



Data Archiving and Networked Services

Linking knowledge spaces

Christophe Guéret (@cgueret)



Take home message

- Best practices for data are so 90's ...
but, no worries, there are alternatives ;-)
- “Linked Data” is not a new data exchange standard. It is a way to publish and link data using the Web
- Linked knowledges spaces are richer and easier to map & explore



Moving back in time...





Dealing with documents until 1989

- 4 simple, natural, steps (using the Internet) :
 - Get a document from a source
 - Find a software able to process it
 - Process and write down links to other documents
 - Keep an eye on updates
- Somewhat cumbersome
 - Authors can not easily link documents
 - Hard to process & keep up with updates
 - Hard to get a “big picture” out



Then came the Web ...

- **Easy**
 - Web browsers display Web documents served by Web servers and wrote using a common language
- **Convenient**
 - Latest version of a document available from the Web server
 - Links between unique identifiers assigned to Web documents (Uniform Resource Identifier)
- **Scalable**
 - Decentralised document publication platform



This had a tremendous success!

- > 40 billion indexed web documents
- Numerous standards and tools
- Dedicated services to find and use documents



mozilla
Firefox®



We could hardly go back now

- Would you dare not creating a web site for your research group or yourself ?
- Web technologies are reaching out beyond simple documents



Helping you to find,
access, and reuse data



Now it is data that matters

Label your
Professional Networks

- Social media
- Médias, comm, contenus
- Divers
- Explorateurs éclairés
- Les historiques
- Fing, institutions, recherche
- Experts & consultants
- Marketing, eMarketing



Dealing with data until, well, now

- 4 simple, natural, steps (using the Internet) :
 - Get a dataset from a source
 - Find a software able to process it
 - Process and write down links to other datasets
 - Keep an eye on updates
- Somewhat cumbersome
 - Authors can not easily link datasets
 - Hard to process & keep up with updates
 - Hard to get a “big picture” out



Sounds familiar ?

- We deal with data the way we dealt with documents 20 years ago
- Lots of different formats, no links, hard to have up-to-date data, model de-coupled from the data...

MODS

 **sdmx**
Statistical Data and Metadata eXchange

< ddi >
Member of the DDI Alliance





Linked Data



- 4 design principles, introduced in 2006
 - Use URIs as names for things
 - Use HTTP URIs so that people can look up those names
 - When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
 - Include links to other URIs so that they can discover more things
- Publish data using the Web (not *on* the Web)



Linked Data



- 4 design principles, introduced in 2006
 - Use URIs as names for things
 - Use HTTP URIs so that people can look up those names
 - When someone looks up information, using the HTTP protocol, they get more information
 - Include links to other things

