



OPEN DATA SUPPORT

Training Module 1.2

Introduction to Linked Data



PwC firms help organisations and individuals create the value they're looking for. We're a network of firms in 158 countries with close to 180,000 people who are committed to delivering quality in assurance, tax and advisory services. Tell us what matters to you and find out more by visiting us at www.pwc.com. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

Presentation metadata

Open Data Support is funded by the European Commission under SMART 2012/0107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals'(Contract No. 30-CE-0530965/00-17).

© 2014 European Commission

This presentation has been created by PwC

Authors:

Michiel De Keyzer, Nikolaos Loutas, Christophe Colas and Stijn Goedertier

Disclaimers

1. The views expressed in this presentation are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this presentation, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.
2. This presentation has been carefully compiled by PwC, but no representation is made or warranty given (either express or implied) as to the completeness or accuracy of the information it contains. PwC is not liable for the information in this presentation or any decision or consequence based on the use of it.. PwC will not be liable for any damages arising from the use of the information contained in this presentation. The information contained in this presentation is of a general nature and is solely for guidance on matters of general interest. This presentation is not a substitute for professional advice on any particular matter. No reader should act on the basis of any matter contained in this publication without considering appropriate professional advice.

Learning objectives

By the end of this training module you should have an understanding of:

- 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.

Content

This module contains ...

- An introduction to the linked data principles;
- An introduction to linked data technologies;
- An outline of the 5-star scheme for publishing linked data;
- An example of how tabular data can be published as linked data using Open Refine;
- The expected benefits of linked data for governments;
- An overview of linked data initiatives in Europe.

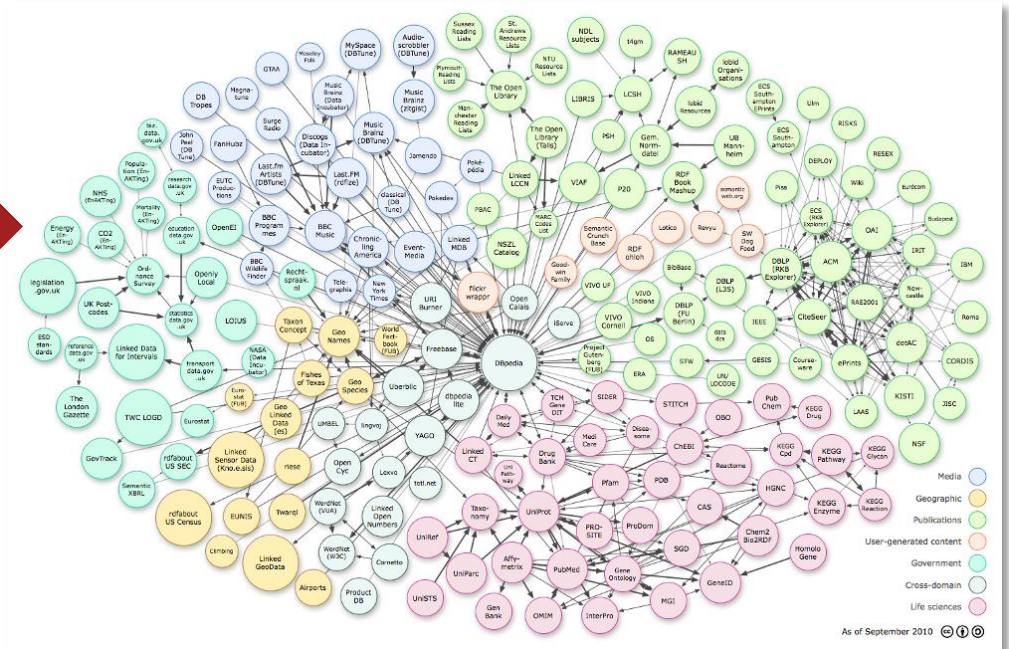
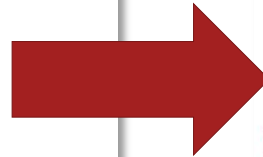
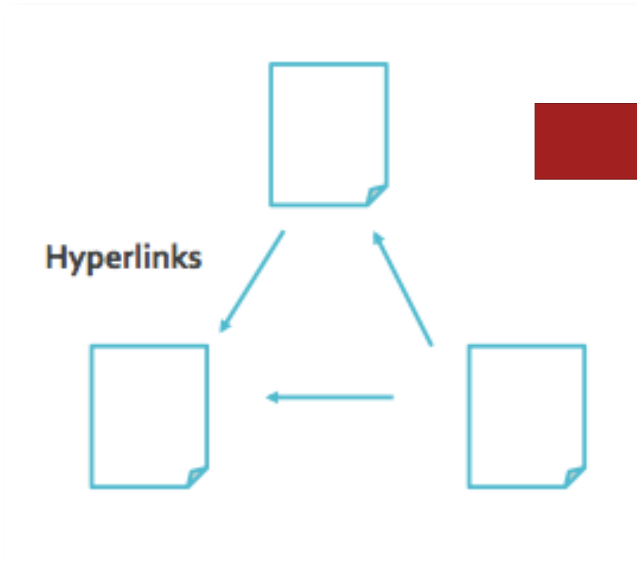
What is linked data?

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”... (1/2)

Web of documents...

Web of linked data...



The Web is evolving from a “Web of linked documents” into a “Web of linked data”... (2/2)

- 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 enabling the access to this data, by making it available in machine-readable 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.

See also:

http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html
<http://linkeddatabook.com/editions/1.0/>

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.

See also:

http://www.youtube.com/watch?v=4x_xzT5eF5Q
<http://www.w3.org/DesignIssues/LinkedData.html>
<http://www.youtube.com/watch?v=uju4wT9uBIA>

Linked (open) government data – 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.

See also:

ISA Study on Business Models for LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

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">Унищф - Детски фонд на ООН</skos:prefLabel>  
  <skos:prefLabel xml:lang="cs">UNICEF - Dětšký 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

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



Linked data

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

See also:

Cobden et al., A research agenda for Linked Closed Data
http://ceur-ws.org/Vol-782/CobdenEtAl_COLD2011.pdf

Linked data foundations

URIs for naming things, RDF for describing data and SPARQL for querying it.

