHTML
  - Microformats – ampliamente usado y apoyado, usan etiquetas XHTML antiguas

```
<a href="http://jane-blog.example.org/" rel="sweetheart date met">Jane</a>
```

  - Microdata – más nuevo, soportado por los buscadores, nivel de complejidad intermedio

```
<div itemscope itemtype="http://schema.org/SoftwareApplication">
  <span itemprop="name">Angry Birds</span> -
  REQUIRES <span itemprop="operatingSystem">ANDROID</span><br>
  <link itemprop="applicationCategory" href="http://schema.org/GameApplication"/>
</div>
```

¡¡Todas juntas nos ayudarán a alcanzar la visión de una web con más significado, pero todavía comprensible tanto a humanos como máquinas!!

Democratizando las ontologías: Schema.org

- Initiative launched in 2011 by Bing, Google, Yahoo and then Yandex
- Objective: "**create and support a common set of schemas for structured data mark-up on web pages**."
  - Propose to use their schemas to **annotate contents in a web page with metadata**
    - Metadata are recognized by search engines and other parsers, thus accessing to the "meaning" of portals
    - Their vocabularies were inspired by earlier formats like Microformats, FOAF, GoodRelations and OpenCyc
- Offer schemas in the following domains (http://schema.org/docs/schemas.html):
  - Events, health, organization, person, place, product, offer, revisión and so on.
- To map declarations in microdata to RDF the following tools can be used:
  - http://tools.seochat.com/category/schema-generators
- More info at: http://schema.org/
- Examples:
  - http://schema.org/CreativeWork
  - http://paginaspersonales.deusto.es/dipina/ (microdata.reveal Chrome plugin)

gnoss
Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Defining Linked Data

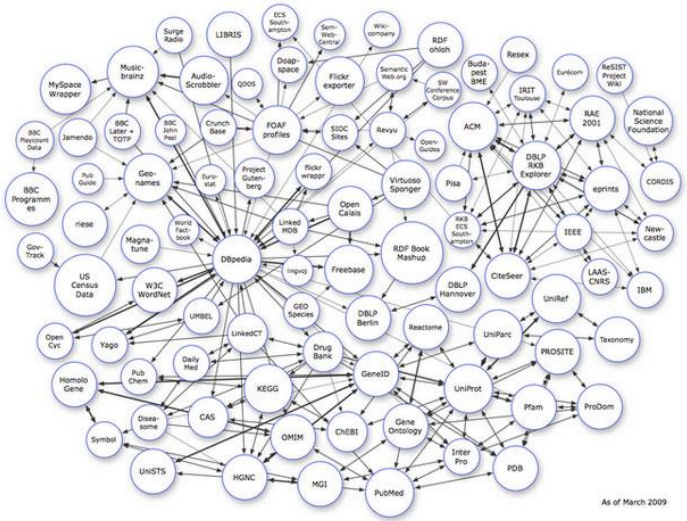
- *"Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens."*
  - *EC ISA Case Study: How Linked Data is transforming eGovernment*

❑ The four design principles of Linked Data (by Tim Berners Lee):

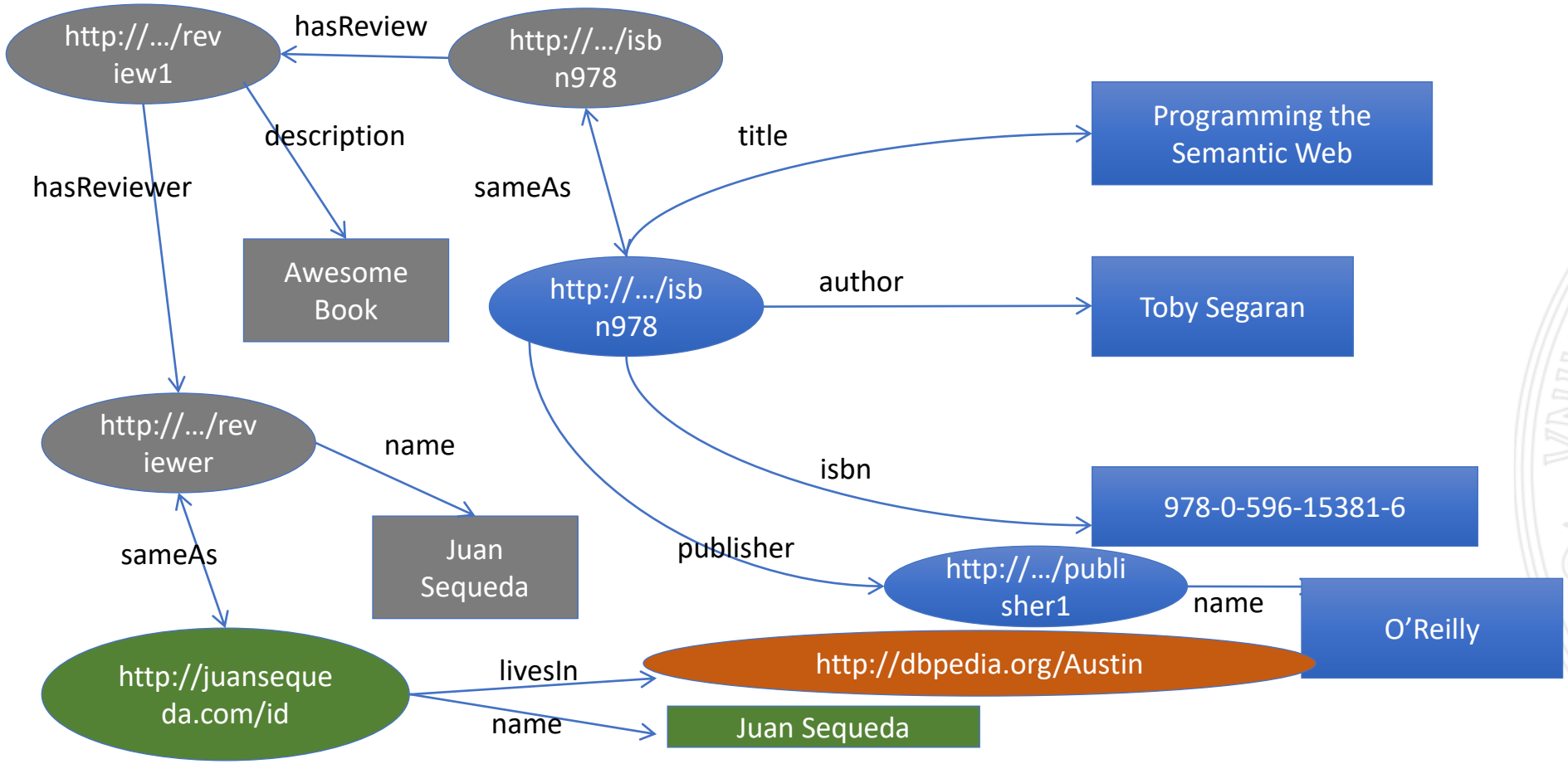
1. Use Uniform Resource Identifiers (URIs) as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs so that they can discover more things.