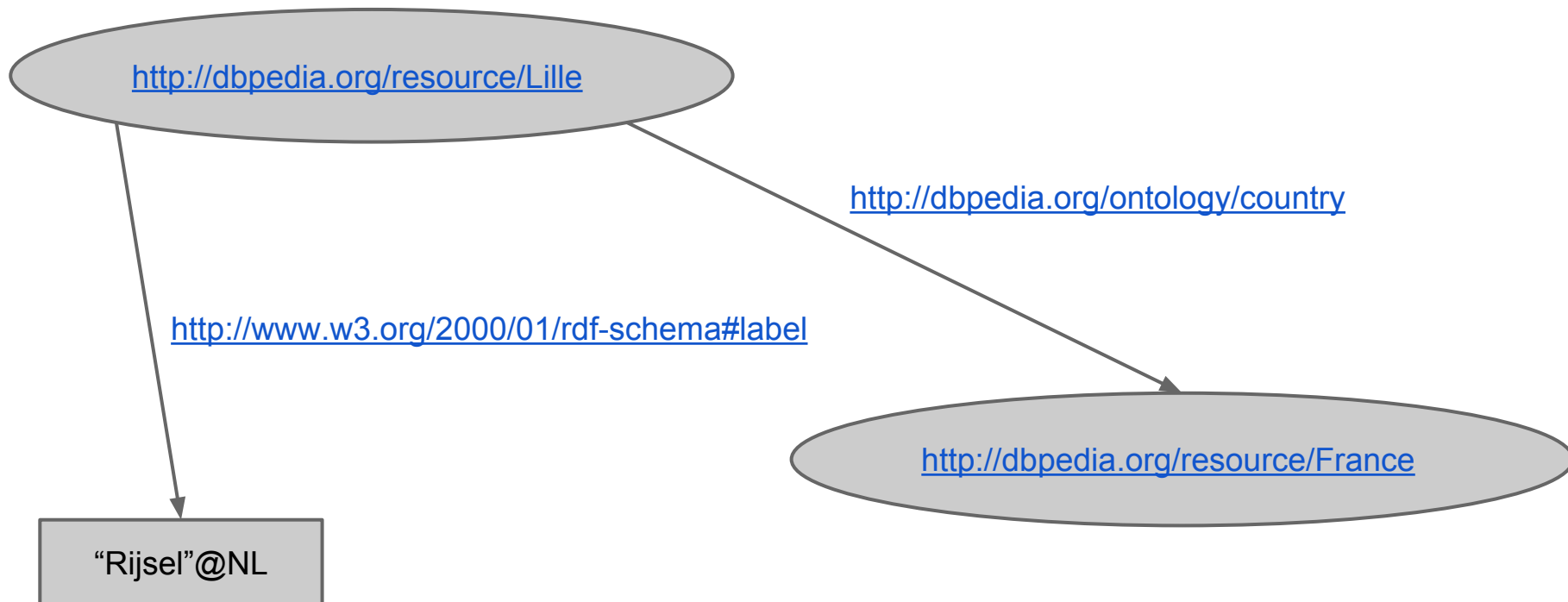
Packed with good stuff:
Open standards
HTTP
ReST
De-centralised publication

- Publish data using the Web (not *on* the Web)



Concretely...

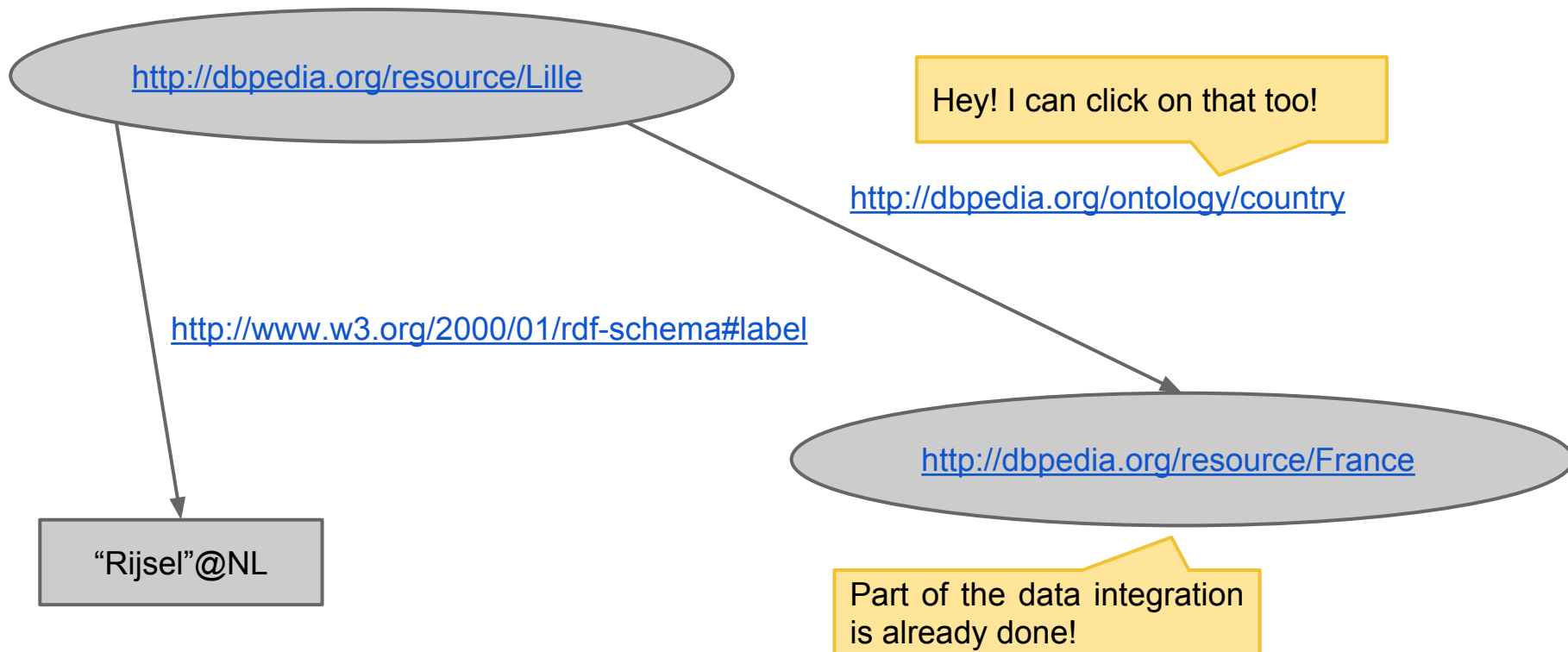
- Lille is in France and called “Rijsel” in Dutch





