

WWW 2011

29th March 2011, Hyderabad, India

4th Linked Data on the Web Workshop (LDOW 2011)

Christian Bizer, Freie Universität Berlin, Germany

Tom Heath, Talis, UK

Tim Berners-Lee, W3C/MIT, USA

Michael Hausenblas, Richard Cyganiak, DERI, Irland

Programme

■ Introduction

- 9:00-9:25: **Introduction to the Workshop and Overview of the State of the Web of Data**
(Christian Bizer, Tom Heath, Tim Berners-Lee, Michael Hausenblas)

■ Session 1: Publishing Linked Data

- 9:25-9:45: **A Privacy Preference Ontology (PPO) for Linked Data**
(Owen Sacco, Alexandre Passant)
- 9:45-10:10: **Publishing Provenance Information on the Web using the Memento Datetime Content Negotiation**
(Sam Coppens, Erik Mannens, Davy Van Deursen, Patrick Hochstenbach, Bart Janssens, Rik Van De Walle)

■ Coffee Break

- 10:10-10:40

■ Session 2: Infrastructure and Architectures

- 10:40-11:00: **Augmenting the Web of Data using Referers**
(Hannes Mühleisen , Anja Jentzsch)
- 11:00-11:25: **RESTful writable APIs for the web of Linked Data using relational storage solutions**
(Antonio Garrote, María N. Moreno García)
- 11:25-11:50: **How Caching Improves Efficiency and Result Completeness for Querying Linked Data**
(Olaf Hartig)
- 11:50-12:15: **A Main Memory Index Structure to Query Linked Data**
(Olaf Hartig, Frank Huber)

■ Lunch Break

- 12:15-14:00

■ Session 3: Linked Data Applications

- 14:00-14:20: **LiDDM: A Data Mining System for Linked Data**
(Venkata Narasimha Pavan Kappara, Ryutaro Ichise, Vyas O.P.)
- 14:20-14:45: **Talash: Friend Finding In Federated Social Networks**
(Ruturaj Dhekane, Brion Vibber)
- 14:45-15:05: **Automatically Annotating Text with Linked Open Data**
(Delia Rusu, Blaz Fortuna, Dunja Mladenic)

■ 15:05-15:30: Coffee Break

■ Session 4: Exploiting the Web of Data as a Whole

- 15:30-15:50: **Identifying Relevant Sources for Data Linking using a Semantic Web Index**
(Andriy Nikolov, Mathieu D'Aquin)
- 15:50-16:15: **Re-using Cool URIs: Entity Reconciliation Against LOD Hubs**
(Fadi Maali, Richard Cyganiak, Vassilios Peristeras)
- 16:15-16:40: **Open eBusiness Ontology Usage: Investigating Community Implementation of GoodRelations**
(Jamshaid Ashraf, Richard Cyganiak, Sean O'Riain, Maja Hadzic)

■ Discussion

- 16:40-17:40: **Next Steps and Research Challenges for Linked Data**

■ LOD Gathering / Workshop Dinner

- 19:00: **La Cantina (next to the pool)**

LDOW 2011

29th March 2011, Hyderabad, India

State of the Web of Data

Christian Bizer, Freie Universität Berlin, Germany

Anja Jentzsch, Freie Universität Berlin, Germany

Richard Cyganiak, DERI, Ireland

Statistics based on ...

■ LOD Data Set Catalog on CKAN

- <http://www.ckan.net/group/lodcloud>

■ LOD Dataset Page in ESW Wiki

- <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets>



■ Detailed statistics available at

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

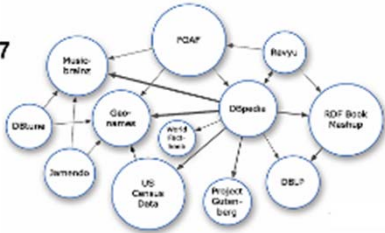
■ Guidelines for adding your own datasets

