OPEN
DATA
SUPPORT

Presentation metadata

Training Module 1.2

Introduction to Linked Data

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;

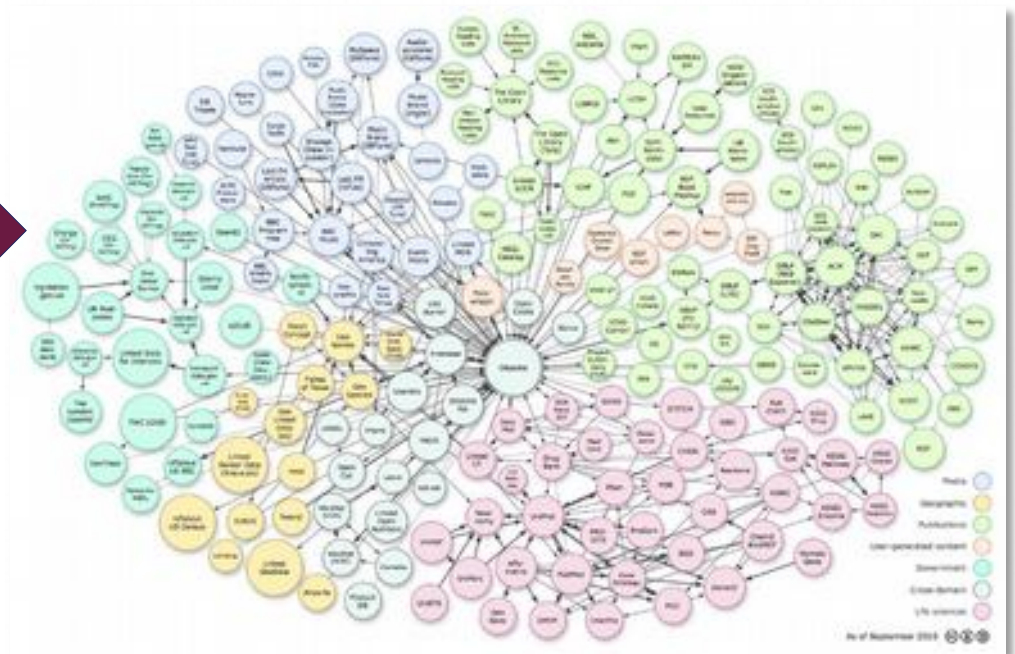- An overview of linked data initiatives in Europe.

What is linked dat

Evolution from a document-based Web to a Web of inte data.

The Web is evolving from a "Web of linked docume a "Web of linked data"… (1/2)

Web of documents…
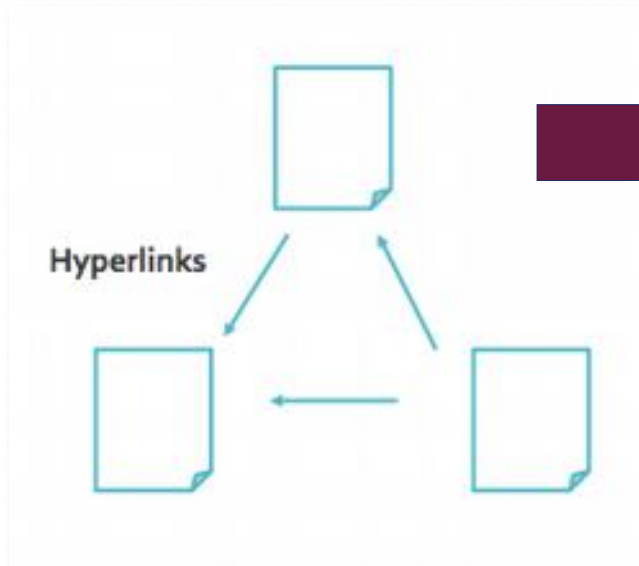
Web of linked data…

The Web is evolving from a "Web of linked docume 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/

EUROPEAN DATA PORTAL

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

EUROPEAN
DATA PORTAL

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

OPEN DATA SUPPORT

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">Уницеф - 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.