Concretely...

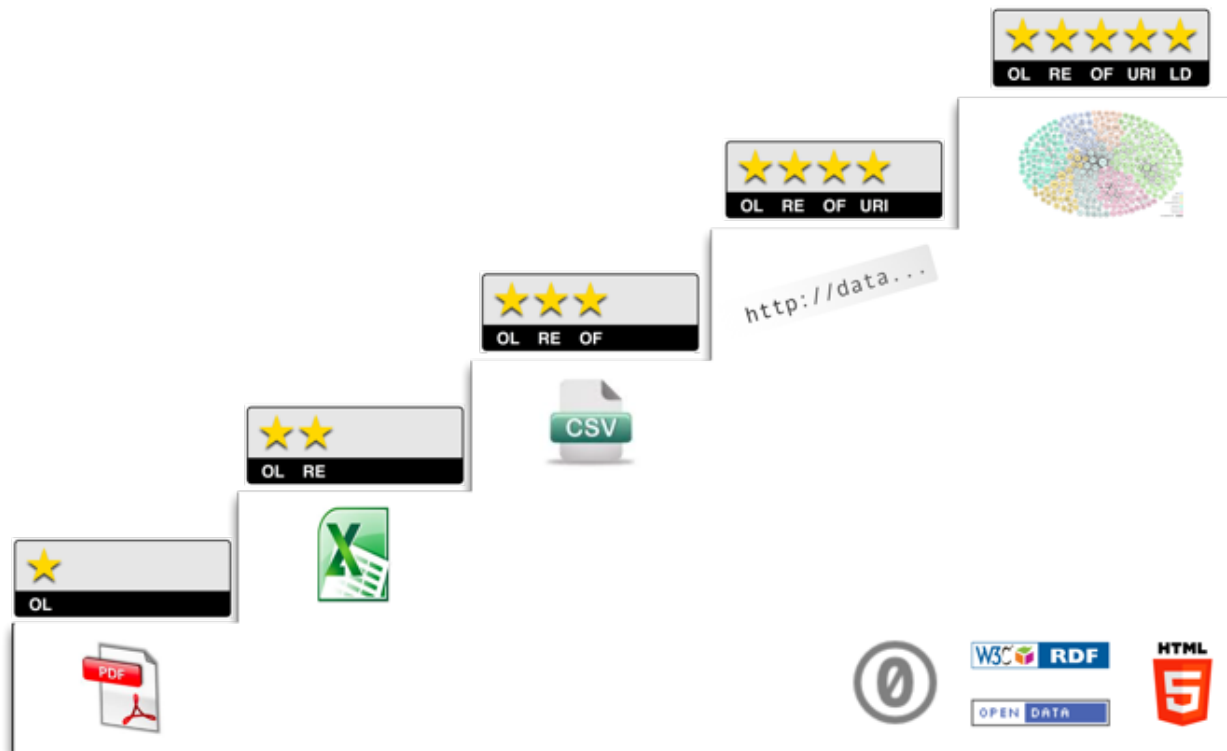
- Lille is in France and called “Rijsel” in Dutch