- <http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKANmetainformation>

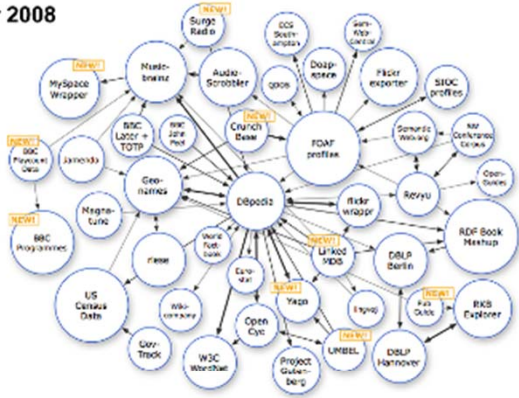
1. Growth

Growth of the Web of Data

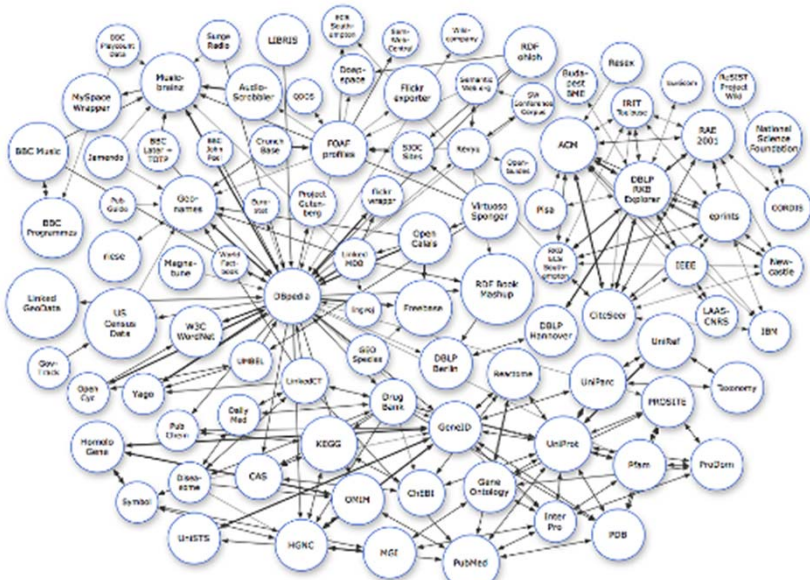
May 2007



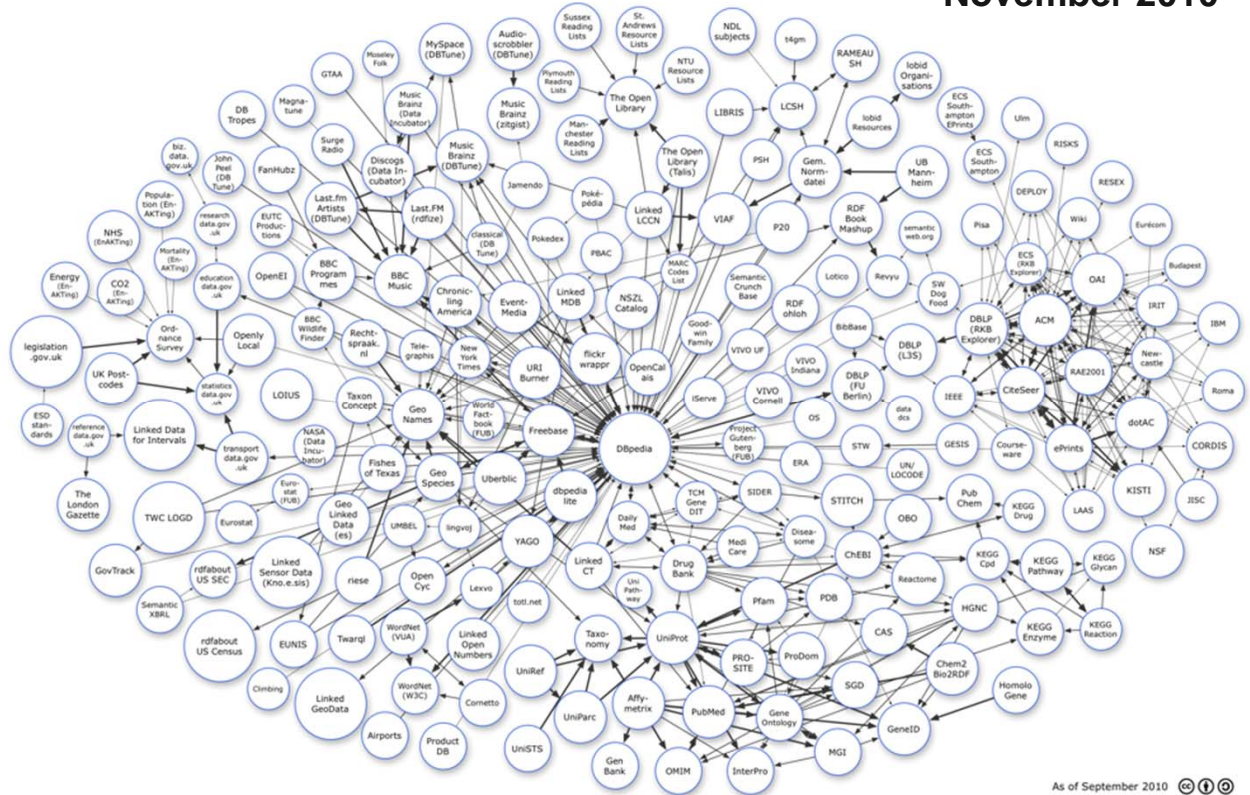
