

This detail from Sandro Botticelli's 'Birth of Venus' focuses on the central figure, Venus, as she reclines on a large, scallop-shaped shell. She is depicted with a fair complexion and long, flowing blonde hair, wearing a pinkish-red mantle that is partially draped over her. Her gaze is directed towards the viewer. To her left, a figure in a green garment is partially visible, and to her right, another figure in a pinkish-red mantle is seen. The background is a lush, green landscape with a small body of water in the foreground. The overall style is characteristic of the Italian Renaissance, with soft, flowing lines and a focus on idealized beauty.

HPI Hasso Plattner Institut

Hasso Plattner Institute for IT Systems Engineering
University of Potsdam
Autumn 2015

How to get Data from the Web?

- Data can only be found on the Web, if it is available at some website



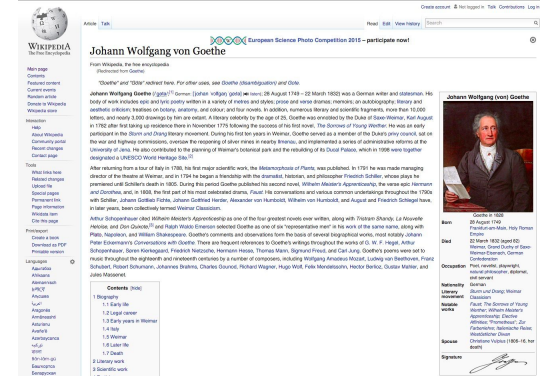
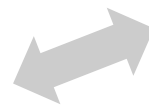
Database

JDBC



Web Server

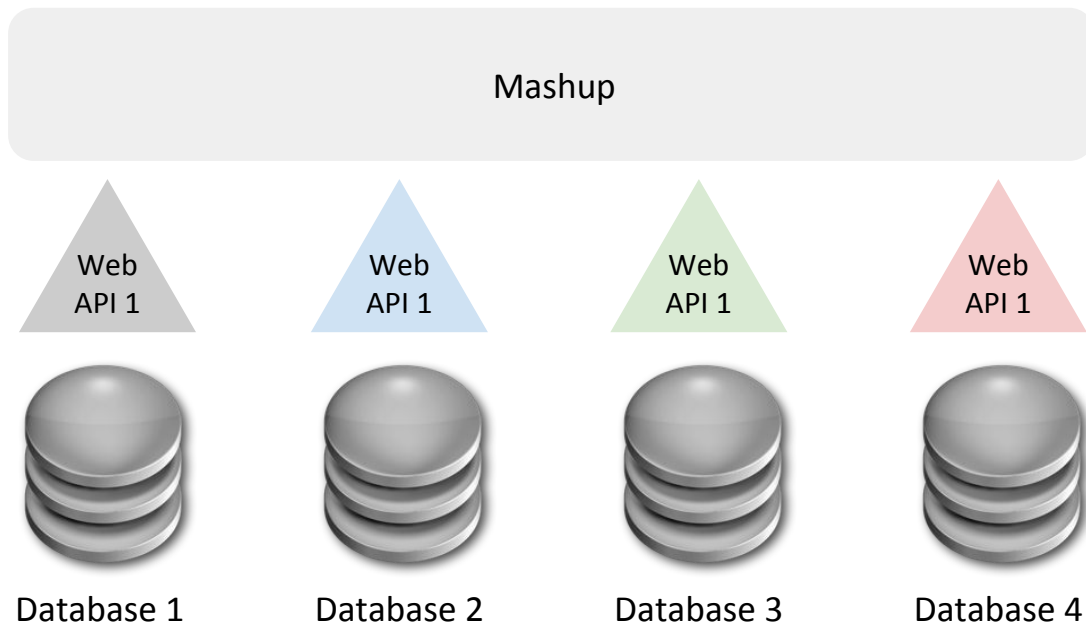
HTTP



Browser

How to get Data from the Web?