LinkedData

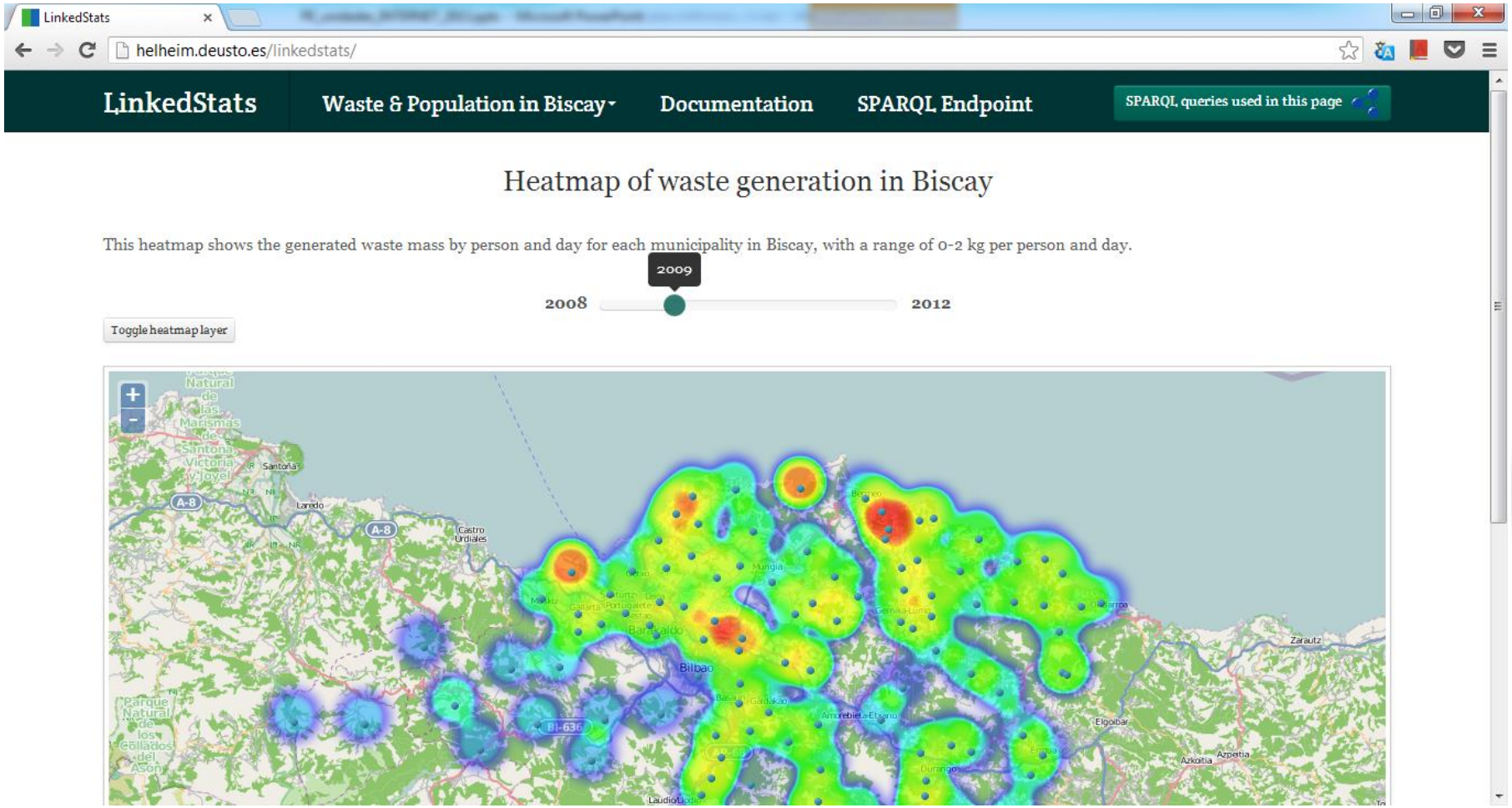
- "A term used to describe a **recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF**."
- Permite **descubrir, conectar, describir y reutilizar todo tipo de datos**.
  - Pasa de una Web de Documentos a una **Web de Datos**
    - En Septiembre 2011 ya contenía 31 billones de tripletas RDF, ligadas por 504millones de enlaces
- Pensado para abrir y conectar diversos vocabularios e instancias semánticas, para que puedan ser utilizados por la comunidad semántica
- URL: https://www.w3.org/standards/semanticweb/data



Ejemplo de Linked Data



Visualizing Linked Data



gnoss
Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Linked (open) government data (LOGD) – value proposition

- **Flexible data integration**: LOGD facilitates data integration and enables the interconnection of previously disparate government datasets.
- **Increase in data quality**: The increased (re)use of LOGD triggers a growing demand to improve data quality. Through crowd-sourcing and self-service mechanisms, errors are progressively corrected.
- **New services**: The availability of LOGD gives rise to new services offered by the public and/or private sector.
- **Cost reduction**: The reuse of LOGD in e-Government applications leads to considerable cost reductions.