September 2008



July 2009



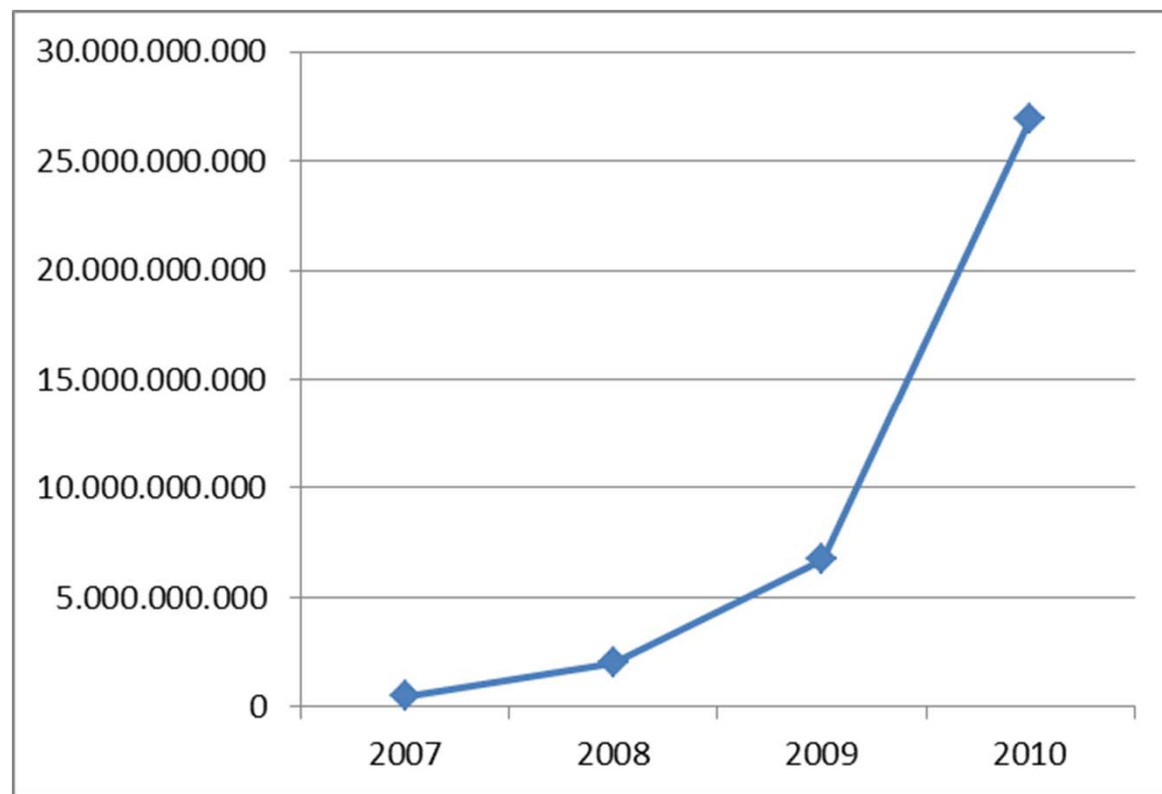
November 2010



As of September 2010 

The Growth in Numbers

Year	Datasets	Triples	Growth
2007	12	500.000.000	
2008	45	2.000.000.000	300%
2009	95	6.726.000.000	236%
2010	203	26.930.509.703	300%