Uniform Resource Identifier (URI)

“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

A country, e.g. Belgium

- <http://publications.europa.eu/resource/authority/country/BEL>



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/>



See also:

<http://www.slideshare.net/OpenDataSupport/design-and-manage-persitent-uris>

RDF & SPARQL

The **Resource Description Framework** (RDF) is a syntax for representing data and resources in the Web

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/Brussels> is the capital of “Belgium”.

OR

<http://example.org/place/Brussels> is the capital of <http://example.org/place/Belgium>.

Subject

Predicate

Object

SPARQL is a standardised language for querying RDF data.

How to publish linked data?

Paving the way towards 5-star linked data

5 star-schema of Linked (Open) Data

- ★ Make your stuff available on the Web (whatever format) under an open license. optional
- ★★ 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 licence*



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 2015.

Licensed under **OGI** Open Government Licence.

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)
	<i>percentage of dwellings within group</i>						
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	



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-ins to export the data from the proprietary format.

★ ★ ★ ★ *Use URIs to denote things*

For example, creating an URI for one of the units of the Greek Ministry of the Administrative Reform and e-Governance.



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Υπουργείο Διοικητικής Μεταρρύθμισης
και Ηλεκτρονικής Διακυβέρνησης

ΥΠΟ <http://data.ydmed.gov.gr/doc/organization/16180>

Type Organization

Raw data [HTML](#) | [RDF/XML](#) | [Turtle](#)

Category ΥΠΟΥΡΓΕΙΑ

See also:

<http://www.slideshare.net/OpenDataSupport/design-and-manage-persitent-uris>

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	<ul style="list-style-type: none">http://org.testproject.eu/mareg/def/orgunit/Office
preferred label	<ul style="list-style-type: none">Office of the Deputy Minister for Administrative Reform and e-governanceΓραφείο Υφυπουργού Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
hasUnit	<ul style="list-style-type: none">Office of the Secretary General for Administrative Reform and e-governanceManaging 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	<ul style="list-style-type: none">http://org.testproject.eu/mareg/def/orgunit/Office
preferred label	<ul style="list-style-type: none">Office of the Secretary General for Administrative Reform and e-governanceΓραφείο Γενικού Γραμματέα Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
hasUnit	<ul style="list-style-type: none">Directorate General of Financial and Administrative ServicesDirectorate General of Administrative Reform and e-GovernanceDirectorate General of Human Resources Management

Pros & cons of ★ ★ ★ ★ ★ open data

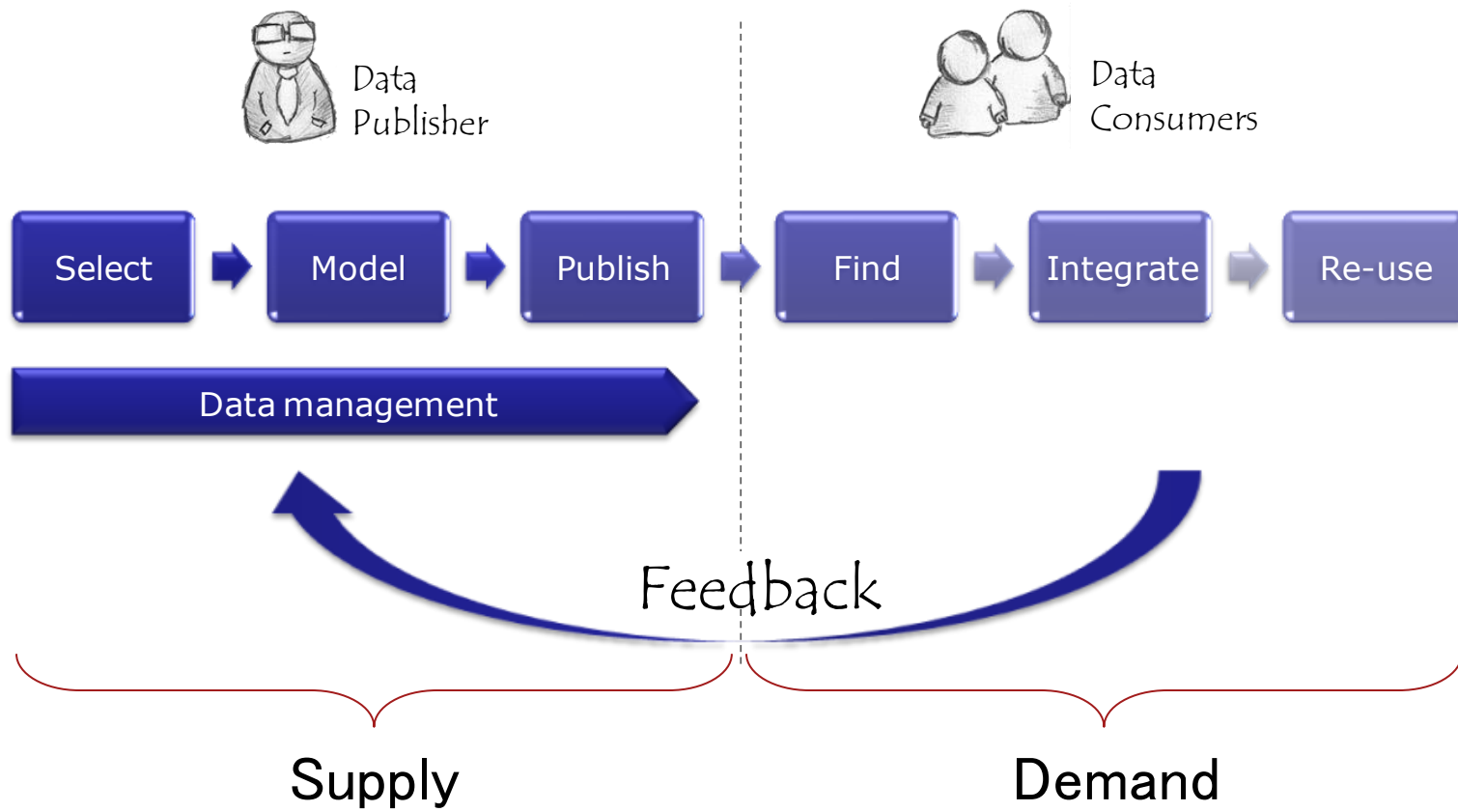
All the benefits of ★ ★ ★ ★ open data; **plus**

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

The LOGD 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.