Linked Data + Open Data = LOD

- 5-star scheme to get from closed data to open linked data <http://5stardata.info/>





LOD + Semantics = Semantic Web

- Tell a bit about the Semantics of your data and a computer will derive new facts for you
- For instance, “All the cities in France are in Europe” => “Lille is in Europe”

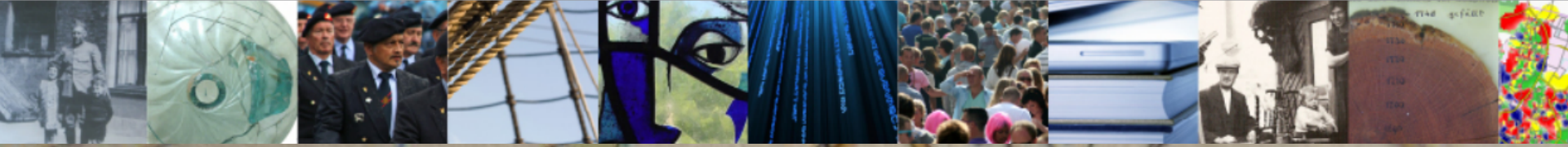


Let's take a step back

- A quick comparison of some features...

	Web of Documents	Web of Data	Any data on the Web
Model	Tree	Statements	<i>Varied</i>
Identifiers	URI	URI	URN + URI
Serialisation	XML	XML, TTL, ...	XML, CSV, ...
Granularity	Page	Statement	Data set
Access	Look up	Look up	Download
Schema	HTML	<i>Varied</i>	<i>Varied</i>
Query language	XQuery / XPath	SPARQL	<i>Varied</i>


Sweet spot for data integration !



Linking & Mapping knowledge spaces





Mapping knowledge spaces

- Without Linked Data
 - Download individual data sets
 - Integrate them as another data set
 - Map the output
 - (return to the first step on every update)
- With Linked Data
 - Index the different data sources
 - Map the output using “live” data
 - Eventually, cache the data for speed/accessibility



Example: Research landscape

- Without www.narcis.nl

The screenshot shows the NARCIS website in a browser window. The address bar displays 'www.narcis.nl'. The page features a dark blue header with the NARCIS logo and navigation links: HOME, ABOUT NARCIS, and NEDERLANDS. Below the header, a dark blue banner contains the text 'The gateway to scholarly information in the Netherlands' and a link to 'Submit Content to NARCIS'. A search bar with the placeholder 'Search...' and a 'SEARCH' button is present. Below the search bar, a row of statistics is displayed:

Category	Count
PUBLICATIONS	811,792
DATA SETS	28,993
RESEARCH	59,419
PEOPLE	50,634
ORGANISATIONS	2,894
ENHANCED PUBLICATIONS	1,733 <small>beta release</small>

Below the statistics, there is a section titled 'SCIENTIFIC NEWS FROM OTHER SOURCES (DUTCH)' with a list of news items:

- > Zoon schuldig aan hulp zelfdoding moeder (99), geen celstraf
22-10-2013 | de Volkskrant
- > Mart.Spruijt Theaterafficheprijs
22-10-2013 | UvA Nieuws
- > Fascinatie voor de bijzondere opvoedsituatie
22-10-2013 | UvA Nieuws
- > More scientific news

At the bottom of the page, the text 'Royal Netherlands Academy of Arts and Sciences' is visible, followed by the full name 'KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN' in a stylized font. The DANS logo is located in the bottom right corner of the slide.



Example: Research landscape

- With : <http://narcis-vivo.appspot.com/>

People browse

This application let you browse researchers listed in ISIDORE and the "eur", "kun", "ouh", "rug", "rul", "rum", "ruu", "tud", "tue", "tum", "uva", "vut", "vua" and "wur" universities. Just type a name and follow the links.