The Growth by Domain 2009-2010

Domain	Triples (June 2009)	Triples (Nov 2010)	Growth
Geographic	3.097.000.000	5.904.980.833	91%
Libraries	212.000.000	2.237.435.732	955%
Media	698.000.000	2.453.898.811	252%
Life sciences	2.429.000.000	2.664.119.184	10%
Cross-domain	214.000.000	1.999.085.950	834%
User-generated	76.000.000	57.463.756	-24%
Government	0	11.613.525.437	-
Total	6.726.000.000	26.930.509.703	300%

Uptake in the Government Domain

HM Government data.gov.uk

Home Blog Data SPARQL Apps Ideas Forum Wiki Resources About

Unlocking innovation
Working with UK Public Sector information and data

Advised by Sir Tim Berners-Lee and Professor Nigel Shadbolt and others, government is opening up data for reuse. This site seeks to give a way into the wealth of government data and is under constant development. We want to work with you to make it better.

We're very aware that there are more people like you outside of government who have the skills and abilities to make wonderful things out of public data. These are our first steps in building a collaborative relationship with you.

Latest news:

- Read about our latest site changes
- find out how the data.gov.uk team has been getting involved with the community
- listen to a Podcast on setting up data.gov.uk

Search Data
Enter keyword(s) **Search**
e.g. education, NHS, crime, transport, environment
Powered by: CKAN

Browse for Data
List all datasets
By Public Body
Common tags

Subscribe by RSS

Community
Log in / Sign up

Local Data Panel

What is the Semantic Web?
Combining different data sources has never been easy but the Semantic Web will enable data to be joined easily across boundaries.
Read more

Digital Engagement
Twitter stream
In line with Cabinet Office guidance

DATA.GOV
EMPOWERING PEOPLE

HOME DATA TOOLS COMMUNITY METRICS DIALOGUE

LINKING OPEN GOVERNMENT DATA
VIEW MORE

Most Popular Datasets

1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. OSHA Data Initiative - Establishment...
5. IT Dashboard - Federal IT Spending (major...

SEARCH OUR CATALOGS
 SEARCH

APPS

National identity management tool
National identity management tool
National identity management tool

With so much government data to work with, developers are creating a wide variety of applications, mashups, and visualizations. From crime statistics by neighborhood to the best towns to find a job to seeing the environmental health of your community—these applications arm citizens with the information they need to make decisions every day. Enjoy these highlights of the hundreds of applications available

COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and other countries. After just one year a community is born around open government data.

Just look at the numbers:

- 6 Other nations establishing open data
- 8 States now offering data sites
- 8 Cities in America with open data
- 236 New applications from Data.gov datasets

SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Thanks to our collaboration with the **Tetherless World Constellation** at the **Rensselaer Polytechnic Institute**, Data.gov is now hosting one of the largest open collections of RDF datasets in the world! Check out some of their semantic mashups **we're featuring** and read our blog entry to learn more about where we are,

■ The EU is pushing Linked Data (LOD2, LATC, EuroStat)

■ W3C eGovernment Interest Group

Uptake in the Libraries Community

■ Institutions publishing Linked Data

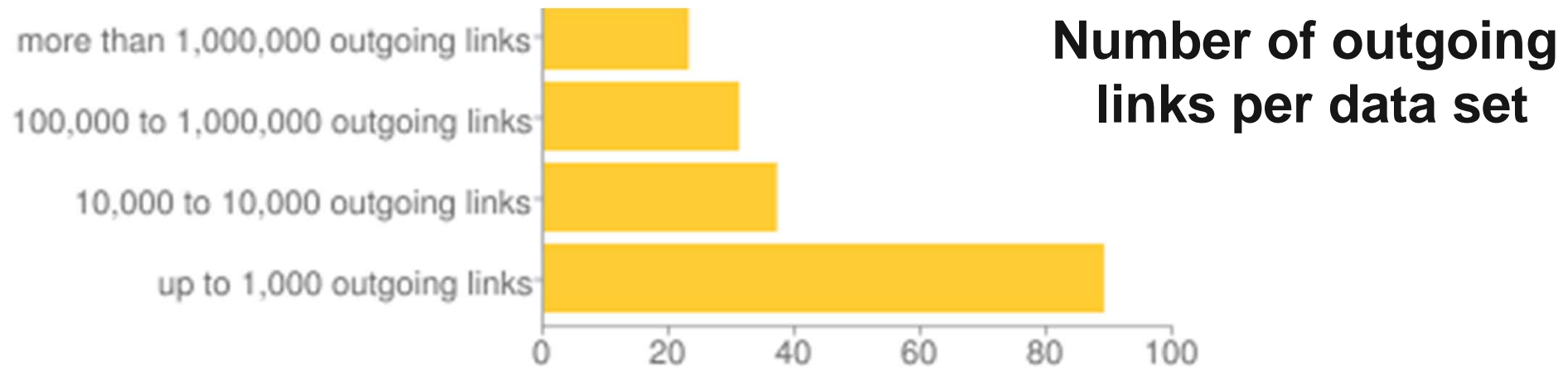
- Library of Congress (subject headings)
- German National Library (PND dataset and subject headings)
- Swedish National Library (Libris - catalog)
- Hungarian National Library (OPAC and Digital Library)
- Deutschen Zentralbibliothek für Wirtschaftswissenschaften (subject headings)