See also:

<http://www.slideshare.net/OpenDataSupport/introduction-to-rdf-sparql>

Cleanse your data with Open Refine (Google Refine) -

<https://code.google.com/p/google-refine/>

Cleansing 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 publically 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.

Collecting feedback from the reusers of your data

Ask feedback to the (potential) users of data:

- Which data do they need.
- How did they use the data.
- What did they think about the quality.
- Make sure that requests and fixes reach you – crowdsource data quality!

Behoeftte aan meer data?

Via onderstaand formulier kunt u een verzoek om overheidsdata insturen. We zullen dan bemiddelen tussen u en de overheidsinstantie van wie u data zou willen hebben. We kunnen geen garantie geven dat een aanvraag ook leidt tot het beschikbaar komen van data.

Voor vragen kunt u contact opnemen met paul.suijkerbuijk@minbzk.nl.

We houden ons aan de [e-mailgedragslijn voor overheden](#) van burger@overheid.nl. Lees hier hoe Overheid.nl met uw [persoonsgegevens](#) omgaat. Velden aangegeven met een * zijn verplicht.

Uw naam *

Uw organisatie *

data.overheid.nl



data.gov.uk

Example

Using Open Refine for RDF to publish tabular data as Linked Data.

What is Open Refine RDF extension



Open Refine RDF extension, allows you to easily import data in different formats such as :

- CSV;
- Excel(.xls and .xlsx);
- JSON;
- XML; and
- RDF/XML.

And then determine the intended structure of an RDF dataset, by drawing a template graph.

See also:

LOD 2 Webinar – Open Refine

<http://www.youtube.com/watch?v=4Ve93C238gl>

Case study: Linking data about plant protection products

We will show how a [dataset](#) of the Greek [Ministry of Rural Development and Food](#) was described using an [ontology](#) developed by [DG Health and Consumers](#) and was then published as Linked Data.

The dataset was in CSV format.

The screenshot shows a web interface titled "Linking data about applications and decisions for authorisation of plant protection products". It features a search bar with the keyword "Acanto" and a "Search" button. Below the search bar is a SPARQL query editor containing the following query:

```
SPARQL Query:
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ppp: <http://ec.europa.eu/open-data/planthealth/ppp/>
SELECT DISTINCT ?s ?p ?o
FROM <http://health.testproject.eu/ppp/>
WHERE {
  ?s a ppp:Product;
  ?p ?o.
  FILTER (regex(?s, 'Acanto', 'i')).
}
LIMIT 100
```

Below the query editor is a "Run Query" button and a link to "Search using the Faceted Browser". On the left side of the interface, there are sections for "PPP semantic asset" (listing "PPP Ontology" and "PPP Taxonomies"), "Sample queries" (with several links to find products and substances), and "Find out more about Linked Data" (with links to understand linked data, case studies, and rules for persistent URIs).

At the bottom of the interface, there is a footer with the text: "This work is supported by Action 1.1 of the interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission." Below this, there are logos for various data pilots: DCAT, ADMS, ADMS SW, CORE BUSINESS VOCABULARY, CORE PERSON VOCABULARY, CORE LOCATION VOCABULARY, CORE PUBLIC SERVICE VOCABULARY, and isa. To the right of these logos is the logo for the Directorate-General for Health & Consumers.

<http://health.testproject.eu/PPP/>

See also:

http://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp

Creating the project in Open Refine

- Make sure that Open Refine and the RDF extension are installed on your machine.
- Launch Open Refine.
- Upload the spreadsheet and selected the sheets that you want.
- Confirm the creation of the project.

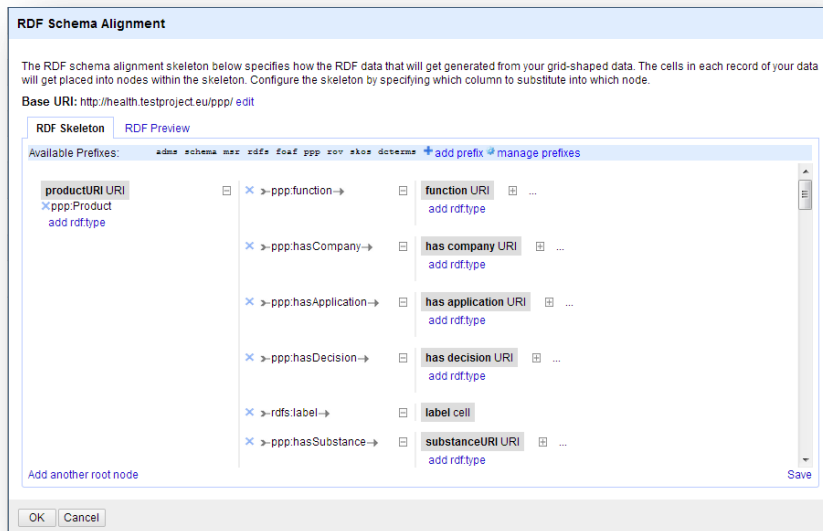
The screenshot shows the Open Refine web interface in a browser window. The address bar shows the URL 127.0.0.1:3333. The page title is "Google Refine" and the subtitle is "A power tool for working with messy data." The interface is divided into several sections:

- Left sidebar:** Contains "Create Project" and "Import Project" buttons.
- Main table:** A table with 12 rows and 4 columns: "productURI", "label", "trade name", and "function". The rows contain various product identifiers and their corresponding labels and functions.
- Bottom section:** A "Parse data as" panel with several options:
 - Excel (.xlsx) files:** Selected.
 - Worksheets to Import:** "Product" (6674 rows), "Formulation" (1 row), "Substance" (6584 rows), "Variant" (1 row), "Application" (1 row), and "Decision" (6584 rows) are listed. "Product" and "Substance" are checked.
 - Other options:** "Ignore first line(s) at beginning of file" (checked), "Parse next as column headers" (checked), "Discard initial data" (unchecked), "Load at most row(s) of data" (unchecked), "Store blank rows" (checked), "Store blank cells as nulls" (checked), and "Store file source (file names, URLs) in each row" (unchecked).