- Example:
  - EU Open Data Portal: https://data.europa.eu/euodp/en/home

The four principles in practice... (1)

1. Use Uniform Resource Identifiers (URIs) as names for things.
2. Use HTTP URIs so that people can look up those names.
   - E.g. for an organisation: UNICEF
     http://publications.europa.eu/resource/authority/corporate-body/UNICEF

The four principles in practice... (2)

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.

```
<skos:Concept rdf:about="http://publications.europa.eu/resource/authority/corporate-body/UNICEF"
              at:deprecated="false">
  <skos:inScheme rdf:resource="http://publications.europa.eu/resource/authority/corporate-body"/>
  <skos:broader rdf:resource="http://publications.europa.eu/resource/authority/corporate-body/UNO"/>
  <at:authority-code>UNICEF</at:authority-code>
  <at:op-code>UNICEF</at:op-code>
  <atold:op-code>UNICEF</atold:op-code>
  <dc:identifier>UNICEF</dc:identifier>
  <at:start.use>1951-01-01</at:start.use>
  <skos:prefLabel xml:lang="bg">Унисеф - Детски фонд на ООН</skos:prefLabel>
  <skos:prefLabel xml:lang="cs">UNICEF - Dětský fond Organizace spojených národů</skos:prefLabel>
  <skos:prefLabel xml:lang="da">UNICEF - De Forenede Nationers Børnefond</skos:prefLabel>
  <skos:prefLabel xml:lang="de">Unicef - Kinderhilfswerk der Vereinten Nationen</skos:prefLabel>
  <skos:prefLabel xml:lang="el">Unicef - Ταμείο των Ηνωμένων Εθνών για τα Παιδιά</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Unicef - United Nations Children's Fund</skos:prefLabel>
```

Linked Data vs. Open Data

"Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike."

- OpenDefinition.org

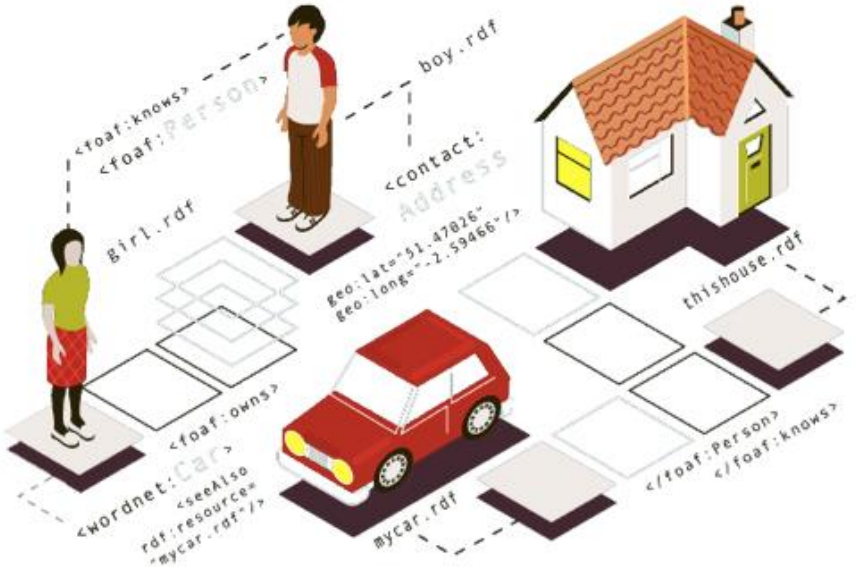
Open data ≠ **Linked data**

Data can be published and be publicly available under an open licence without linking to other data sources.

Data can be linked to URIs from other data sources, using open standards such as RDF without being publicly available under an open licence.

4 reglas de Linked Data

1. Usa URIs para identificar cosas
2. Usa URIs HTTP para que estas cosas puedan ser referenciadas y dereferenciadas por gente y agentes de usuario
3. Proporciona información útil (descripción estructurada y metadatos) sobre la cosa/concepto al que referencia la URI
4. Incluye enlaces a otras URIs para mejorar el descubrimiento de información relacionada en la Web



Linked Data Foundations: URIs

- URIs for naming things, RDF for describing data and SPARQL for querying it
  - "A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource."
    - ISA's 10 Rules for Persistent URIs
- Examples:
  - A country, e.g. Spain
    - http://publications.europa.eu/resource/authority/country/ESP
  - An organisation, e.g. the Publications Office
    - http://publications.europa.eu/resource/authority/corporate-body/PUBL
  - A dataset, e.g. Countries Named Authority List
    - http://publications.europa.eu/resource/authority/country

Linked Data Foundations: RDF & SPARQL

- The **Resource Description Framework** (RDF) is a syntax for representing data and resources in the Web
- **SPARQL** is a standardised language for querying RDF data.
- RDF breaks every piece of information down in **triples**:
  - Subject – a resource, which may be identified with a URI.
  - Predicate – a URI-identified reused specification of the relationship.
  - Object – a resource or literal to which the subject is related.

http://example.org/place/Madrid is the capital of "Spain".

OR

http://example.org/place/Madrid is the capital of http://example.org/place/Madrid.

Subject | Predicate | Object

Manifestaciones de Linked Data

- Los datos publicados como LinkedData puede seguir la siguiente clasificación, según Tim Bernes-Lee:
  - **1 estrella**: datos disponibles en la web (en cualquier formato), pero con una licencia abierta
  - **2 estrellas**: datos disponibles son estructurados y legibles por máquinas. Por ejemplo, Microsoft Excel en vez de una imagen escaneada de una tabla.
  - **3 estrellas**: los datos disponibles como en (2) pero no siguen un formato propietario. Por ejemplo, CSV en vez de Excel.
  - **4 estrellas**: los datos son dispuestos de manera abierta usando un estándar abierto de W3C (RDF y SPARQL) para identificar cosas, de modo que la gente los pueda enlazar.
  - **5 estrellas**: los datos son dispuestos siguiendo lo anterior, incluyendo enlaces externos a los datos de otra gente.