■ The Europeana project is moving towards Linked Data

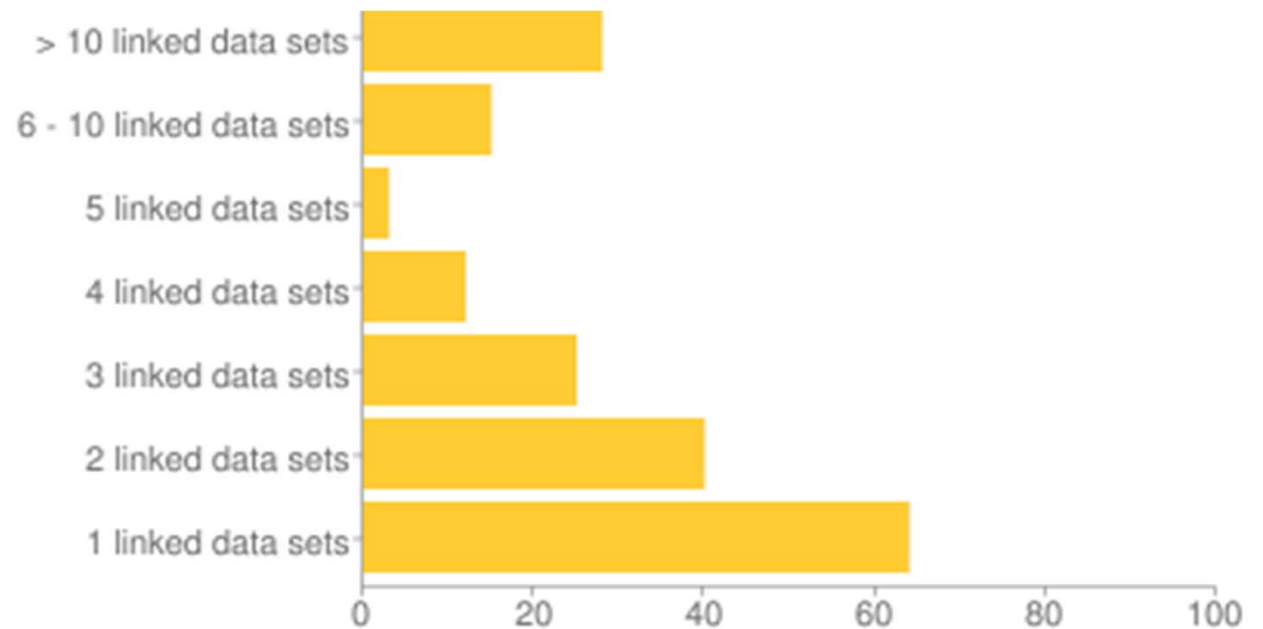
■ W3C Library Linked Data Incubator Group

2. Compliance with Best Practices

RDF Links between Datasets



Number of linked target data sets



Provenance and Licensing Metadata

■ Licensing Metadata

- 18 (9.05 %) out of the 207 data sources provide machine-readable licensing information.
- 181 (90.95 %) out of the 207 data sources do not provide machine-readable licensing information.

■ Provenance Metadata

- 50 (25.25 %) out of the 207 data sources provide machine-readable provenance information.
- 148 (74.75 %) out of the 207 data sources do not provide machine-readable provenance information.