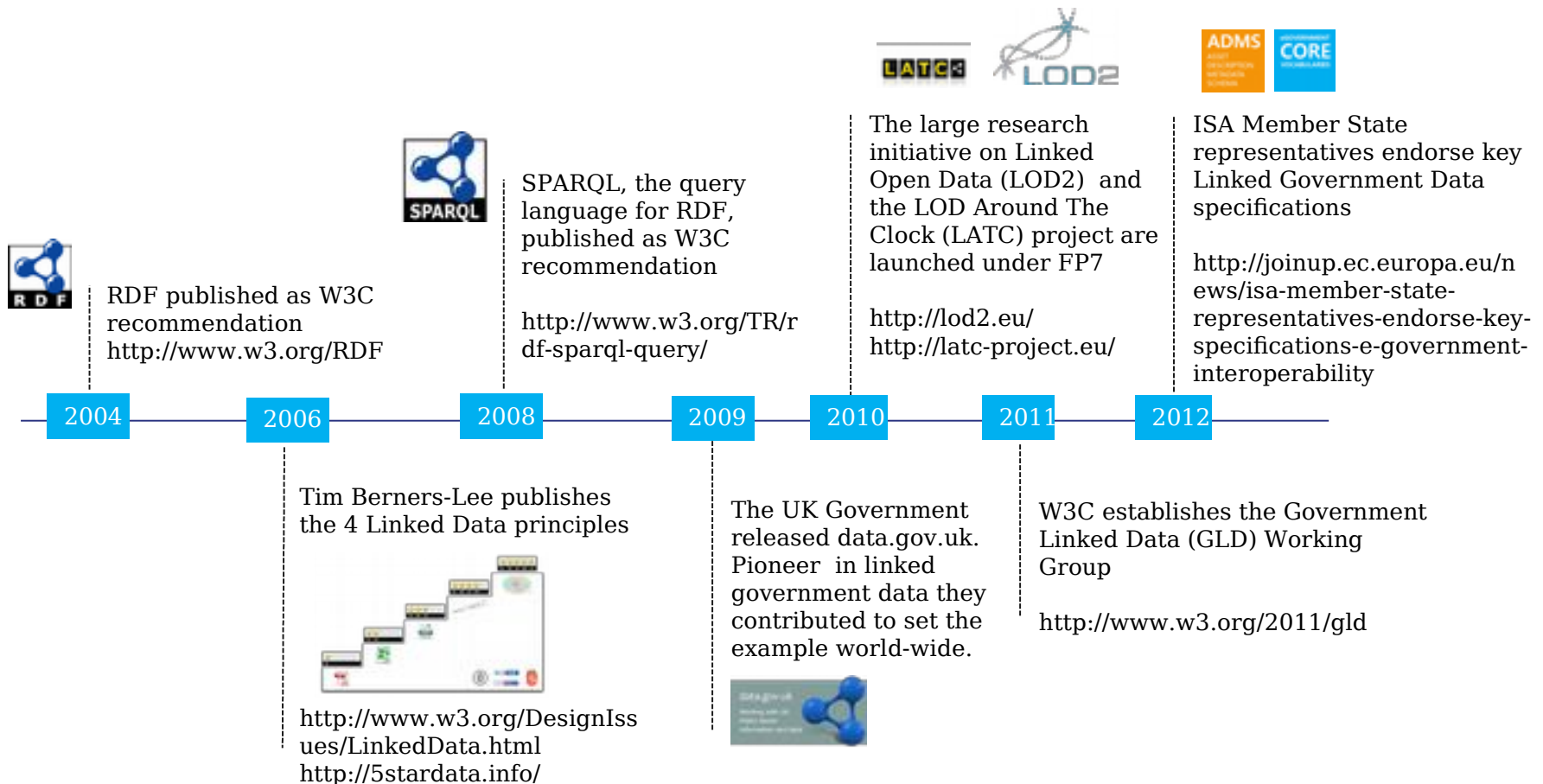Data can be linked to URIs from other data sources, using open standards such as RDF without being publicly available under an open licence.