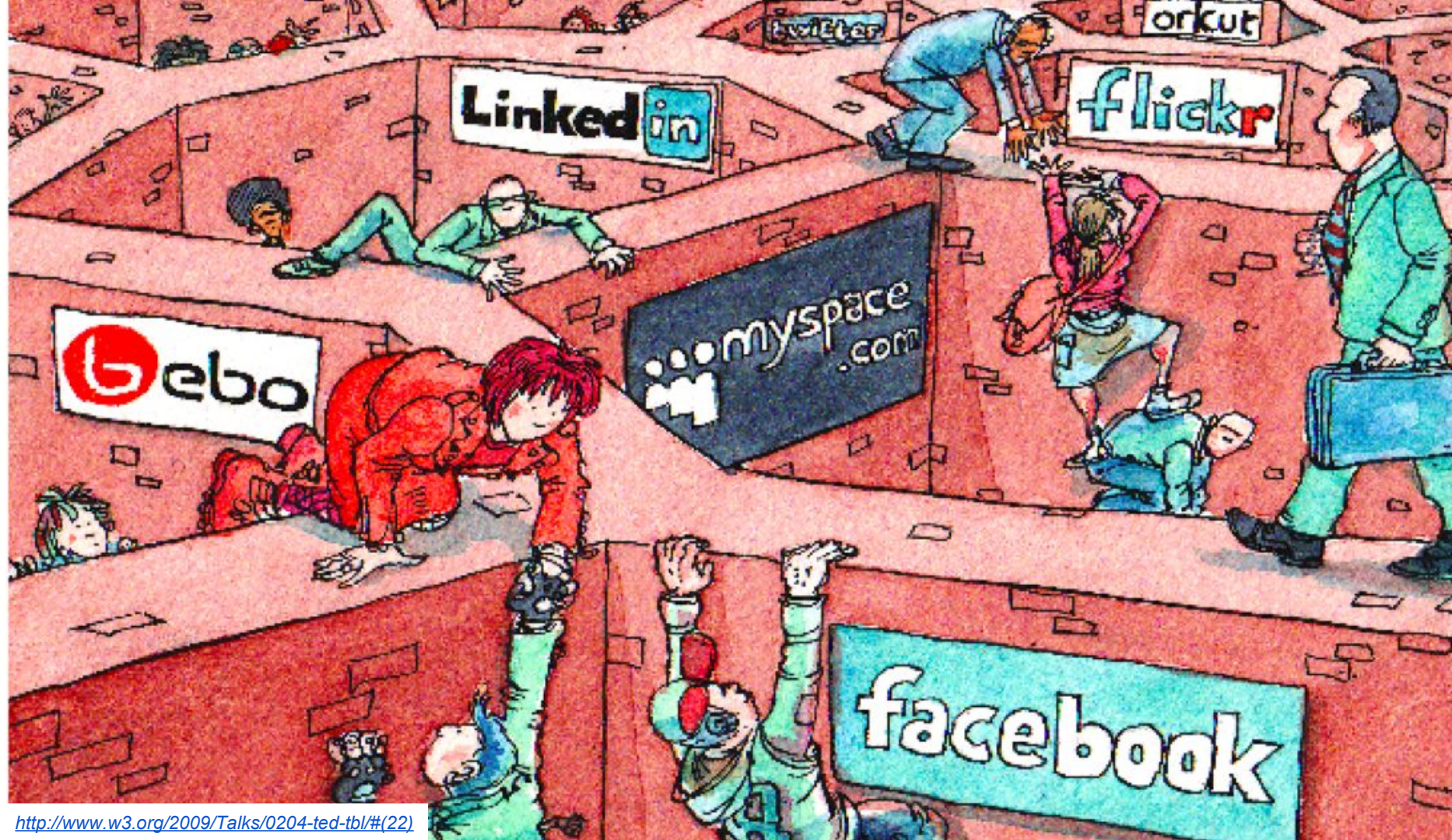
- There is a a number of different (proprietary) **Web APIs**, data exchange formats and **Mashups** on top of that



In the Web today...

- Data is locked up in small data islands
- Other applications usually cannot access this data...