You can also click on one of the following examples: [J. Leeuwen \(ruu\)](#) [F.A.H. Harmelen \(vua\)](#) [Magali Hersant \(isidore\)](#)

Centres of interest

[Computer science](#) [Mathematics](#)

Description

People with shared interests

#	Predicate	Object	#	Source	Person
0	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://vivoweb.org/ontology/core#FacultyMember	0	isidore	Sivaji Bandyopadhyay (isidore)
1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person	1	tue	H. Waardt (tue)
2	http://www.w3.org/2000/01/rdf-schema#label	Leeuwen; Prof.dr. J. van	2	isidore	Alexis Nasr (isidore)
3	http://xmlns.com/foaf/0.1/firstName	J.	3	uva	R.G. Strom (uva)
4	http://xmlns.com/foaf/0.1/lastName	Leeuwen	4	uva	H.P.M. Jägers (uva)
5	http://dans.knaw.nl/ontology/vivon#daild	068413025	5	vua	I.R. Hellsten (vua)
6	http://vivoweb.org/ontology/core#preferredTitle	Prof.dr.	6	kun	A.J.A. Felling (kun)
7	http://vivoweb.org/ontology/core#middleName	van	7	eur	H.W.G.M. Heck (eur)
8	http://vivoweb.org/ontology/core#personInPosition	http://ruu.example.org/individual/Position_PP	8	isidore	Daniel Kroh (isidore)

Dutch + French data

Running without data



Live browsing of the Web of Data

- LODLive at <http://en.lodlive.it/>

lodlive IT / EN / FR

simple search ?

CHOOSE A DATASET

choose...

INSERT A KEYWORD

start >>

insert URI ?

PASTE A RESOURCE ADDRESS

start >>

browse your data ?

PASTE SOME RDF DATA

PASTE AN ADDRESS

coming soon

start >>

Live on LodLive

CHOOSE AN ENDPOINT

choose...

fr.dbpedia.org ?

RESOURCE

choose...

start >>

linked.opendata.cz ?

RESOURCE

choose...

start >>

LodLive project provides a demonstration The basic principle underlying LodLive is to prove that resources published according to the W3C

Information relevant to FAO efforts

- OpenAGRIS : <http://agris.fao.org/openagris/index.do>

The screenshot displays the OpenAGRIS website interface. At the top, there is a navigation bar with the FAO logo and the text "Food and Agriculture Organization of the United Nations". Below this, there is a search bar labeled "OpenAGRIS BETA" and a search input field containing "Find resources...". The main content area is divided into several sections:

- Statistics:** A box indicating the total number of records: 7,636,604.
- Number of records added to the AGRIS collection:** A line graph showing the number of records added per year from 2005 to 2013. The y-axis represents the number of records, ranging from 32,768 to 2,097,152. The x-axis represents the year. The data points are: 2005 (65,536), 2006 (40,000), 2007 (35,000), 2008 (75,000), 2009 (45,000), 2010 (55,000), 2011 (35,000), 2012 (1,500,000), and 2013 (2,097,152).
- What is Linked Open Data?:** A section explaining that Linked Open Data (LOD) is Linked Data distributed under an open license that allows its reuse for free. It includes a "Read More" link.
- About OpenAGRIS:** A section stating that OpenAgris is a Web application that aggregates information from different Web sources to expand the AGRIS knowledge. It includes a "Read More" link.
- LODE-BD Recommendations:** A section explaining that LODE-BD aims to support the selection of appropriate encoding strategies for producing meaningful Linked Open Data (LOD)-enabled bibliographical data (directly or indirectly). It includes a "Read More" link.

At the bottom left, there is a section titled "Some random documents" with a link to "Strategies for Controlling Volunteer Corn and Wheat". At the bottom right, there is a social media share button for Google+ with a count of 2.



Take home message

- Modern best practices are so 90's ...
but this can be changed ;-)
- “Linked Data” is not a new data exchange standard. It is a way to publish and link data using the Web
- Linked knowledges spaces are richer and easier to map & explore