How to publish Linked Data?

- Paving the way towards 5-star linked data

★	Make your stuff available on the Web (whatever format) under an open license.
★★	Make it available as structured data (e.g., Excel instead of image scan of a table)
★★★	Use non-proprietary formats (e.g., CSV instead of Excel)
★★★★	Use URIs to denote things, so that people can point at your stuff
★★★★★	Link your data to other data to provide context

★ Make your stuff available on the web under an open license

The National Archives

Sustainable development targets for 2011-12

Our business plan for 2011–2015 sets out our strategic objectives for the next four years and our specific business priorities for 2011–12. Our aim of 'a more sustainable Kew' sets out actions which will deliver significant sustainability benefits.

1. **Reduce carbon emissions at Kew by 6% from 2010–11 levels, balancing record preservation and environmental conditions**

Period	Electricity (KWh)	Gas (KWh)	Carbon (tonnes)	Change on 2010–11 (%)
April	762,625	354,062	479	-8.6
May	757,291	348,324	475	-6.5
June	846,364	388,369	530	-13.0
July	908,864	338,278	555	-17.1
August	928,827	384,925	574	-6.1
September	868,526	463,960	556	-2.7
October	810,768	376,137	509	-11.0
November	697,957	439,482	459	-17.1
December	536,080	472,718	378	-24.9

Performance to the end of December 2011 is -11.6%, well ahead of target. Our long-term commitment, which we are on track to meet or exceed, is to reduce emissions from buildings and business-related travel by 25% from 2009–10 levels by April 201

gnoss

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Pros & Cons of ★ Open Data

As a consumer...	As a publisher...
✓You can access the data.	✓It is simple to publish.
✓You can store it locally.	✓You do not have explain repeatedly to others that they can use your data.
✓You can enter the data into any other system.	
✓You can change the data.	
✓You can share the data with anyone.	

★ ★ Make it available as structured data

Table DA2301 (SST2.10): Security and fire safety - dwellings, 2010

all dwellings

	smoke alarm*	burglar alarm	door viewer	external lighting	secure windows and doors	**all dwellings in group (000s)**	*sample size (unweighted)*
	percentage of dwellings within group						
tenure							
owner occupied	-	36,9	51,9	63,3	77,3	**14.860**	*8.791*
private rented	-	20,0	48,5	53,2	66,0	**3.706**	*3.096*
local authority	-	11,9	67,4	60,9	76,7	**1.801**	*2.276*
housing association	-	11,9	75,3	68,0	78,7	**2.018**	*2.507*
all private	-	33,6	51,2	61,3	75,1	**18.567**	*11.887*
all social	-	11,9	71,6	64,7	77,8	**3.819**	*4.783*
dwelling age							
pre-1919	-	25,4	44,3	41,9	58,4	**4.865**	*3.249*
1919-44	-	33,1	51,1	54,9	72,3	**3.751**	*2.684*
1945-64	-	27,2	54,3	60,2	79,6	**4.397**	*3.609*
1965-80	-	26,0	56,6	67,8	81,8	**4.602**	*3.593*
1981-90	-	31,2	57,9	77,5	78,6	**1.880**	*1.429*
post 1990	-	42,6	72,2	87,5	90,3	**2.892**	*2.106*
dwelling type							
end terrace	-	28,6	51,9	51,3	75,3	**2.251**	
mid terrace	-	24,4	49,6	40,4	72,3	**4.105**	
small terraced house	-	22,1	49,5	42,1	71,7	**2.171**	
medium/large terraced house	-	27,8	51,0	45,4	74,3	**4.185**	

gnoss

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto

Pros & Cons of ★ ★ Open Data

- ❑ All the benefits of ★ open data; plus

As a consumer...	As a publisher...
✓You can directly process it with proprietary software to aggregate it, perform calculations, visualise it, etc.	✓It is still simple to publish.
✓You can export it into another (structured and/or non proprietary) format.	

★ ★ ★ Use non-proprietary formats

- ❑ Proprietary: Excel, Word, PDF...
- ❑ Non-proprietary: XML, CSV, RDF, JSON, ODF...