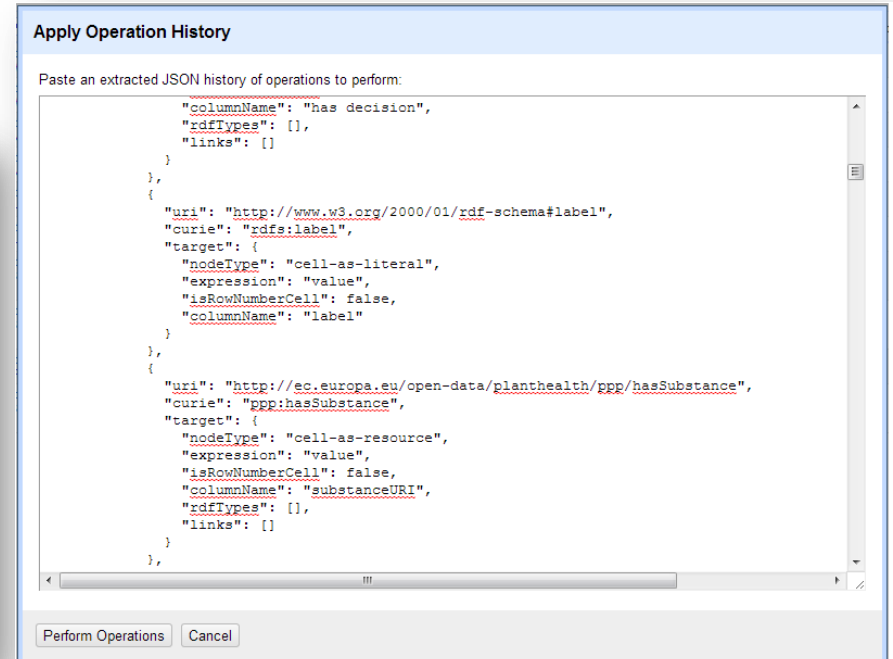
Mapping the raw data to the ontology

You can map the data to the ontology using a simple graphical interface to create or edit an existing RDF skeleton.

You can set the base URI for the data.



Graphical interface to edit an RDF skeleton



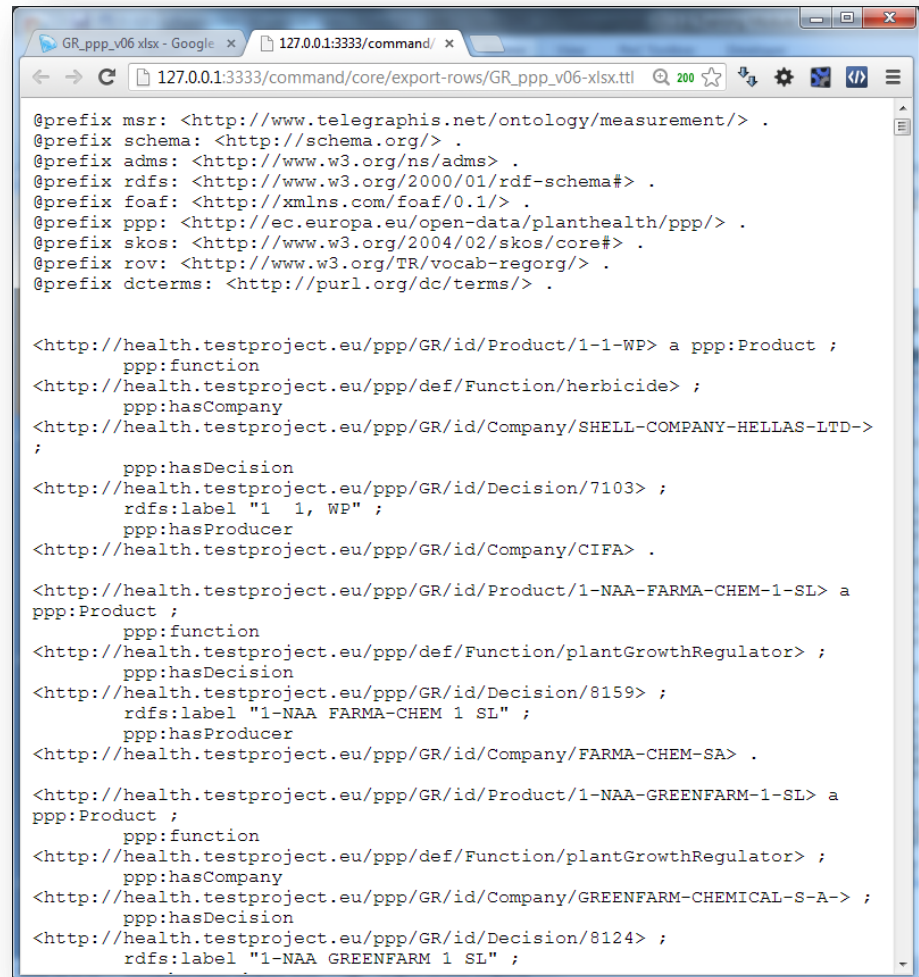
Graphical interface to copy/paste an existing RDF skeleton

Exporting the data in RDF – Linked Data

You can now export your data in:

- RDF/XML; or
- Turtle

Export of the data
in Turtle



The screenshot shows a web browser window with the address bar displaying '127.0.0.1:3333/command/core/export-rows/GR_ppp_v06.xlsx.ttl'. The main content area displays RDF data in Turtle format, including various prefixes and triples. The data describes products, functions, companies, and decisions, such as 'herbicide' and 'plantGrowthRegulator'.

```
@prefix msr: <http://www.telegraphis.net/ontology/measurement/> .
@prefix schema: <http://schema.org/> .
@prefix adms: <http://www.w3.org/ns/adms> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ppp: <http://ec.europa.eu/open-data/planthealth/ppp/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rov: <http://www.w3.org/TR/vocab-regorg/> .
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://health.testproject.eu/ppp/GR/id/Product/1-1-WP> a ppp:Product ;
  ppp:function
  <http://health.testproject.eu/ppp/def/Function/herbicide> ;
  ppp:hasCompany
  <http://health.testproject.eu/ppp/GR/id/Company/SHELL-COMPANY-HELLAS-LTD->
  ;
  ppp:hasDecision
  <http://health.testproject.eu/ppp/GR/id/Decision/7103> ;
  rdfs:label "1 1, WP" ;
  ppp:hasProducer
  <http://health.testproject.eu/ppp/GR/id/Company/CIFA> .

<http://health.testproject.eu/ppp/GR/id/Product/1-NAA-FARMA-CHEM-1-SL> a
ppp:Product ;
  ppp:function
  <http://health.testproject.eu/ppp/def/Function/plantGrowthRegulator> ;
  ppp:hasDecision
  <http://health.testproject.eu/ppp/GR/id/Decision/8159> ;
  rdfs:label "1-NAA FARMA-CHEM 1 SL" ;
  ppp:hasProducer
  <http://health.testproject.eu/ppp/GR/id/Company/FARMA-CHEM-SA> .

<http://health.testproject.eu/ppp/GR/id/Product/1-NAA-GREENFARM-1-SL> a
ppp:Product ;
  ppp:function
  <http://health.testproject.eu/ppp/def/Function/plantGrowthRegulator> ;
  ppp:hasCompany
  <http://health.testproject.eu/ppp/GR/id/Company/GREENFARM-CHEMICAL-S-A-> ;
  ppp:hasDecision
  <http://health.testproject.eu/ppp/GR/id/Decision/8124> ;
  rdfs:label "1-NAA GREENFARM 1 SL" ;
```

LOGD enablers & roadblocks

*From the study on Business Models for LOGD of the ISA
Programme of the European Commission.*

LOGD enablers

- Efficiency gains in data integration – the network effect.
- Forward-looking strategies.
- Increased linking and integrated services.
- Ease of model updates.
- Ease of navigation.
- Open licensing and free access.
- Enthusiasm from ‘champions’.
- Emerging best practice guidance.

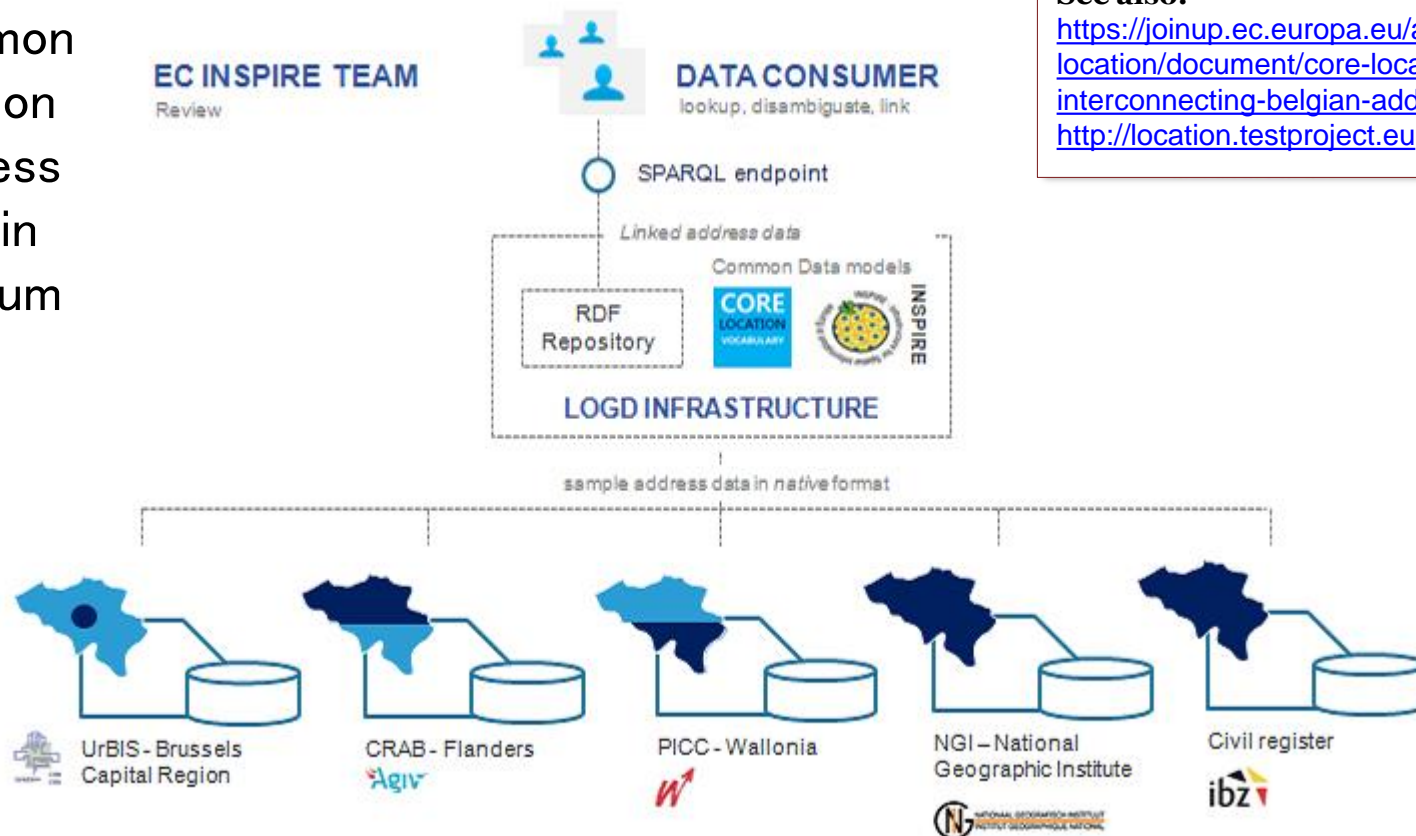
See also:

ISA Study on Business Models for LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Linked data can help you publish structured data and integrate data from different sources

Building a common view on address data in Belgium



See also:

https://joinup.ec.europa.eu/asset/core_location/document/core-location-pilot-interconnecting-belgian-address-data
<http://location.testproject.eu>

LOGD roadblocks

- Necessary investments.
- Lack of necessary competencies.
- Perceived lack of tools.
- Lack of service level guarantees.
- Missing, restrictive, or incompatible licences.
- Surfeit of standard vocabularies.
- The inertia of the status quo – change is accomplished slowly.

See also:

ISA Study on Business Models for LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Linked data initiatives in Europe

Some examples on supra-national, national, regional and private initiatives in the area of linked (open) data across Europe.

Member State initiatives – some examples

DE – Bibliotheksverbund Bayern

Linked data from 180 academic libraries in Bavaria, Berlin and Brandenburg.

IT – Agenzia per l'Italia digitale

Three datasets published as linked data: the Index of Public Administration, the SPC contracts for web services and conduction systems and the Classifications for the data in Public Administration.

NL – Building and address register

The Dutch Address and Buildings base register published as linked data.

UK – Ordnance Survey

Three OS Open Data products published as linked data: the 1:50 000 Scale Gazetteer, Code-Point Open and the administrative geography taken from Boundary Line.

UK – Companies House

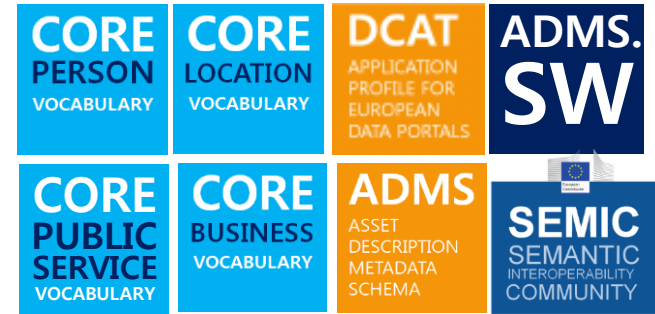
Publishing basic company details as linked data using a simple URI for each company in their database.

See also:

ISA Study on Business Models for LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Linked Government Data & Metadata initiatives funded by the European Commission



European Union Open Data Portal



DIRECTIVES

DIRECTIVE 2013/37/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 26 June 2013
amending Directive 2003/98/EC on the re-use of public sector information

The Linked Government Data Pilots of ISA

Linking data about applications and decisions for authorisation of plant protection products

PPP semantic asset

- PPP Ontology
- PPP Taxonomies

Sample queries

- Find the country where the product is authorised
- Find a product made with a given substance
- Find products made by a company
- Find the product to use on a given pest

Find out more about Linked Data

- Understanding Linked Data by example
- Case study on how Linked Data is transforming eGovernment
- Describe organizations in RDF with Core Business Vocabulary and ORG Ontology
- 10 Rules for Persistent URIs

Type a keyword:

SPARQL Query:

```

PREFIX sds: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ppp: <http://ec.europa.eu/open-data/planthealth/ppp/>
SELECT DISTINCT ?c ?p ?o
FROM <http://health.testproject.eu/ppp/>
WHERE {
  ?c a ppp:Product;
  ?p ?o.
  FILTER(regex(?c,'Acanto','^|')) .
}
LIMIT 100
    
```

Search using the [Faceted Browser](#)

This work is supported by Action 1.1 of the [Interoperable Solutions for European Public Administrations \(ISA\)](#) Programme of the European Commission. [Click on the link to the public dataset.](#)

Linked Data pilots: [Core Location Pilot](#) / [Core Public Service Pilot](#) / [Catalan Open Data Pilot](#) / [E-Health Information Pilot](#) / [Health Surveillance Pilot](#)

DCAT ADMINSW CORE BUSINESS CORE PERSON LOCATION CORE PUBLIC SERVICE VOCABULARY isa Directorate-General Health & Consumer Affairs CORE PUBLIC SERVICE VOCABULARY

<http://health.testproject.eu/PPP/>

Linked maritime surveillance data

CISE semantic asset

- EU Fishing Fleet Ontology

Sample queries

- Find the fishing vessel from a given country
- Find the fishing vessel from a specific port
- Find all the AIS transmission for a fishing vessel
- Find incidents linked to a fishing vessel
- Find the trace for a fishing vessel

Find out more about Linked Data

- Understanding Linked Data by example
- Case study on how Linked Data is transforming eGovernment
- Describe organizations in RDF with Core Business Vocabulary and ORG Ontology
- 10 Rules for Persistent URIs

Type a keyword:

SPARQL Query:

```

PREFIX sds: <http://health.testproject.eu/fishing/ese/ontology/>
PREFIX sds: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?fishingVessel ?CountryCodeIRI ?subsector ?CountryCodeIRI ?
WHERE {
  ?fishingVessel a sds:FishingVessel;
  dcterms:spatial ?CountryCodeIRI;
  dcterms:title ?fishingVessel.
  FILTER(regex(?fishingVessel,'sitanic','^|')) .
}
LIMIT 100
    
```

This work is supported by Action 1.1 of the [Interoperable Solutions for European Public Administrations \(ISA\)](#) Programme of the European Commission. [Click on the link to the public dataset.](#)

Linked Data pilots: [Core Location Pilot](#) / [Core Public Service Pilot](#) / [Catalan Open Data Pilot](#) / [E-Health Information Pilot](#) / [Health Surveillance Pilot](#)

DCAT ADMINSW CORE BUSINESS CORE PERSON LOCATION CORE PUBLIC SERVICE VOCABULARY isa Directorate-General Health & Consumer Affairs CORE PUBLIC SERVICE VOCABULARY

<http://health.testproject.eu/CISE/>

Core Public Service Pilot: describe public services only once

Governments use local, regional and national access portals to give businesses, citizens, and public administrations basic information about their public services. Unfortunately, this basic information often duplicated, unstructured, and not machine-readable. This fragmentation makes it difficult for citizens, businesses, and public administrations to find information about the public service that they need. It also leads to situations where basically the same information about a public service is re-created in different applications or by different governments.