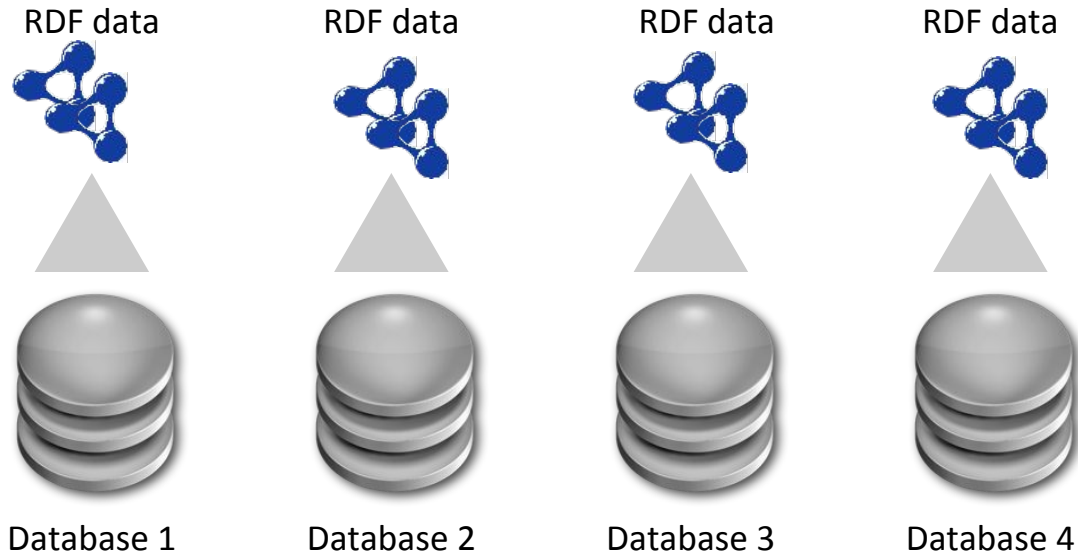


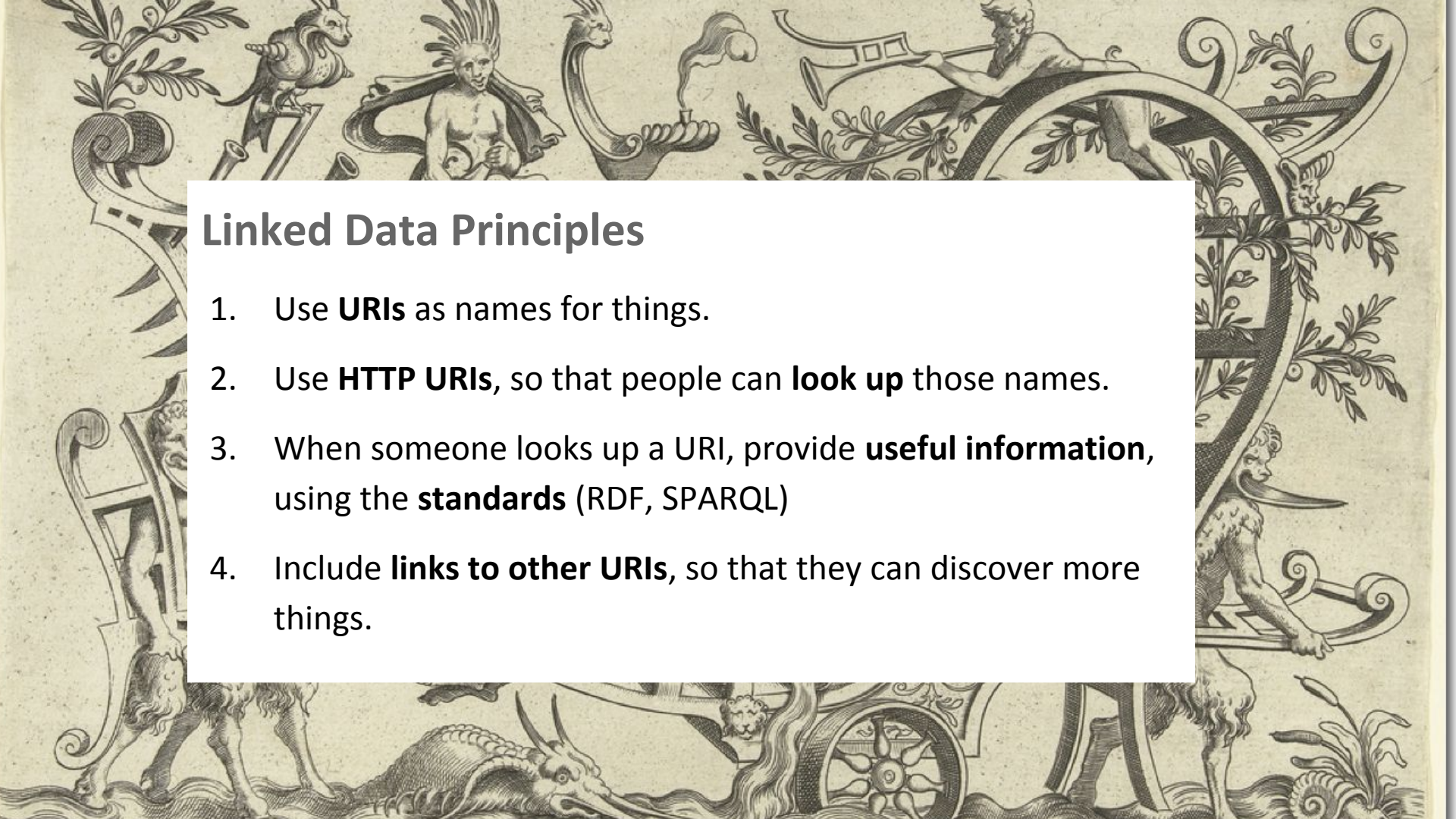


How to get rid of Closed Data Islands?

- **Apply Semantic Web technologies**

- to publish (structured) data on the web
- to draw connections from one data source to data from other data sources



The background of the slide is a detailed, sepia-toned illustration. It features a variety of figures and symbols: a unicorn-like creature on the left, a figure with a large feathered headdress in the upper center, a hand holding a smoking pipe in the upper middle, a figure playing a trumpet on the right, and a figure with a long, curved horn or staff at the bottom right. The scene is framed by elaborate, swirling acanthus leaves and other decorative elements.

Linked Data Principles

1. Use **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.

Linked Data and the ‘Web of Data’

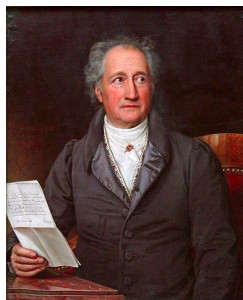
- Term refers to an idea originally from Tim Berners-Lee
(Tim Berners-Lee, *Linked Data*, 2006, <http://www.w3.org/DesignIssues/LinkedData.html>)
- Set of **best practices** for publication and linking of structured data on the web
- Basic assumption: The value of data on the web increases when they are connected to other data sources



The Web of data is about
a data
and naming
model on the Web

Linked Data Principles (1/4)

1. Use **URIs** as names for things.
 - URIs do not only identify documents but also arbitrary objects of the real world as well as abstract concepts



<https://viaf.org/viaf/24602065/>



<http://musicbrainz.org/artist/5700dcd4-c139-4f31-aa3e-6382b9af9032>

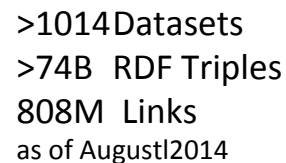
<http://www.imdb.com/title/tt0017136/>

Linked Data Principles (2/4)