- ❑ Road safety- Accidents 2006:

```
Acc_Index,Vehicle_Reference,Casualty_Reference,Casualty_Class,Sex_of_Ca
ge_Band_of_Casualty,Casualty_Severity,Pedestrian_Location,Pedestrian_Mo
ar_Passenger,Bus_or_Coach_Passenger,Pedestrian_Road_Maintenance_Worker,
_Type,Casualty_Home_Area_Type
200601BS70001,1,1,1,1,6,3,0,0,0,0,-1,4,1
200601BS70002,1,1,1,1,7,2,0,0,0,0,-1,3,1
200601BS70002,1,2,3,1,6,3,1,1,0,0,-1,0,1
200601BS70003,2,1,1,1,9,3,0,0,0,0,-1,9,1
200601BS70005,1,1,3,2,5,3,5,3,0,0,-1,0,1
200601BS70006,2,1,1,1,7,3,0,0,0,0,-1,3,1
200601BS70007,1,1,3,2,10,3,5,9,0,0,-1,0,3
200601BS70009,1,1,3,1,11,3,5,1,0,0,-1,0,-1
200601BS70010,2,1,1,1,8,2,0,0,0,0,-1,1,1
200601BS70013,1,1,1,1,7,3,0,0,0,0,-1,9,1
200601BS70015,1,1,3,2,10,2,1,1,0,0,-1,0,1
200601BS70017,1,1,1,1,7,3,0,0,0,0,-1,5,1
200601BS70018,1,1,1,2,6,3,0,0,0,0,-1,9,1
200601BS70019,2,1,1,2,5,3,0,0,0,0,-1,1,1
200601BS70020,2,1,1,1,6,2,0,0,0,0,-1,3,1
200601BS70021,1,1,1,1,8,3,0,0,0,0,-1,3,1
200601BS70021,1,2,2,1,2,3,0,0,0,0,-1,3,-1
200601BS70022,1,1,1,1,5,3,0,0,0,0,-1,5,1
200601BS70023,2,1,1,1,6,3,0,0,0,0,-1,5,1
200601BS70024,1,1,3,2,5,3,1,3,0,0,-1,0,1
200601BS70025,1,1,1,2,9,3,0,0,0,0,-1,9,1
200601BS70027,1,1,3,1,6,2,1,1,0,0,-1,0,-1
200601BS70028,1,1,1,2,7,3,0,0,0,0,-1,9,1
200601BS70029,2,1,1,1,5,2,0,0,0,0,-1,3,-1
200601BS70030,1,1,3,2,6,3,4,1,0,0,-1,0,1
200601BS70031,2,1,1,1,6,2,0,0,0,0,-1,5,1
200601BS70033,2,1,1,1,6,3,0,0,0,0,-1,3,-1
200601BS70034,1,1,3,1,5,3,4,3,0,0,-1,0,1
200601BS70035,1,1,1,1,6,3,0,0,0,0,-1,1,1
200601BS70036,1,1,1,2,6,2,0,0,0,0,-1,9,1
200601BS70037,1,1,1,1,8,3,0,0,0,0,-1,9,1
200601BS70037,1,2,2,1,-1,3,0,0,1,0,-1,9,-1
200601BS70038,1,1,1,1,6,3,0,0,0,0,-1,5,1
200601BS70039,1,1,1,1,7,3,0,0,0,0,-1,9,1
200601BS70040,1,1,3,2,11,2,4,1,0,0,-1,0,1
200601BS70041,1,1,1,2,6,3,0,0,0,0,-1,9,1
200601BS70042,2,1,1,1,7,2,0,0,0,0,-1,1,1
200601BS70043,1,1,1,1,5,3,0,0,0,0,-1,2,1
200601BS70044,2,1,1,2,9,3,0,0,0,0,-1,9,1
200601BS70045,1,1,3,2,2,3,1,1,0,0,-1,0,1
200601BS70046,2,1,1,2,6,3,0,0,0,0,-1,3,1
200601BS70047,1,1,3,2,3,2,9,5,0,0,-1,0,1
200601BS70048,1,1,3,1,2,3,5,2,0,0,-1,0,1
200601BS70050,1,1,1,1,8,3,0,0,0,0,-1,9,1
200601BS70051,1,1,1,1,10,3,0,0,0,0,-1,9,-1
200601BS70052,1,1,1,1,5,3,0,0,0,0,-1,2,1
200601BS70053,2,1,1,1,8,3,0,0,0,0,-1,9,1
```

Pros & Cons of ★ ★ ★ Open Data

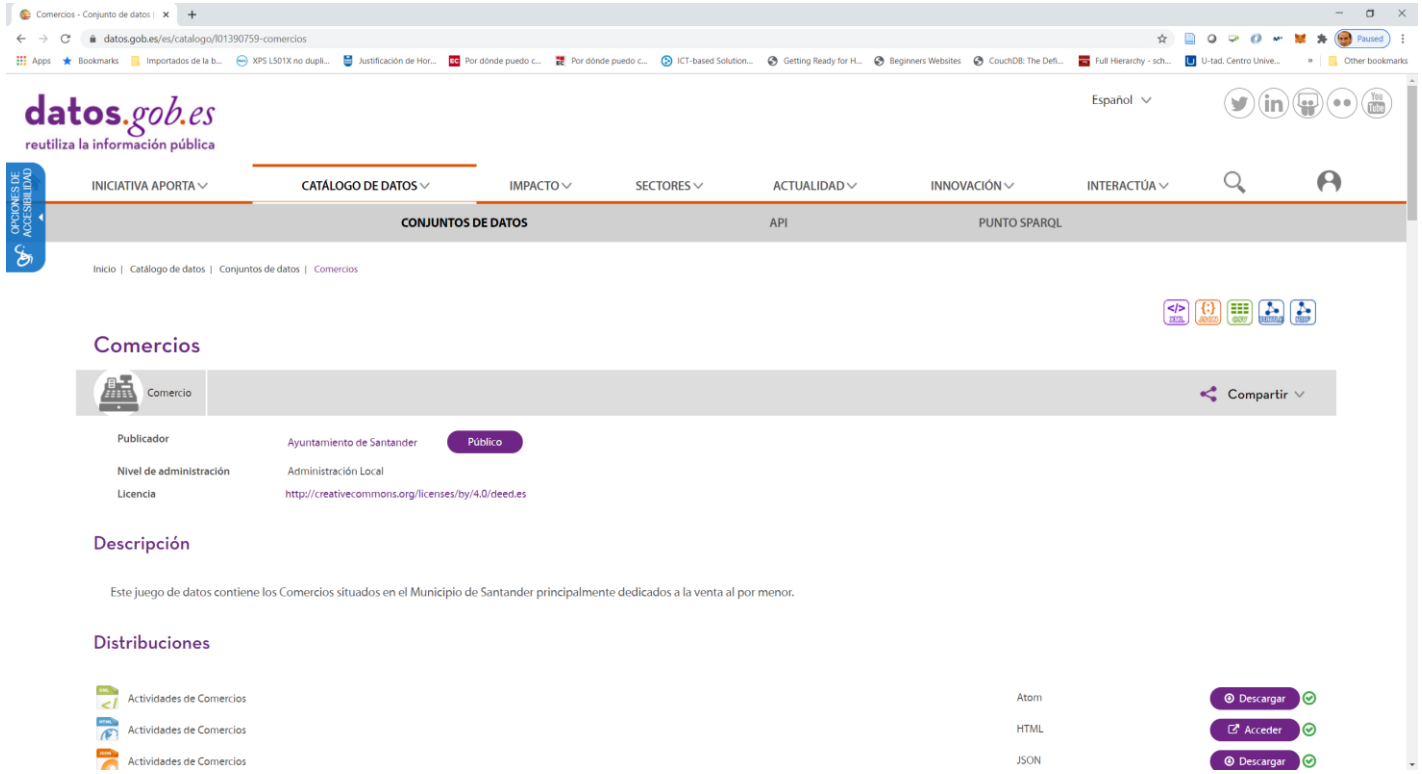
- All the benefits of ★ ★ open data; plus