The [Core Public Service Vocabulary](#) allows public administrations to describe their service only once using a standard, extensible, and machine-readable vocabulary and make these descriptions re-used on many governmental access portals. This pilot implementation shows how this can be done using a Linked Data infrastructure for a small sample of public service descriptions originating from various European public administrations.

Sample public service descriptions

- From the [SPOCS large-scale pilot](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/RealEstateAgencyLicence>
 - <http://cpsv.testproject.eu/dmbl/PublicService/CataniaEstablishmentLicence>
 - <http://cpsv.testproject.eu/dmbl/PublicService/ArchitectRegistration>
 - <http://cpsv.testproject.eu/dmbl/PublicService/TECRecreation>
 - <http://cpsv.testproject.eu/dmbl/PublicService/ArchitectRegistration>
- From the [Flemish Intergovernmental Product and Service Catalogue](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/24>
 - <http://cpsv.testproject.eu/dmbl/PublicService/117>
 - <http://cpsv.testproject.eu/dmbl/PublicService/179>
 - <http://cpsv.testproject.eu/dmbl/PublicService/279>
 - <http://cpsv.testproject.eu/dmbl/PublicService/291>
 - <http://cpsv.testproject.eu/dmbl/PublicService/201>
 - <http://cpsv.testproject.eu/dmbl/PublicService/246>
 - <http://cpsv.testproject.eu/dmbl/PublicService/247>
 - <http://cpsv.testproject.eu/dmbl/PublicService/415>
 - <http://cpsv.testproject.eu/dmbl/PublicService/588>
 - <http://cpsv.testproject.eu/dmbl/PublicService/932>

- From the [Irish Citizens Information portal](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/DivinaTest>
 - <http://cpsv.testproject.eu/dmbl/PublicService/EuropeanHealthInsuranceCard>
 - <http://cpsv.testproject.eu/dmbl/PublicService/PublicServiceCard>

- From the [e-CODEX large scale pilot](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/SmallClaim>

<http://cpsv.testproject.eu/CPSV/>

Non-governmental applications

BBC MUSIC

HOME | SHOWCASE | REVIEWS | GENRES

Search by Artist

The Beatles

Formed 1967 | Disbanded 10 April 1970

Share This Page 15 so far

BBC Music Showcase

BBC MUSIC SHOWCASE

Watch and listen to exclusive music clips

Latest Tracks Played On The BBC

Twist & Shout
BBC RADIO 2 | JOE BALL AND RICHARD BACON GO FOR GOLD 03/06/2012

Tomorrow Never Knows
BBC & MUSIC | RADCLIFFE AND MACMONE WEDNESDAY - JOHNNY G-RASP

She Loves You
BBC RADIO 2 | KEN BRUCE CLAUDIA WINKLEMAN SITS IN

Sgt Pepper's Lonely Hearts/With A Little Help

LinkedGeoData.org

Search results for: berlin

four search was: 'berlin'

1. Berlin
2. Berlin
3. Berlin
4. Berlin
5. Berlin
6. Berlin
7. Berlin
8. Berlin
9. Berlin
10. Berlin

View node 63733593

Link on OpenStreetMap

Name: [redacted]

Description: [redacted]

Image: [redacted]

Source, ref: natural stone monument

opencorporates

The Open Database of The Corporate World

Search companies: 30 million companies, 65 jurisdictions

We have information on **55,021,554** companies

search companies search officers

Filter by jurisdiction

- 4,297 Abu Dhabi (UAE)
- 110,031 Alaska (US)
- 40,455 Albania
- 782,204 Arizona (US)
- 44,739 Aruba
- 654,016 Bahrain
- 30,420 Barbados
- 43,614 Bermuda
- 3,024,466 California (US)
- 664,849 Canada
- 636,484 Connecticut (US)
- 175,482 Croatia
- 220,744 District of Columbia (US)
- 337,339 Dubai (UAE)
- 4,399 Dubai (UAE)
- 377,444 Finland
- 5,997,243 Florida (US)
- 1,672,028 Georgia (US)
- 104,852 Gibraltar
- 8,014 Greenland
- 133,552 Iceland
- 224,035 Idaho (US)
- 4,268,304 India
- 430,944 Iowa (US)
- 506,230 Ireland
- 441,232 Isle of Man
- 83,465 Jamaica
- 141,769 Jersey
- 134,169 Liechtenstein
- 908,802 Louisiana (US)

Just released: OpenCorporates API

Use Google Refine to match companies in your data

Watch screenshot, or read all about it

"This is the kind of resource the (Digital) Single Market needs"

Néelie Kroes, Vice-President of the European Commission

Recently updated corporate groupings

BARCHESTER HEALTHCARE 21

LLOYDS BANKING GROUP 21 ECONOMIST GROUP 1

AEGON 1 ROYAL BANK OF SCOTLAND 2

STAGECOACH 10 MAERSK 235 ALIB 0

NEW! Search officers/directors

SIGMA

SEMANTIC INFORMATION MASHUP

Help About Forum

tin berners-lee

Tim Berners-Lee

picture: [redacted]

given name: Tim [1,11,12,14,15,17,18,19]

family name: Berners-Lee [1,11,12,14,15,17,18,19,20]

comment: Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FREng, FRSA (born 8 June 1955, also known as "TimBL"), is a British engineer and computer scientist and ART professor credited with inventing the World Wide Web, making the first proposal for it in March 1989. On 25 December 1990, with the help of Robert Calliau and a young student of CERN, he implemented the first successful communication between an MTP client and server via the Internet.