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

Key milestones for linked government data

The large research initiative on Linked Open Data (LOD2) and the LOD Around The Clock (LATC) project are launched under FP7

http://lod2.eu/
http://latc-project.eu/

ISA Member State representatives endorse key Linked Government Data specifications

http://joinup.ec.europa.eu/news/isa-member-state-representatives-endorse-key-specifications-e-government-interoperability

SPARQL, the query language for RDF, published as W3C recommendation

http://www.w3.org/TR/rdf-sparql-query/

RDF published as W3C recommendation
http://www.w3.org/RDF

| 2004 | 2006 | 2008 | 2009 | 2010 | 2011 | 2012 |

Tim Berners-Lee publishes the 4 Linked Data principles

http://www.w3.org/DesignIssues/LinkedData.html
http://5stardata.info/

The UK Government released data.gov.uk. Pioneer in linked government data they contributed to set the example world-wide.

W3C establishes the Government Linked Data (GLD) Working Group

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

Linked data foundations

URIs for naming things, RDF for describing data and SP 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/**

EUROPEAN DATA PORTAL

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.

See also:
http://www.slideshare.net/OpenDataSupport/introduction-to-rdf-sparql

How to publish linked data?

Paving the way towards 5-star linked data

EUROPEAN DATA PORTAL

OPEN DATA SUPPORT

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*

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 redu... business-related travel by 25% from 2009–10 levels by April 201...

Licensed under **OGL** 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.	

OPEN DATA SUPPORT

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**	

XLS

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:

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

http://data.ydmed.gov.gr/doc/organization/16180

Type   Organization
Raw data   HTML | RDF/XML | Turtle

Category   ΥΠΟΥΡΓΕΙΑ

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*

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.

Example

Using Open Refine for RDF to publish tabular data as L
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=4Ve93C238gI

OPEN DATA SUPPORT

Case study: *Linking data about plant protection pr*

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.

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.
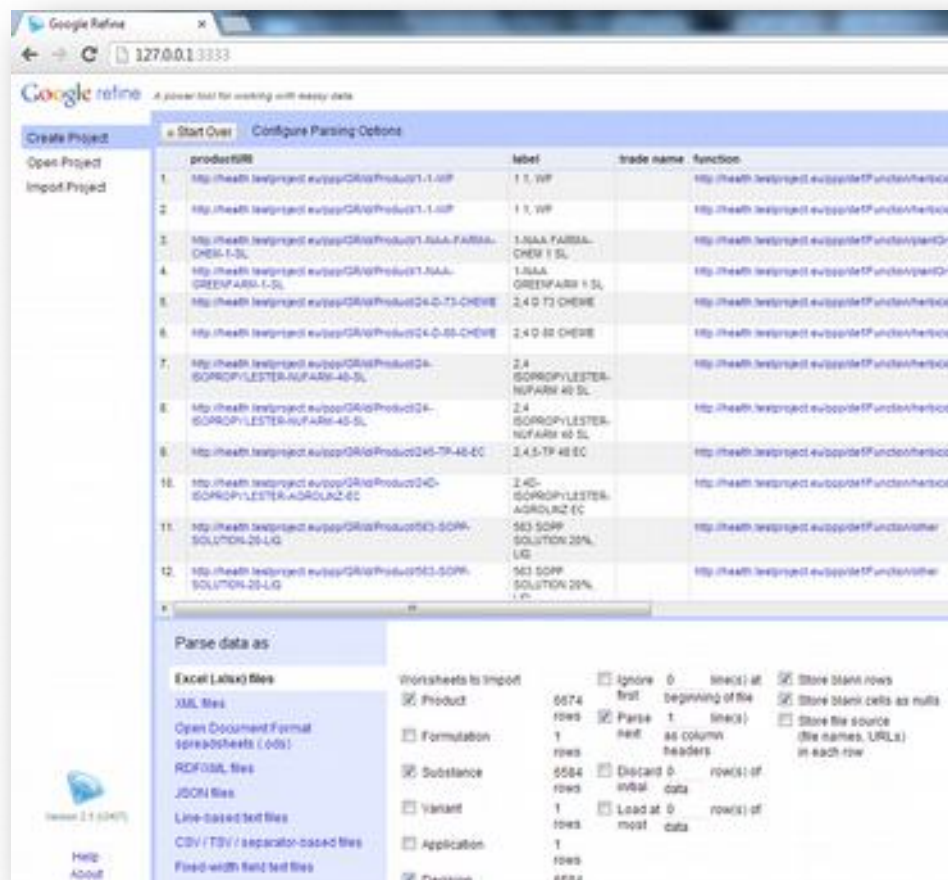
Mapping the raw data to the ontology