As a consumer...	As a publisher...
✓You can manipulate the data in any way you like, without being confined by the capabilities of any particular software.	✓It is still simple to publish.
	- But, you may need converters or plug-informat.s to export the data from the proprietary

★ ★ ★ ★ Use URIs to denote things

- Por ejemplo, Comercios situados en el Municipio de Santander principalmente dedicados a la venta al por menor.

https://datos.gob.es/es/catalogo/l01390759-comercios



Pros & Cons of ★ ★ ★ ★ Open Data

❑ All the benefits of ★ ★ ★ ★ open data; plus

As a consumer...	As a publisher...
✓You can link to it from any other place.	✓Other data publishers can now link into your data, promoting it to 5 star.
✓You can bookmark it.	✓You will be able to reuse vocabularies, data and metadata, and URI design patterns instead of creating them from scratch.
✓You can access information about a particular resource directly through its URI, without having to download the complete dataset.	
✓You may be able to reuse existing tools and libraries.	- But you typically need to invest some time in identifying the resources and assigning URIs.
✓You can combine the data with other data.	- You need to invest in a stable policy, management and infrastructure for persistent URIs.
- But understanding the technology requires effort and can have a steep learning curve.	

★ ★ ★ ★ ★ Link your data to other data to provide context

About: Office of the Deputy Minister for Administrative Reform and e-governance
An Entity of Type : Office,

References | Referenced By

type	http://org.testproject.eu/mareg/def/orgunit/Office
preferred label	Office of the Deputy Minister for Administrative Reform and e-governance Γραφείο Υφυπουργού Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
hasUnit	Office of the Secretary General for Administrative Reform and e-governance Managing Authority of the Operational Programme "Administrative Reform 2007-2013"

About: Office of the Secretary General for Administrative Reform and e-governance
An Entity of Type : Office,

References | Referenced By

type	http://org.testproject.eu/mareg/def/orgunit/Office
preferred label	Office of the Secretary General for Administrative Reform and e-governance Γραφείο Γενικού Γραμματέα Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
hasUnit	Directorate General of Financial and Administrative Services Directorate General of Administrative Reform and e-Governance Directorate General of Human Resources Management