is creator of: [Tabulator](#) [9,10,11,12,13,14,15,17,18,19,20]

alternate: http://ref.freebase.com/ref/an/tim_berners_lee [9]

author name: vianeta181676 [9]

author url: <http://www.slideshare.net/vianeta181676> [9]

admins: 112431582,500054654,220400,612158401,80870553,1502271052,695398126 [1]

birth year: 1955-01-01 00:00:00 [1]

Sources (20) Approved (0) Rejected (0)

- 1 Tim Berners-Lee - Wikipedia, 12 Facts | 2011-05-19
- 2 Tim Berners-Lee - Wikipedia, 9 Facts | 2011-05-19
- 3 Untitled document, 10 Facts | 2011-05-24
- 4 Tim Berners-Lee 2 Facts | 2011-05-19
- 5 Untitled document, 14 Facts | 2014-01-13
- 6 Tim Berners-Lee facts - , 4 Facts | 2011-02-23
- 7 SIOC profile for 'http://...', 2016 Facts | 2013-02-19
- 8 Untitled document, 4 Facts | 2013-02-03
- 9 About: Tim Berners-Lee, 180 Facts | 2011-05-18
- 10 Untitled document, 218 Facts | 2011-01-10
- 11 Tim Berners-Lee, 130 Facts | 2011-05-20
- 12 Timothy Berners-Lee, 137 Facts | 2014-05-18
- 13 About: Timothy Berners-Lee, 224 Facts | 2011-02-09
- 14 Untitled document, 197 Facts | 2011-05-18
- 15 Untitled document, 556 Facts | 2011-05-19
- 16 Berners-Lee, Tim, 3430 Facts | 2011-01-01



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.

Group questions



<http://www.visualpharm.com>

Is there supply and demand for (Linked) Open Government Data in your country?



<http://www.visualpharm.com>

What are, in your opinion, the expected benefits and pitfalls of Linked Data?



<http://www.visualpharm.com>

Are there any Linked (Open) Data initiatives in your country? If so, how many stars would you give them?

Take also the online test [here!](#)

Thank you!
...and now YOUR questions?

References

Slide 6:

- EUCLID. Course 1: Introduction and Application Scenarios. <http://www.euclid-project.eu/modules/course1>
- Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Slides 8:, 13, 36, 38:

- ISA Programme. Case study on how Linked Data is transforming eGovernment. <https://joinup.ec.europa.eu/community/semic/document/case-study-how-linked-data-transforming-egovernment>
- Tim Berners-Lee. Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>

Slide 9:,

- ISA Programme Study on Business Models for LOGD <https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Slide 12:

- The Open Knowledge Foundation. Open Data – An Introduction. <http://okfn.org/opendata/>

Slides 18-28:

- 5 ★ Open Data. <http://5stardata.info/>

Slide 19:

- UK National Archives, Sustainable development targets 2011-12.

Slide 21:

- Data.gov.uk. Housing stock. <http://data.gov.uk/dataset/uk-housing-stock>

Slide 23:

- Data.gov.uk. Road Safety Data. <http://data.gov.uk/dataset/road-accidents-safety-data>

Slide 25 & 27:

- ISA Organization Ontology pilot - Linking public sector's organisational data, https://joinup.ec.europa.eu/asset/core_business/document/organization-ontology-pilot-linking-public-sectors-organisational-data
<http://data.ydmed.gov.gr/>

Slide 37:

- ISA Programme. Core Location Pilot - Interconnecting Belgian Address Data. https://joinup.ec.europa.eu/asset/core_location/document/core-location-pilot-interconnecting-belgian-address-data

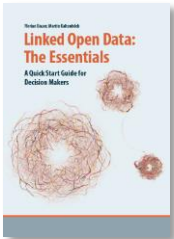
Slides 30-34:

- Open Refine: <https://github.com/OpenRefine>
- RDF Extension: <http://refine.deri.ie/>
- ISA Programme, Linking data about applications and decisions for authorisation of PPP, http://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp

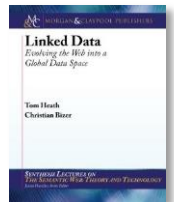
Slide 40

- Bibliotheksverbund Bayern, <http://lod.b3kat.de/doc>
- Agenzia per l'Italia Digitale, <http://spcdata.digitpa.gov.it/data.html>
- NL – Building and address register, <http://lod.Geodan.nl>
- UK Ordnance Survey, <http://data.ordnancesurvey.co.uk/>
- UK Companies House, <http://companieshouse.gov.uk/>

Further reading



Linked Open Data: The Essentials. Florian Bauer, Martin Kaltenböck.
<http://www.semantic-web.at/LOD-TheEssentials.pdf>



Linked Data: Evolving the Web into a Global Data Space. Tom Heath and Christian Bizer.
<http://linkeddatabook.com/editions/1.0/>



Linked Open Government Data. Li Ding Qualcomm, Vassilios Peristeras and Michael Hausenblas.
<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>



EUCLID - Course 1: Introduction and Application Scenarios
<http://www.euclid-project.eu/modules/course1>

Related projects and initiatives



LOD2 FP7 project, <http://lod2.eu/>



The Open Knowledge Foundation, <http://okfn.org/>



W3C Semantic Web, <http://www.w3.org/standards/semanticweb/>



EUCLID, <http://projecteuclid.org/>



ISA Programme, <http://ec.europa.eu/isa/>



W3C LOGD WG, http://www.w3.org/2011/gld/wiki/Main_Page



LOD Around The Clock FP7 project, <http://latc-project.eu/>



Data.gov.uk, <http://data.gov.uk/linked-data>

Be part of our team...

Find us on



[Open Data Support](http://www.slideshare.net/OpenDataSupport)

<http://www.slideshare.net/OpenDataSupport>



[Open Data Support](http://goo.gl/y9ZZI)

<http://goo.gl/y9ZZI>

Follow us



[@OpenDataSupport](https://twitter.com/OpenDataSupport)

Join us on



joinup

<http://www.opendatasupport.eu>

Contact us

contact@opendatasupport.eu