Usage of Common Vocabularies

Prefix	Namespace	Used by
dc	http://purl.org/dc/elements/1.1/	66 (31.88 %)
foaf	http://xmlns.com/foaf/0.1/	55 (26.57 %)
dcterms	http://purl.org/dc/terms/	38 (18.36 %)
skos	http://www.w3.org/2004/02/skos/core#	29 (14.01 %)
akt	http://www.aktors.org/ontology/portal#	17 (8.21 %)
geo	http://www.w3.org/2003/01/geo/wgs84_pos#	14 (6.76 %)
mo	http://purl.org/ontology/mo/	13 (6.28 %)
bibo	http://purl.org/ontology/bibo/	8 (3.86 %)
vcard	http://www.w3.org/2006/vcard/ns#	6 (2.90 %)
frbr	http://purl.org/vocab/frbr/core#	5 (2.42 %)
sioc	http://rdfs.org/sioc/ns#	4 (1.93 %)

Publish Vocabulary Mappings on the Web

■ Map proprietary terms to other vocabularies using

- owl:equivalentClass, owl:equivalentProperty
- rdfs:subClassOf, rdfs:subPropertyOf

```
<http://xmlns.com/foaf/0.1/Person>  
owl:equivalentClass  
<http://dbpedia.org/ontology/Person> .
```

■ Currently 9 (7.32 %) out of the 123 data sources that use proprietary terms provide mappings to other vocabularies for their terms.

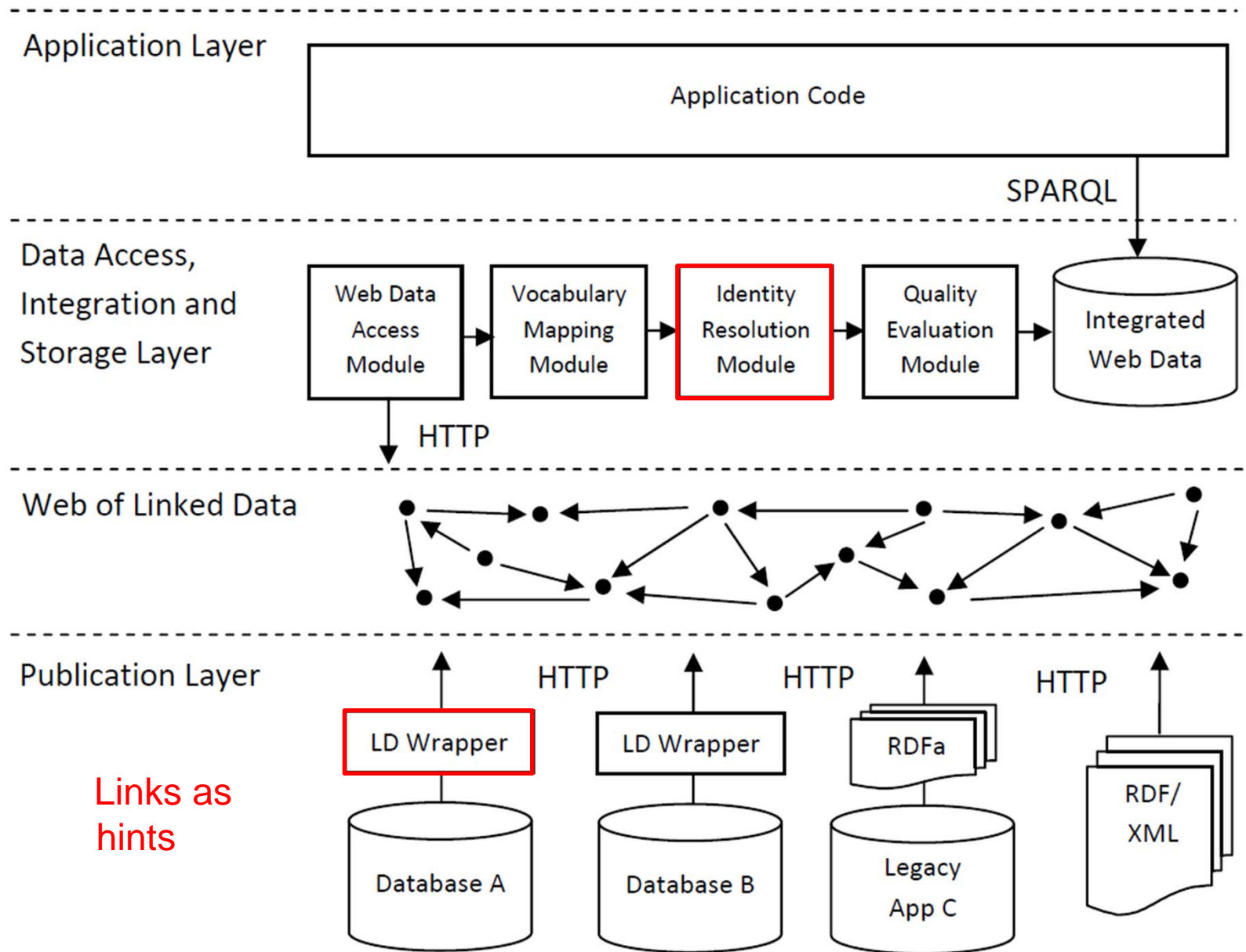
3. Conclusions

For Data Publishers

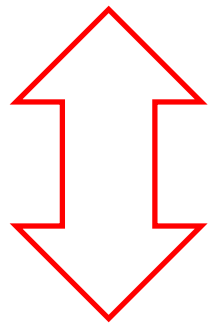
- **Make your data easily consumable by following Best Practices concerning**
 - RDF Links
 - Licensing and Provenance Metadata
 - Widely-used Vocabularies
 - Publication of Vocabulary Mappings on the Web

- **Problem: This requires effort ☹**

Effort Distribution between Publisher and Consumer



Consumer data mines links

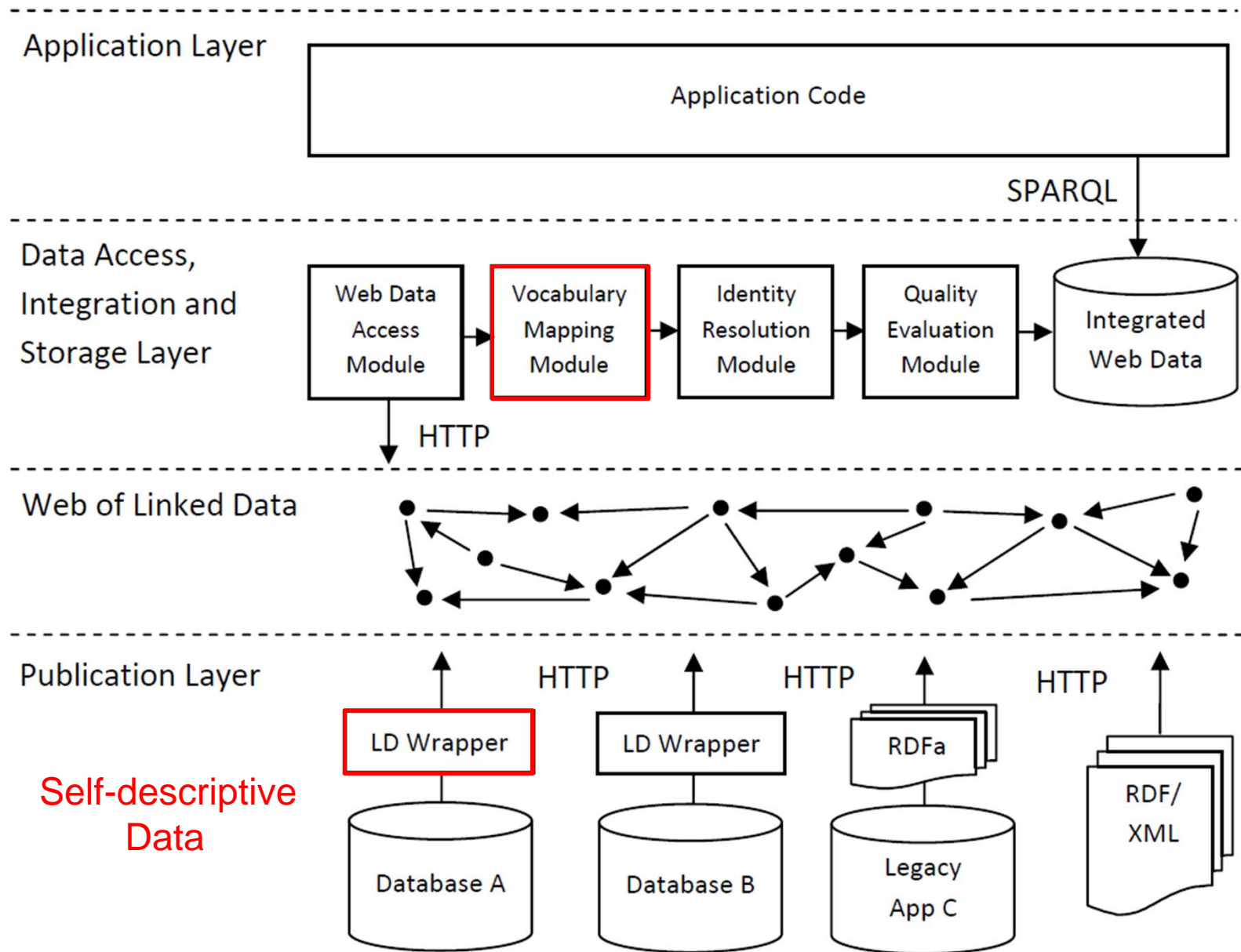


Effort Distribution

Publisher provides links

Links as hints

Effort Distribution between Publisher and Consumer



Consumer data mines mappings

Effort Distribution

Publisher reuses vocabularies and provides mappings

Self-descriptive Data

Somebody-Pays-As-You-Go

The overall data integration effort is **split** between the data publisher, the data consumer and third parties.

■ Data Publisher

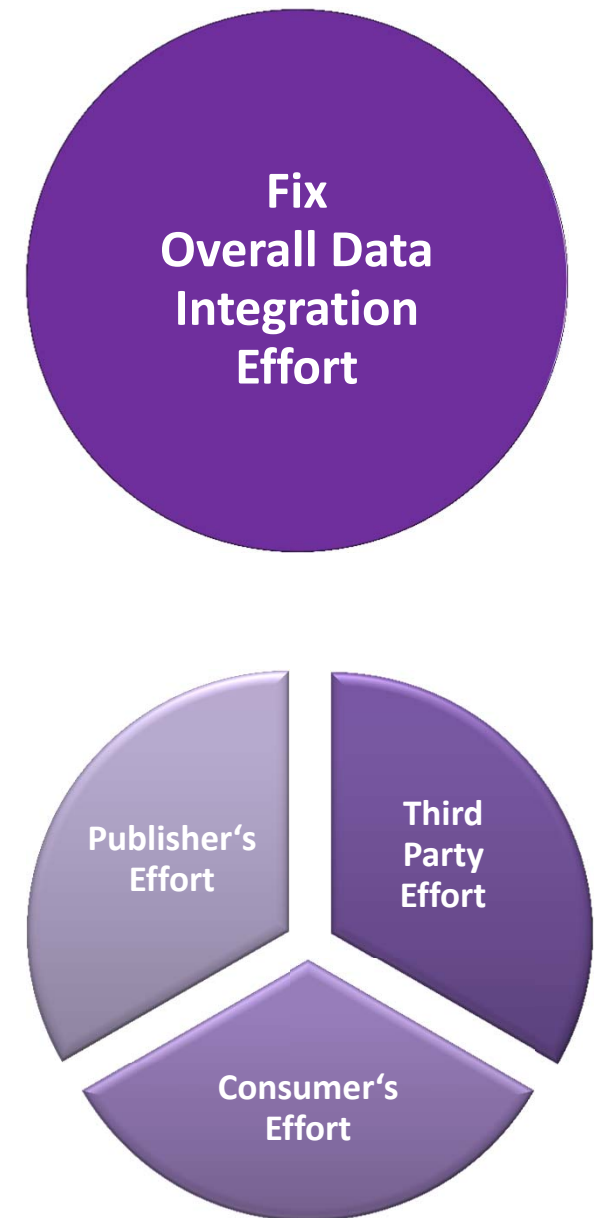
- publishes data as RDF
- publishes data in a self-descriptive fashion
- sets links and publishes mappings

■ Third Parties

- set links pointing at your data
- publish mappings to the Web

■ Data Consumer

- has to do the rest
- using data mining techniques for identity resolution and schema matching



Thanks!

References

- **State of the LOD Cloud Document**
<http://lod-cloud.net/state/>
- **Linked Data - Evolving the Web into a Global Data Space Book**
<http://linkeddatabook.com/>

Lessons Learned and Next Steps

1. Application Architectures (Summary: Tim)

- Lessons Learned
- Future Directions

2. Ontology and Vocabulary Deployment (Summary: Ivan)

- Lessons Learned
- Future Directions

3. Studying the Web of Data (Summary: Nigel)

- What approaches should we use?
- What does Web Science contribute?

4. Is Linked Data over-engineered and too complicated for the real-world? (Summary: Hugh)

- Should the standards be simplified?
- Should the expectations concerning data providers be lowered?