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
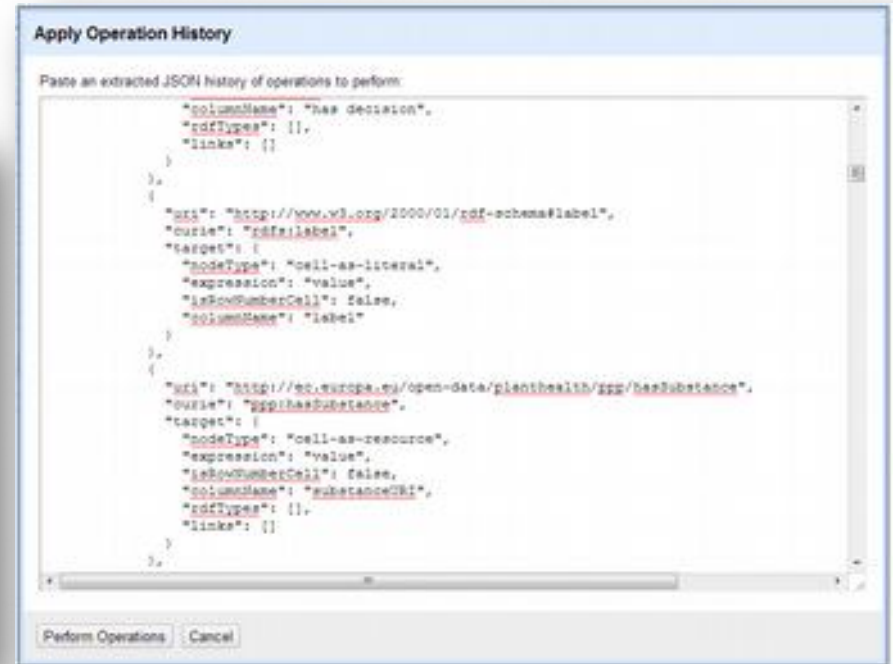
Graphical interface to copy/paste
an existing RDF skeleton

Exporting the data in RDF – Linked Data

You can now export your data in:

- RDF/XML; or

- Turtle

Export of the data in Turtle

OPEN DATA SUPPORT

LOGD enablers & roadblocks

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

OPEN DATA SUPPORT

EUROPEAN DATA PORTAL

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'.