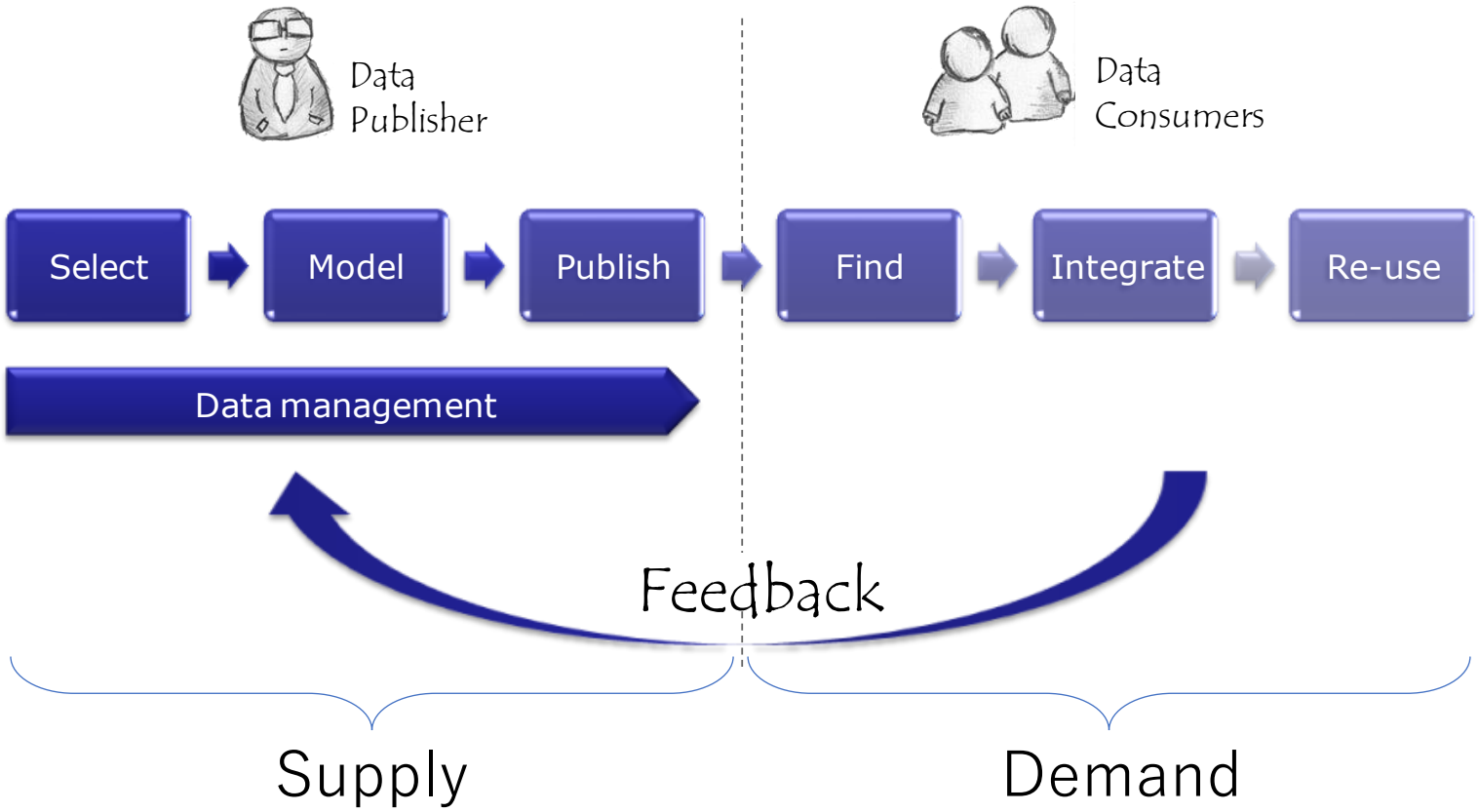
Pros & Cons of ★ ★ ★ ★ Open Data

- ❑ All the benefits of ★ ★ ★ ★ open data; plus

As a consumer...	As a publisher...
✓You can discover more (related) data while consuming the data.	✓You make your data discoverable.
✓You can directly learn about the data schema.	✓You increase the context, expressivity, quality and value of your data (and consequently you give visibility to your organisation).
✓You can combine data from different source, be innovative, gain new knowledge, be an entrepreneur...	- This requires an investment in time, money, technology and competencies/ skills.
- But, you now have to deal with broken data links. Not all publishers/data sources will be reliable.	

Prepare your data for publishing – LOD lifecycle

- LOGD and metadata lifecycle focusing on supply and demand



Selection of high-value data

- Several dimensions can be considered in the selection process of Linked Open Government Data, both from the publisher's and the re-user's point of view:
  - **Transparency**: Does the publication of the dataset increase transparency and openness of the government towards its citizens?
  - **Legal requirements**: Is there a law that makes open publication mandatory or is there no specific obligation?
  - **Relation to public task**: Is the data the direct result of the primary public task of government or is it a product of a non-essential activity?
  - **Current status of open publication**: Is the data already openly available or does it still need to be opened up?
  - **Type of value**: Is the data useful for social engagement or does it have commercial value?
  - **Audience**: Is the data primarily intended for the public or is it primarily aimed at back-office integration?

Modelling your data & metadata is about ...

- Making your data available in a structured, comprehensible and machine-readable way.
- Reusing what already exists in terms of vocabularies and reference data.
- Reaching the right quality level by cleansing your data.
- Providing licensing information so that data consumers know what the conditions of reuse are.
- Providing a rich description (metadata).
- Using semantic technologies (RDF, HTTP URIs...) for describing your data.

Cleansing your data & metadata

- To ensure data and metadata can be published with an appropriate level of quality and minimum errors.
  - This means:
    - Fixing errors.
    - Transforming/homogenising formats.
    - Aligning inconsistencies in data and metadata.
    - Removing duplicate/redundant information.
    - Adding lacking information.
    - Making sure the information is up-to-date.

OpenRefine tool

- Portal: https://openrefine.org/
- Video demostrador:
  - https://www.youtube.com/watch?v=tzXExfZCA1w
- GitHub: https://github.com/OpenRefine/OpenRefine

OpenRefine

Cleansing your data – example

Duplicate

Formatting issue

error

Company_name	Registration date	Country	E-mail	# Establishments
Nikè	1991-04-28	Belgium	niké	7
BARCO	15 September 1986	BE	Barco@email.be	2
Nikè		België		
Coca-Cola		United States	coca@cola.com	3

Cleansing operations

Missing information

Inconsistent information

Redundant information

Company_name	Registration date	Country	E-mail
Nikè	1991-04-28	BE	niké@sport.org
BARCO	1986-09-05	BE	Barco@email.be
Coca-Cola	1964-03-26	US	coca@cola.com

Publishing linked data is about ...

- **Breaking down the walls of the silos in order to create more value**.
  - Making your data and metadata publicly and easily accessible on the Web.
  - Linking your data and metadata to other data (or metadata) in order to:
    - Attach meaning and content to it.
    - Give context to it.
    - Enrich it.
    - Allow people to discover more.

Ejemplos de fuentes de datos públicas reutilizables

- Datos Abiertos de Zaragoza https://www.zaragoza.es/sede/portal/datos-abiertos/api
- Dades Obertes Manlleu: https://dadesobertes.diba.cat/dades-obertes/documentacio-tecnica/api
- Aemet OpenData API: https://opendata.aemet.es/
- Catálogo de APIs Abiertas ISTAC: https://www3.gobiernodecanarias.org/aplicaciones/appsistac/api
- EMT mobilitylabs: https://mobilitylabs.emtmadrid.es/es/portal/opendata

WikiData & DBpedia

- **Wikidata** is a volunteer-created knowledge base of structured data that anyone can edit
  - **Focused on structured data**: possible for humans and computers alike to use the data
  - Many ways to contribute to Wikidata: translate, write apps, add and edit data.
  - It works with:
    - **Items** – abstract concepts with theirs own and a unique identifier (Q###) and optionally a label, description and aliases
    - **Statements** are added to items: category of data as a property, while the data that describes an item for a given property is known as a value.
  - Example: entry for Everest mountain https://www.wikidata.org/wiki/Q513
  - Documentation: https://www.wikidata.org/wiki/Wikidata:Tours
  - Wikidata query service: https://query.wikidata.org/
- **DBpedia**, a project to create a **graph from Wikipedia data** – allows users to semantically query relationships and properties associated with Wikipedia resources, including links to other related datasets
  - Wikipedia articles consist mostly of free text, but also include structured information embedded in the articles, such as **"infobox" tables**, categorisation information, images, geo-coordinates and links to external Web pages.
    - This structured information is extracted and put in a uniform dataset which can be queried: http://live.dbpedia.org/sparql