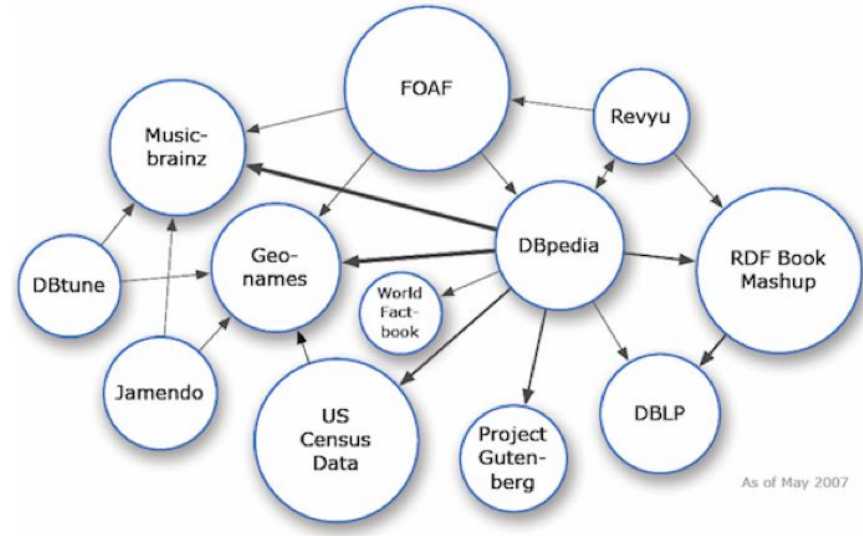
2. Use **HTTP URIs**, so that people can **look up** those names.
 - HTTP URIs (URLs) as globally unique names enable dereferencing of associated information in the Web
 - via http Content Negotiation
 - 303 URIs
http Response Code 303 ‚See Other‘ (redirect)
 - Hash URIs
`http://example.com/Harald#me`

3. When someone looks up a URI, provide **useful information**, using the **standards** (RDF, SPARQL)
 - RDF as universal data model for publishing structured data on the Web
 - Make all URIs in the RDF graph **dereferencible**
 - Avoid RDF constructs that cause problems in Linked Data context
 - RDF Reification
 - RDF Collections und Containers
 - unnamed Blank Nodes

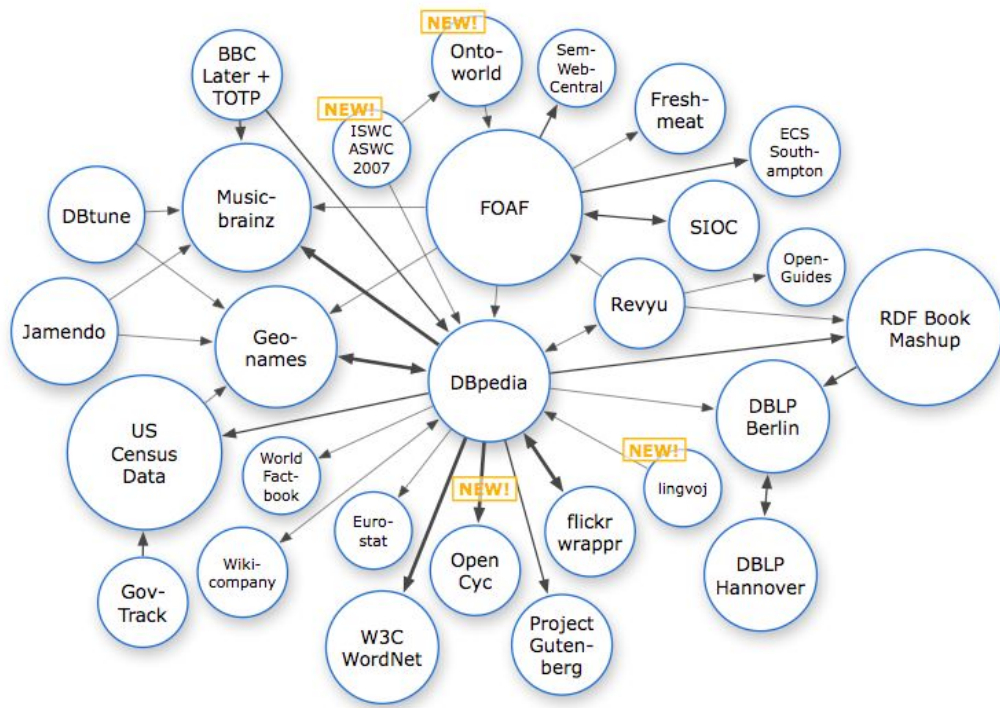
4. Include **links to other URIs**, so that they can discover more things.
 - Link RDF references among data between different data sources, to find information related by content
 - **Relationship Links**
Links to external LOD Entities related with the original entity
 - **Identity Links**
Links to external LOD Entities referring to the same object or concept
 - **Vocabulary Links**
Links to definitions of the original entity



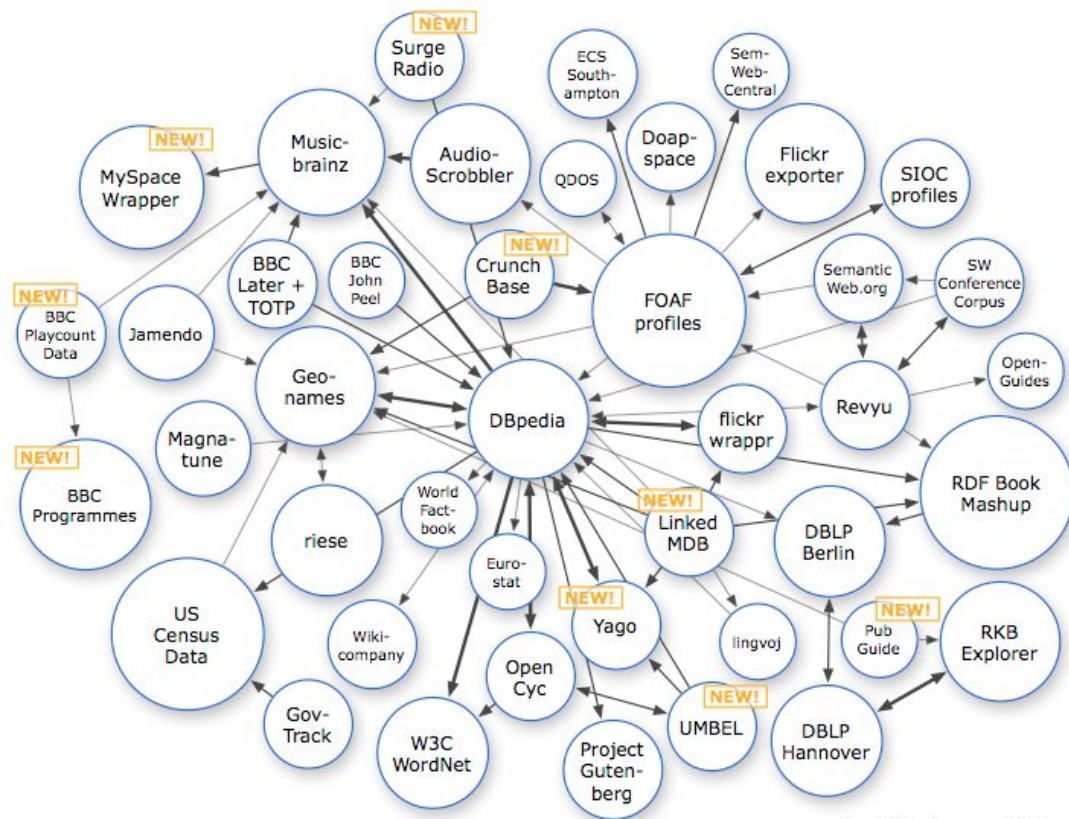
The Development of the Web of Data



The Development of the Web of Data

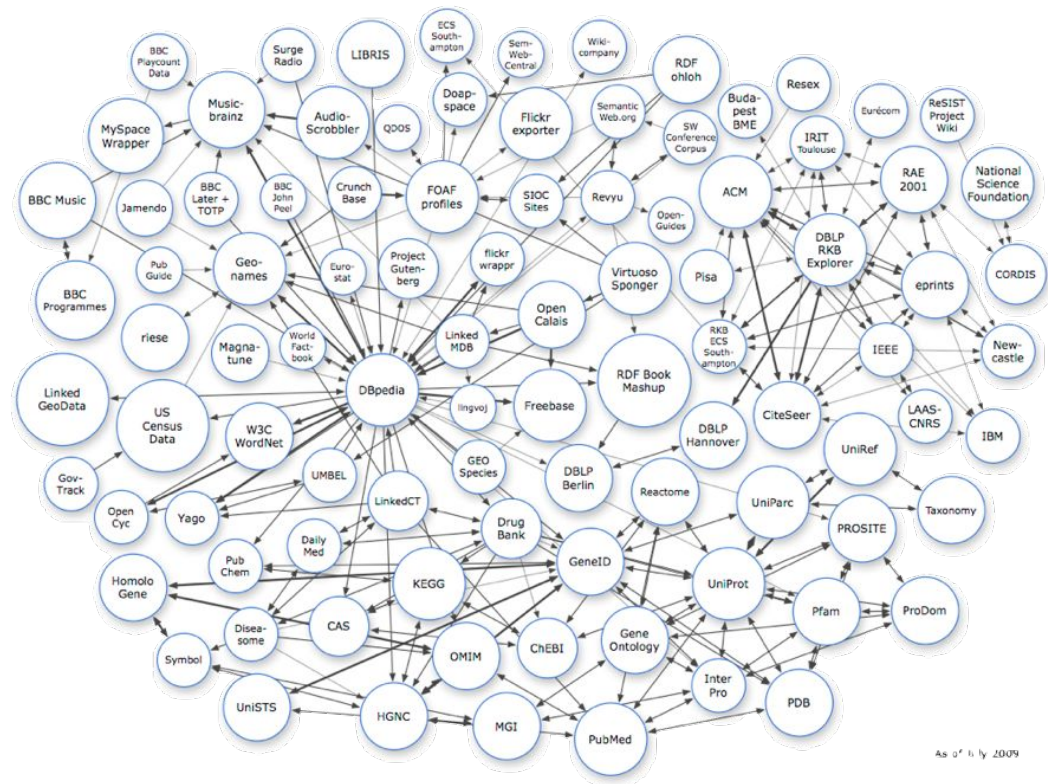


The Development of the Web of Data



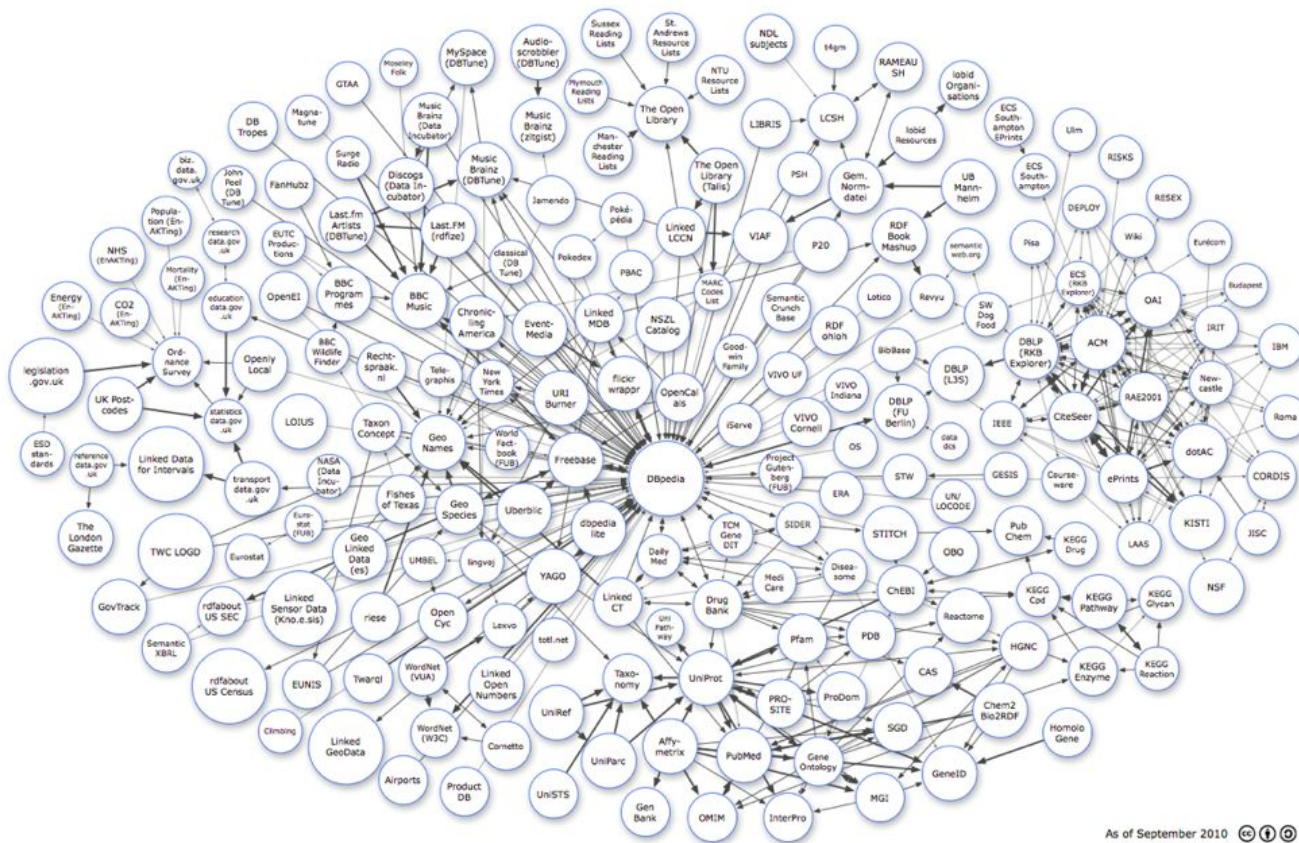
As of September 2008

The Development of the Web of Data



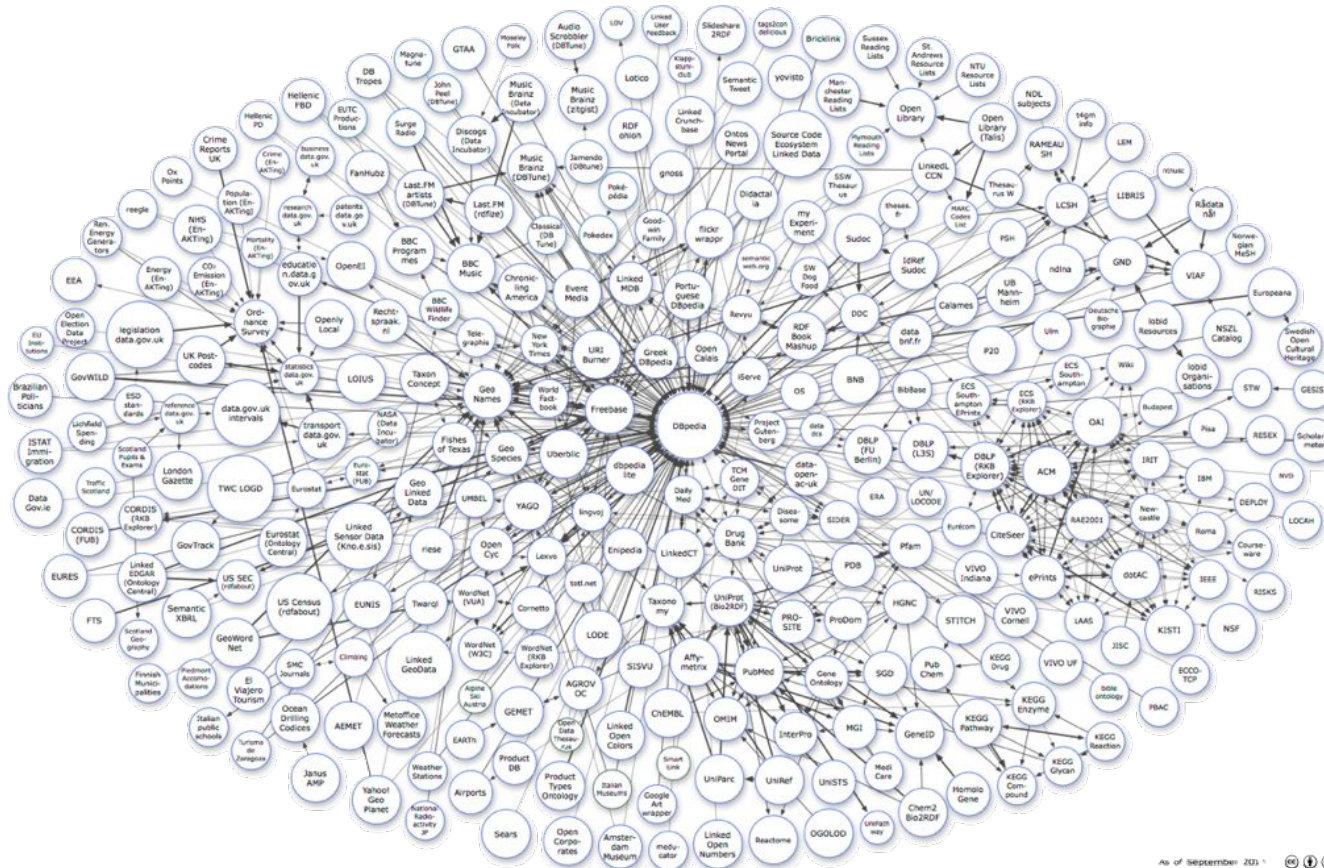
As of July 2009

The Development of the Web of Data

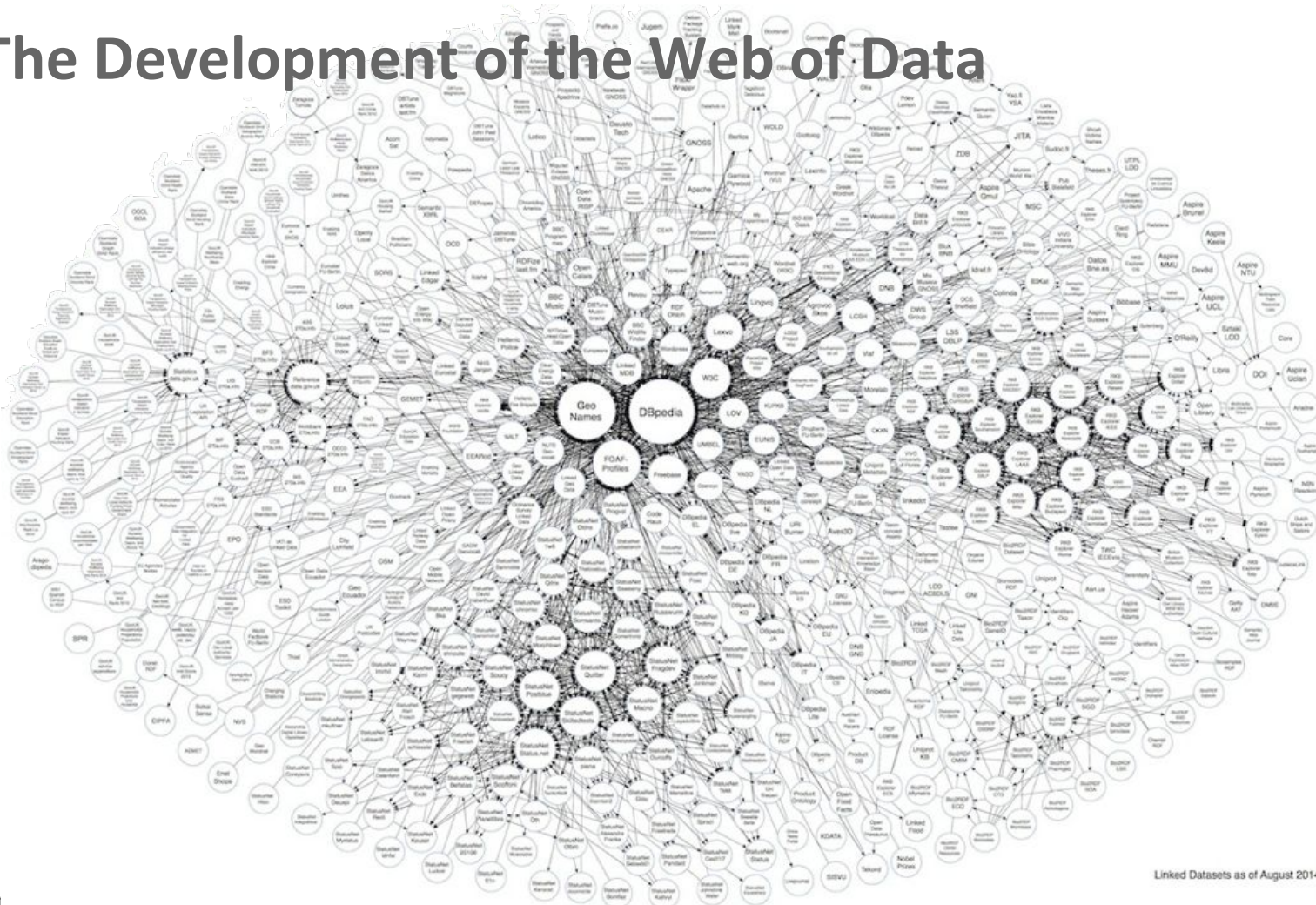


As of September 2010 

HPI Hasso Plattner Institut



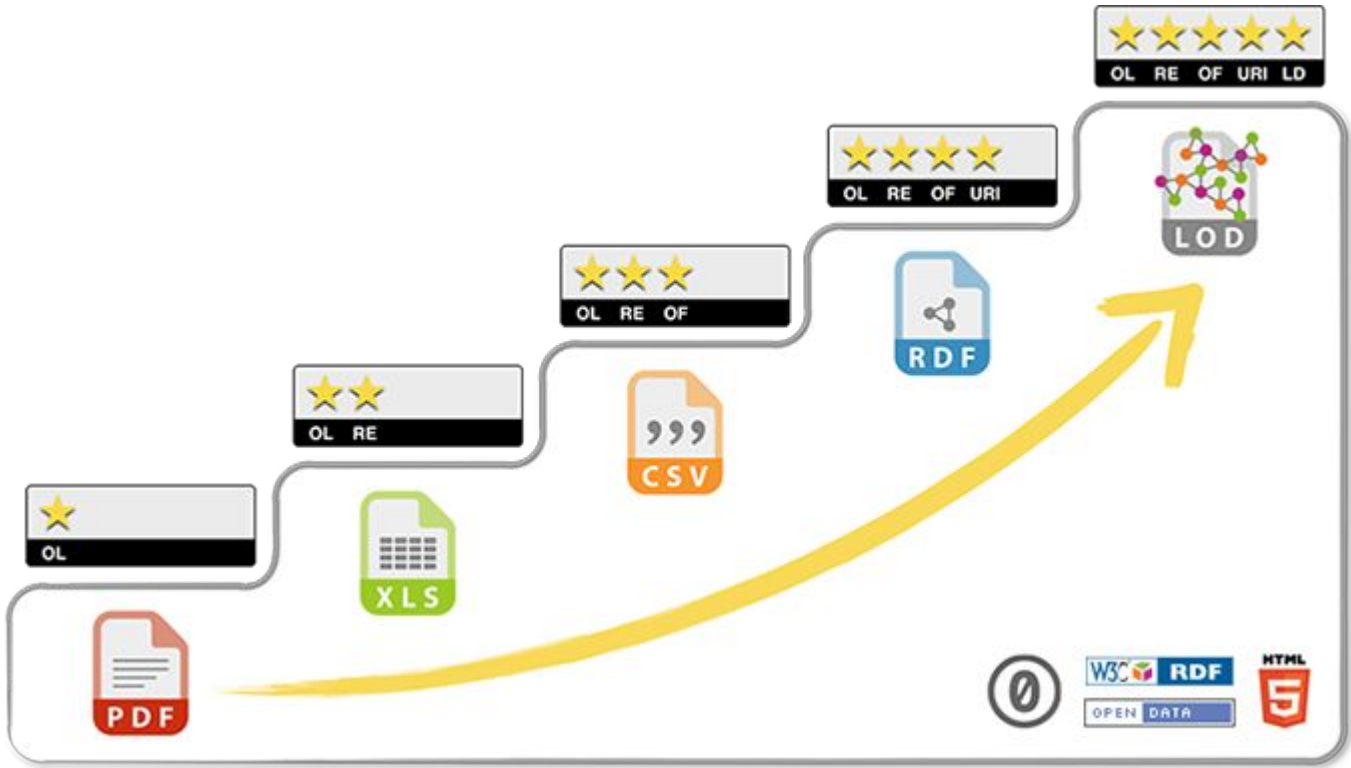
The Development of the Web of Data



Linked Open Data

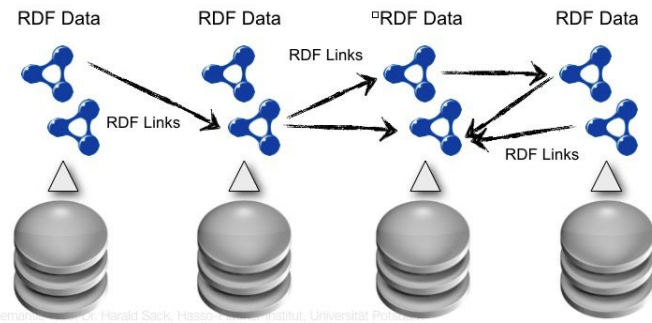
- Public Linked Data resources in the Web, licensed as Creative Common CC-BY
- Tim Berners-Lee's 5-Star Criteria for Linked Open Data
 - ★ Available on the web (whatever format) but with an **open licence**, to be Open Data
 - ★★ Available as **machine-readable structured data**
(e.g. excel instead of image scan of a table)
 - ★★★ as (2) plus **non-proprietary format** (e.g. CSV instead of excel)
 - ★★★★ All the above plus: use **open standards from W3C**
(RDF and SPARQL) to identify things, so that people can point at your stuff
 - ★★★★★ All the above, plus: **link your data to other people's data** to provide context

Linked Open Data



Advantages of Linked Open Data vs. APIs

- **Simple** and **generic API** for various heterogeneous data sources enables **simple reuse** and **data sharing** among applications
- **RDF Data model** guarantees (simple) **extensibility**
- **Transport via http**, standard Port 80, prevents firewall adaption
- **Ontologies** enable **meaningful connections** between data sources
- **Reasoning** over Linked Data enables to **generate new knowledge**, i.e. inference from implicit to explicit knowledge



Lecture 6: Knowledge Engineering