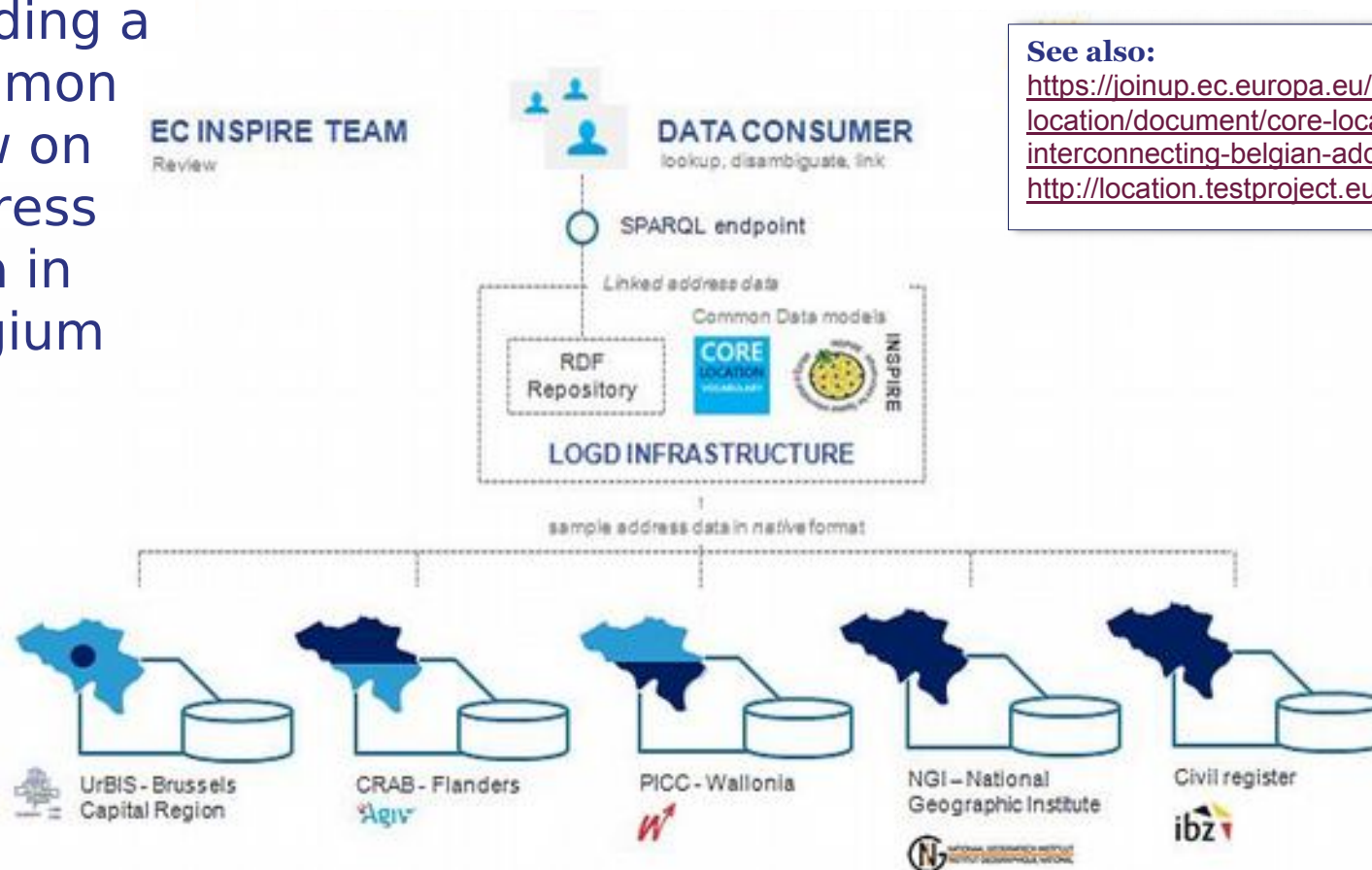- Emerging best practice guidance.

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

OPEN DATA SUPPORT

Linked data can help you publish structured data a 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

EUROPEAN
DATA PORTAL

OPEN DATA SUPPORT

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 digitiale

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

EUROPEAN DATA PORTAL

Linked Government Data & Metadata initiatives fu
the European Commission

The Linked Government Data Pilots of ISA

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

http://maritime.testproject.eu/CISE/

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

Non-governmental applications

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:

  o Data integration with small impact on legacy systems;

  o Enables for semantic interoperability;

  o Enables creativity and innovation through context and knowledge-creation.

Group questions

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

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

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

Take also the online test her

OPEN DATA SUPPORT

Thank you!
...and now YOUR question

EUROPEAN DATA PORTAL

OPEN DATA SUPPORT

This presentation has been created by Open Data S

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.

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

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/

EUROPEAN DATA PORTAL

OPEN DATA SUPPORT

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

Open Data Support
http://goo.gl/y9ZZI

Join us on

http://www.opendatasupport.eu

Follow us

@OpenDataSupport

Contact us

contact@opendatasupport.eu

OPEN DATA SUPPORT
EUROPEAN DATA PORTAL