WikiData & DBpedia

- There are **4 main differences**:
  - **Wikidata provides data to Wikipedia**, while DBpedia extracts data from Wikipedia.
  - **Wikidata's ontology is community curated**, and part of the data maintained on the site, while DBpedia's ontology is statically defined, and much stricter.
  - Formally, **Wikidata only asserts statements** (who claims what), while DBpedia asserts facts, often causing contradictions.
  - Wikidata is licensed CC-0, and is this re-usable without any restrictions, while DBpedia is licensed CC-BY-SA, which requires author attribution - which is a good thing generally, but impractical for a knowledge base automatically derived from text.
- More info: https://www.quora.com/How-is-Wikidata-related-to-Wikipedia-in-a-way-different-from-how-DBpedia-is-related-to-Wikipedia
- Examples:
  - Listar los nombres en castellano de los países en Dbpedia
  - Countries sorted by population in WikiData

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?nombre WHERE {
   ?pais rdf:type dbo:Country .
   ?pais rdfs:label ?nombre .
   FILTER (lang(?nombre)='es')
}
```

Iniciativa Aporta – https://datos.gob.es/es/acerca-de-la-iniciativa-aporta

- La **Iniciativa Aporta es la estrategia nacional de coordinación y de impulso de la apertura de datos** procedentes del sector público, y de promoción del desarrollo de servicios avanzados basados ellos.
- **Objetivos:**
  - Impulsar y coordinar la apertura de los datos generados por el sector público.
  - Estimular un mercado ligado a la reutilización de la información del sector público.
  - Contribuir a favorecer las condiciones para el desarrollo de la Estrategia europea de datos en España.

https://datos.gob.es/ Punto de encuentro entre las administraciones, las empresas y los ciudadanos que forman parte del ecosistema de datos abiertos en España

LOS DATOS, MOTOR DEL MUNDO

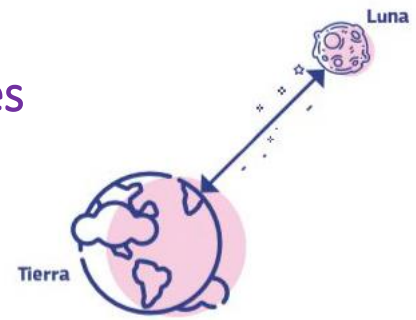
Cada año crece el volumen de datos en el mundo.

Unos datos generados tanto por empresas, como por el conjunto de la sociedad.

CRECERÁ EL VOLUMEN GLOBAL DE DATOS:

2018

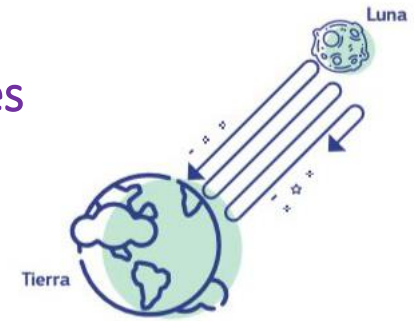
33 zettabytes



almacenados en tabletas de 512 GB, formarían una torre que llegaría a la luna.

2025 5 veces más

175 zettabytes



bastarían para hacer cinco veces el camino de ida y vuelta a la luna.

Estos datos son un elemento fundamental para facilitar la toma de decisiones y generar valor económico y social.

En 2018, el valor de la economía de datos superó los 300 mil millones de euros en la UE28, con un crecimiento del 12% con respecto al año anterior.*

Fuente: *Unión Europea, 2020*

**Fuente:* DATA ASTHE ENGINE OF EUROPE'S DIGITAL FUTURE IDC, 2019*

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
UNIVERSIDAD DE MURCIA
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Unión Europea
Una manera de hacer Europa

EL USO DE LOS DATOS EN LAS EMPRESAS

- **Los datos abiertos son la materia prima de nuevos servicios, productos y aplicaciones.**
- Las empresas infomediarias crean productos y servicios de valor añadido con datos públicos y privados. En España, el sector generó **2.000 millones de euros** y empleó a más de **15.000 personas** en 2018 *

- **+ 15,4%**
  - **Volumen de negocio (2015-2018)**
- **+ 14,3%**
  - **Empleados (2016-2018)**

EJEMPLOS

Agricultura:
datos del suelo o del tiempo para optimizar el riego

Turismo:
análisis de tendencias de viajeros

Energía:
medidores inteligentes que tienen en cuenta las condiciones atmosféricas

Sector Inmobiliario:
análisis de precios del mercado y condiciones de las distintas zonas

**Fuente: informe "Del sector infomediario a la economía del dato. Parte I. Caracterización del Sector Infomediario", ONTSI .*

Conclusions

- Linked data is a set of design principles for sharing machine-readable data on the Web.
- Linked data and open data are not the same.
- URIs, RDF and SPARQL form the foundational layer for Linked data.
- Linked data offers a number of advantages for:
  - Data integration with small impact on legacy systems;
  - Enables for semantic interoperability;
  - Enables creativity and innovation through context and knowledge-creation.

Referencias

- Iniciativa Aporta: https://datos.gob.es/es/acerca-de-la-iniciativa-aporta
  - Presentación: http://ondemand2.redes.ondemand.flumotion.com/redes/ondemand2/EX-24956/PresentacionMultimedia_07.09.20.pptx
- European Data Portal – Introduction to Linked Data, training module: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_1.2_introduction_to_linked_data_en_edp.pdf
- Guía práctica para la publicación de Datos Abiertos usando APIs – Iniciativa Aporta
  - https://datos.gob.es/es/documentacion/guia-practica-para-la-publicacion-de-datos-abiertos-usando-apis

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
UNIVERSIDAD DE MURCIA
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Unión Europea
Una manera de hacer Europa

Red Ontologías Hércules – ROH

Diego López-de-Ipiña
MORElab research group, Universidad de Deusto
dipina@deusto.es

HĒRCULES

MORE lab
ICT for good

gnoss

Deusto
Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto