

7th Central and Eastern European
Software Engineering Conference
in Russia - CEE-SECR 2011



October 31 – November 3, Moscow

Linked data driven development

Jan Božnik*, Vili Podgorelec*, Marjan Heričko*, Črt Gerlec*

*University of Maribor, Slovenia

Sharing is caring



Sharing information is the basic principle of communication.

Sharing is caring

- Documents are the main carriers of information in a digital world.

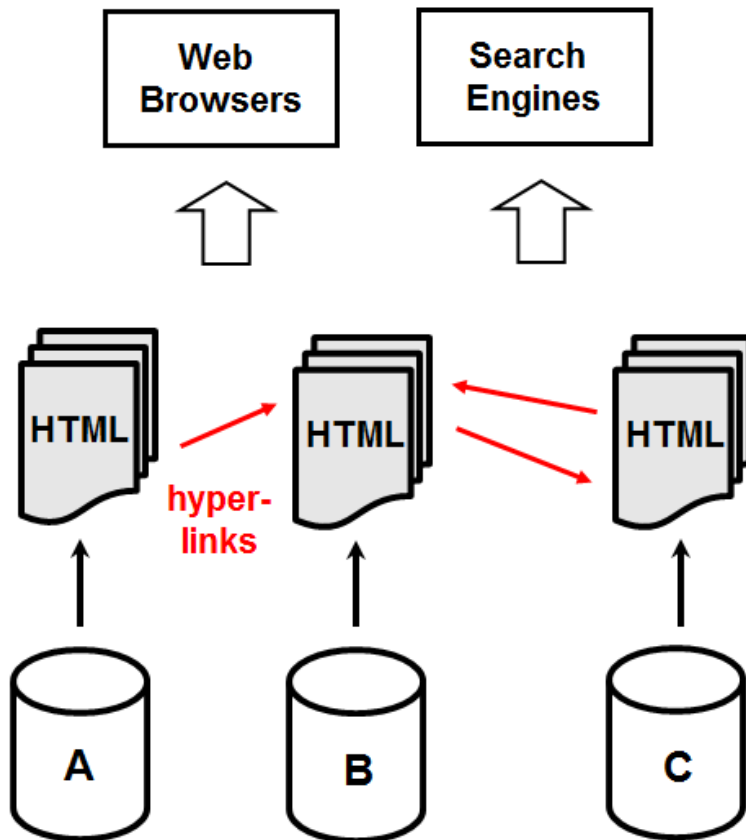


- Humans are aware of the meaning of information.

- How can we transfer this meaning awareness to computers?



Classic Web



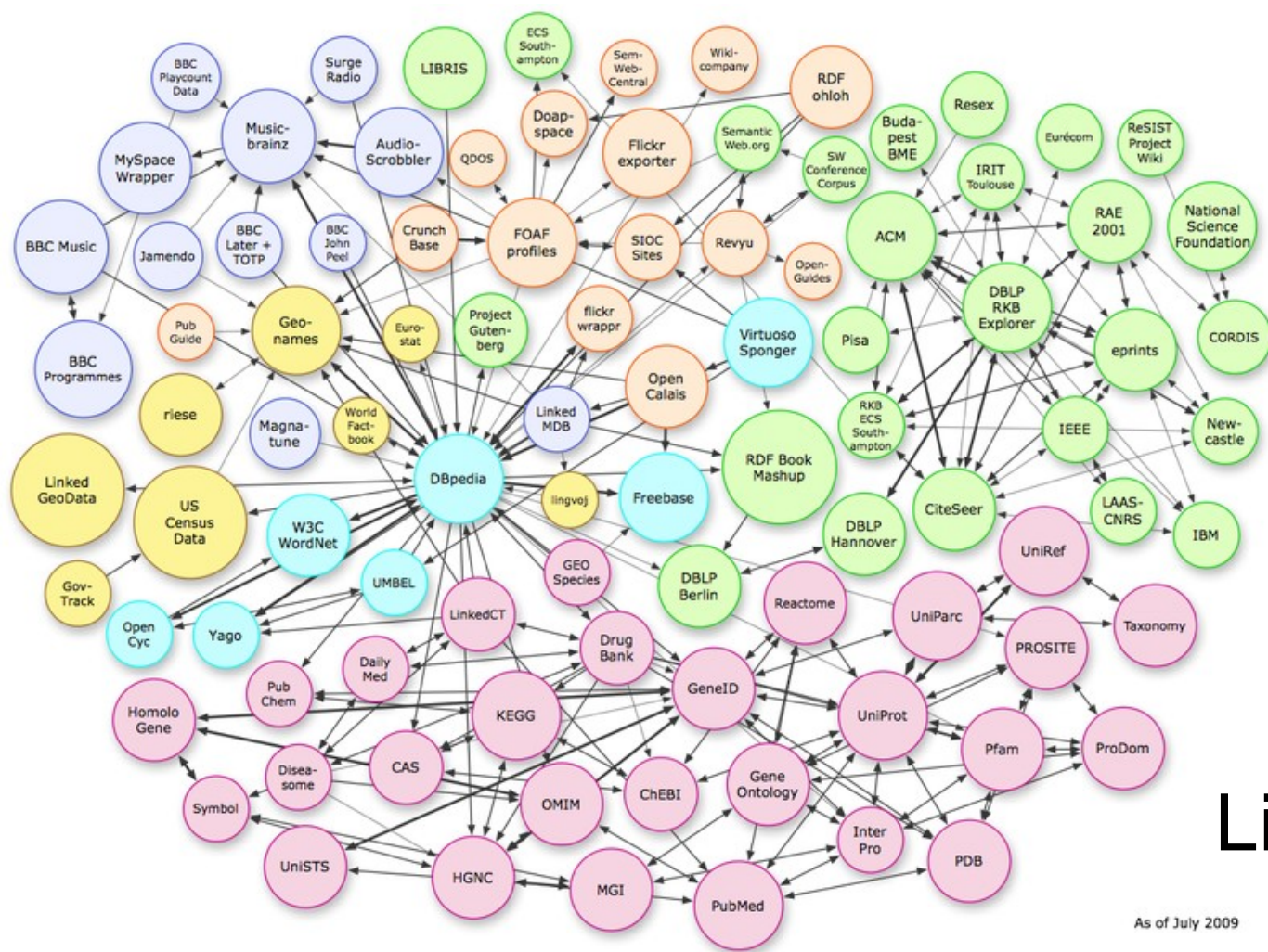
- The Document Web
 - Single Global Information Space
 - Unified Resource Locators as
 - globally unique IDs
 - retrieval mechanism
 - Hypertext Markup Language (HTML) as a widely used content format
 - Hypertext Transfer Protocol (HTTP) as universal access mechanism
 - Hyperlinks between pages

The problem



- Web content is loosely structured
- Difficult for applications to make meaningful connections

The solution is:



Linked
data

As of July 2009

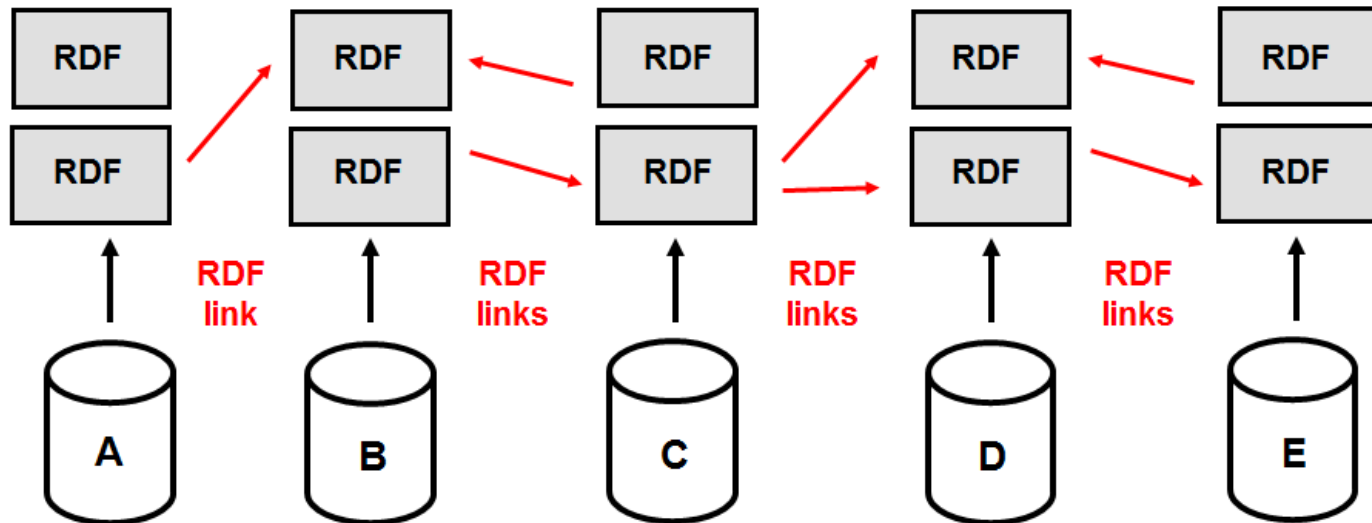
What is Linked Data?

- Set of best practices for publishing structured data on the web.



Linked Data

- Setting links between data from one data source to data within other data sources



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 RDF information.
4. Include RDF statements that link to other URIs so that they can discover related things.

Tim Berners-Lee 2007

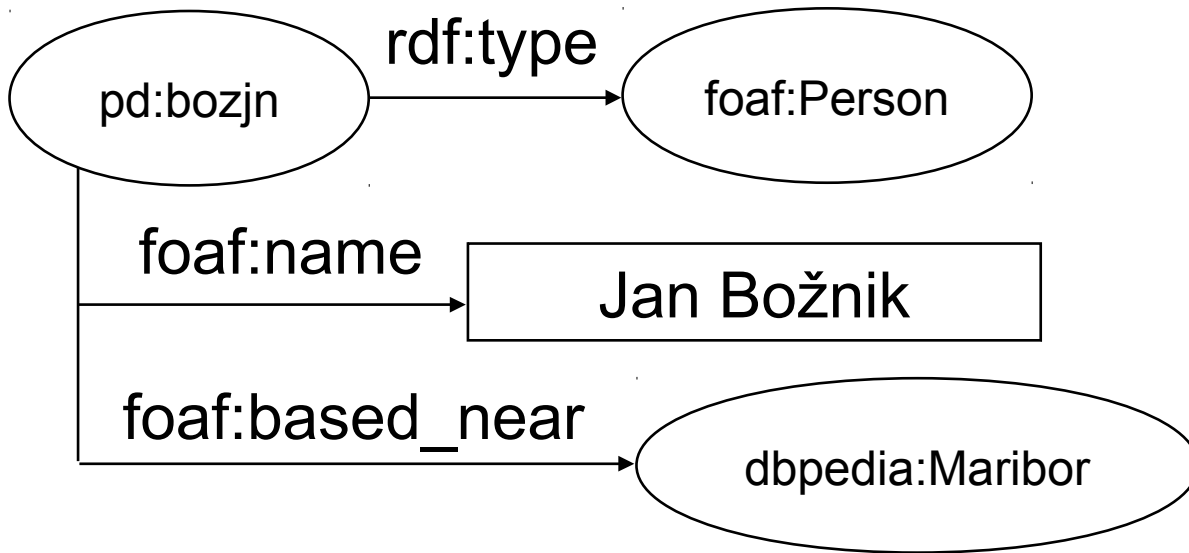
<http://www.w3.org/DesignIssues/LinkedData.html>

Linked Data principles - continued

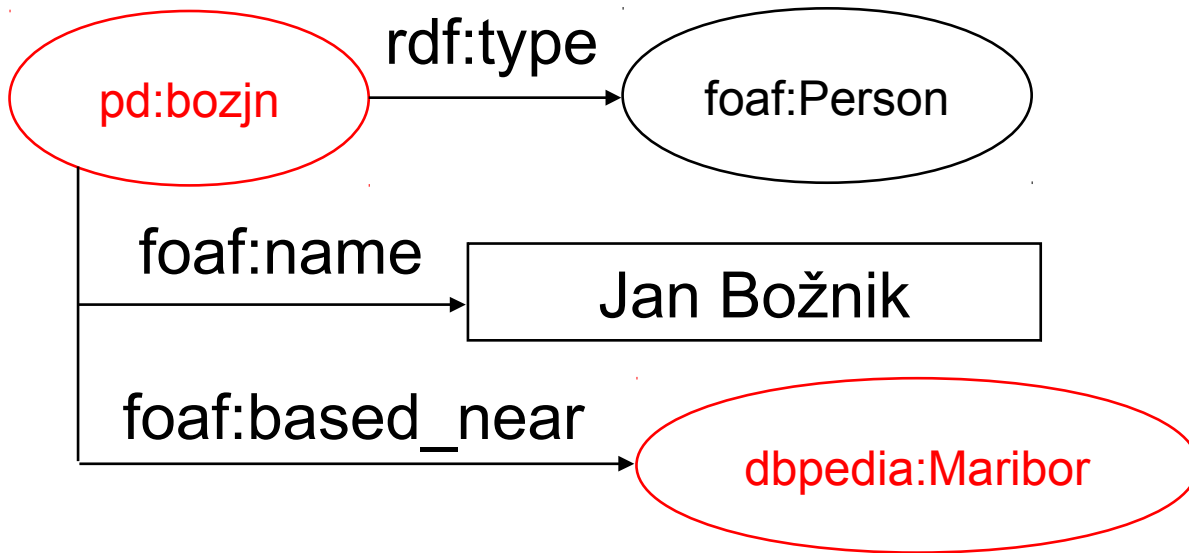
- Anyone can publish data to the Web of Linked Data
- Entities are connected by links
 - creating a global data graph that spans data sources and enables the discovery of new data sources.
- Data is self-describing
 - If an application encounters data represented using an unfamiliar vocabulary, the application can resolve the URIs that identify vocabulary terms in order to find their RDFS or OWL definition.
- The Web of Data is open
 - meaning that applications can discover new data sources at run-time by following links.

By Christian Bizer

The RDF Data Model



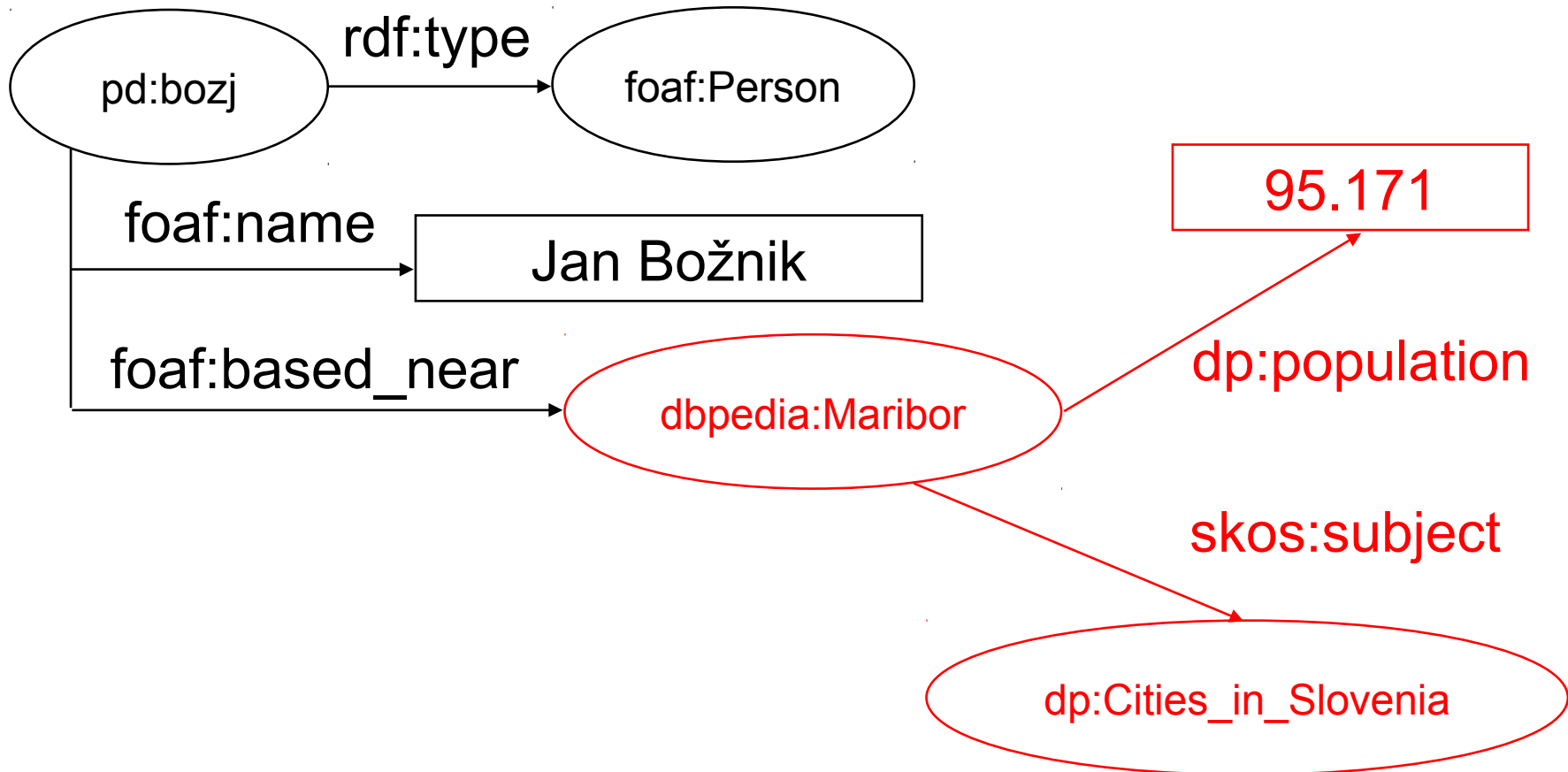
Data items are identified with HTTP URIs



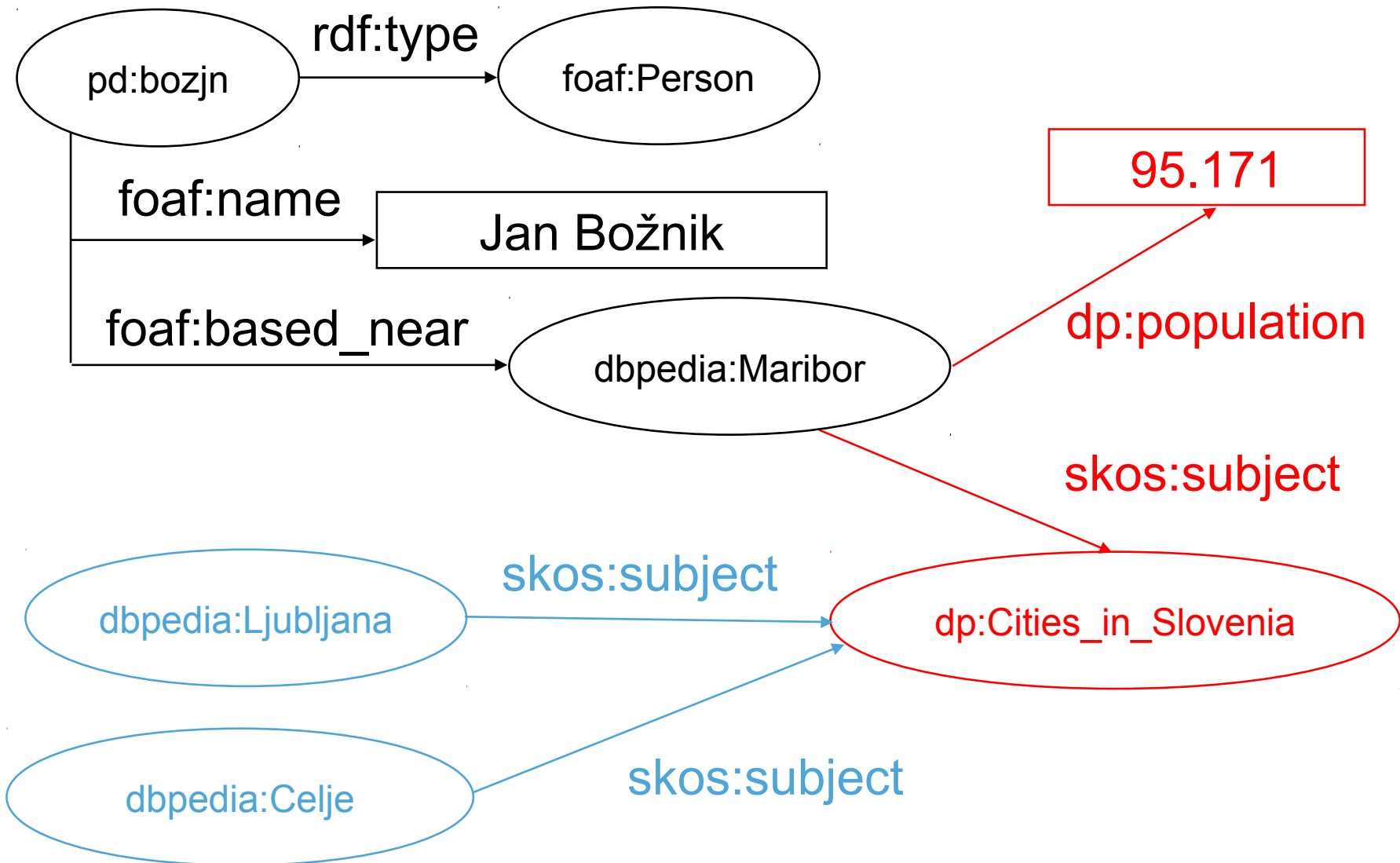
pd:bozjn = <http://jan.boznik.si/foaf.rdf#bozjn>

dbpedia:Maribor = <http://dbpedia.org/resource/Maribor>

Resolving URIs over the Web



Dereferencing URIs over the Web



Vocabularies

The Dublin Core Metadata
Initiative (DCMI)

The Friend-of-a-Friend
(FOAF)

The Semantically-Interlinked
Online Communities (SIOC)

The Programmes Ontology

The Music Ontology

The OAI Object Reuse and
Exchange

The Basic Geo (WGS84)

The Good Relations
Ontology

The Web of Trust (WOT)

The Bibliographic Ontology
(BIBO)

The Review Vocabulary

The Creative Commons
(CC)



Publishing data

- Incoming and outgoing links
- Data consumption
- Search engines
- APIs for discovering RDF files



Application infrastructure patterns

1

Applications that implement the **crawling pattern** in the sense of **traversing** the Web of Data and seeking RDF links.

2

Applications that implement the **dereferencing pattern**, such as Linked Data browsers Disco hyperdata browser, Tabulator and Marbles.

3

Applications that implement **query federation pattern**.

Examples of Linked data driven applications

- DBpedia

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[[ a ?Concept]}
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: (The CXML output is disabled, see [details](#))

Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Examples of Linked data driven applications

- BBC Programmes

The screenshot displays the BBC Programmes website. At the top, there is a navigation bar with the BBC logo, a 'Sign in' link, and links for News, Sport, Weather, iPlayer, TV, Radio, and More. A search bar is located on the right side of the navigation bar. Below the navigation bar, the word 'Programmes' is prominently displayed in a large, white font against a blue background. Underneath, there is a secondary navigation bar with links for Home, A to Z, Genres, Formats, and Topics. The main content area is divided into several sections: a 'Welcome' section with a brief description of the site's purpose and links to more information; a 'Programmes for you' section with a description of the content and links to explore further; an 'A to Z of programmes' section with a navigation bar for letters and numbers; a 'Categories' section with a description of how programmes are categorized; and two columns of lists for 'Genres' and 'Formats'.

BBC Sign in News Sport Weather iPlayer TV Radio More Search the BBC

Programmes

Home A to Z Genres Formats Topics

Welcome

This site aims to ensure that every TV & Radio programme the BBC broadcasts has a permanent, findable web presence.

Find out more [about this site](#), read our [commitment to accessibility](#) and discover [our plans for the future](#).

Programmes for you

Whether it's [games & quizzes](#), all new [drama](#) or the full suite of [music](#), we've got the programmes for you.

Explore the links below to find programmes you're interested in. Browse by name, category or schedule.

A to Z of programmes

Use the A to Z to find a programme by name

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Categories

We categorise our programmes by genre and format

Genres

- Children's
- Comedy
- Drama
- Entertainment
- Factual
- Learning
- Music
- News
- Religion & Ethics
- Sport
- Weather

Formats

- Animation
- Appeals
- Bulletins
- Discussion & Talk
- Docudramas
- Documentaries
- Films
- Games & Quizzes
- Magazines & Reviews
- Makeovers
- Performances & Events
- Phone-ins
- Readings
- Reality
- Talent Shows

Examples of Linked data driven applications

- Faviki

[WHAT IS FAVIKI?](#)

[COMMON TAGGING](#)

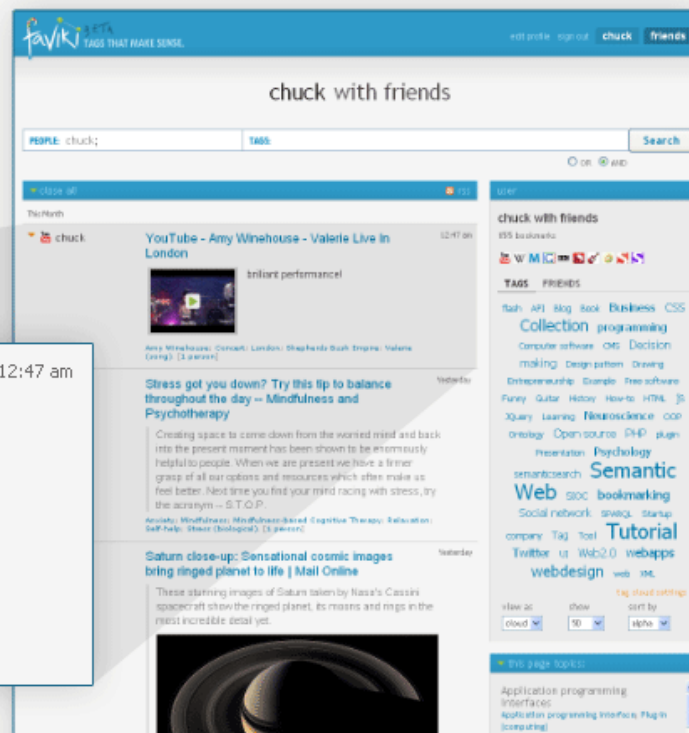
[ORGANIZE BETTER](#)

[TAG IN YOUR LANGUAGE](#)

[JOIN](#)

Faviki is a social bookmarking tool that lets you use [Wikipedia](#) concepts as tags.

Faviki allows you to keep your own tags and connect them to common, universal concepts from the world's largest collection of knowledge!



YouTube - Amy Winehouse - Valerie Live In London

12:47 am



brilliant performance!

Amy Winehouse; Concert; [London](#); [Shepherds Bush Empire](#); Valerie (song). [1 person]

London is the capital of the United Kingdom and the constituent country of England, and is the largest city in the European Union. An important settlement for two millennia, London's history goes ... source: Wikipedia

Connecting the classic Web and Linked Data

- Annotate Web documents with Linked Data URIs

```
<http://www.secr.ru/lang/en-en/talks/linked-data-driven-  
development> dc:subject  
<http://dbpedia.org/resource/Semantic_web> .
```

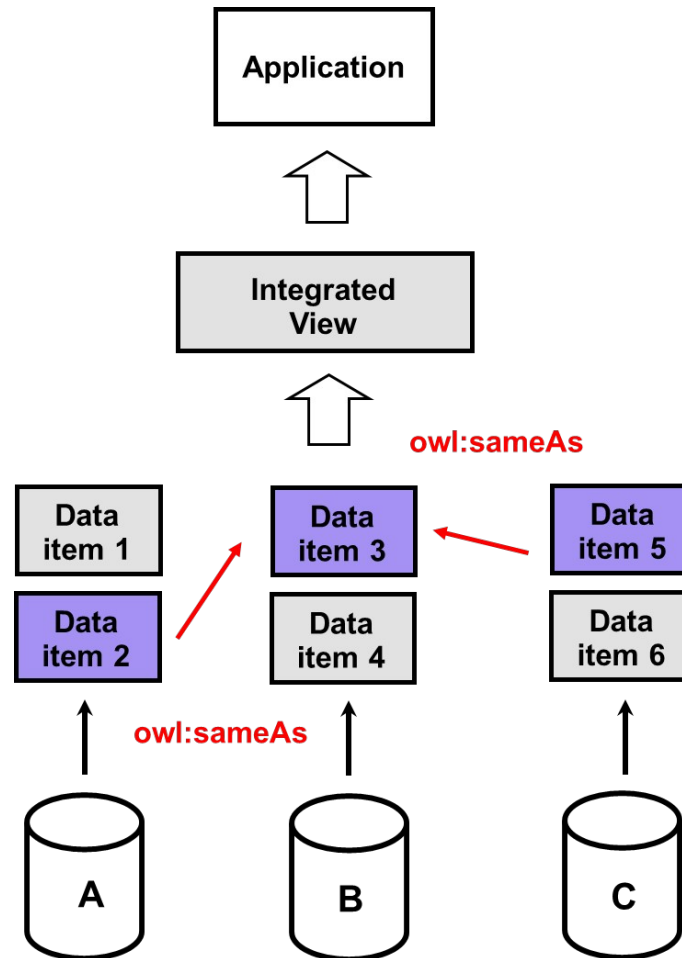
- (Semi-) Automated Annotation Services using Named Entity Recognition
 - Open Calais (Thomsons Reuters) for news
 - Zemanta for blog posts



Next steps

- More data is becoming available
 - government data
 - statistical data
 - bibliographic data
- What is still missing?

Linked Data Fusion



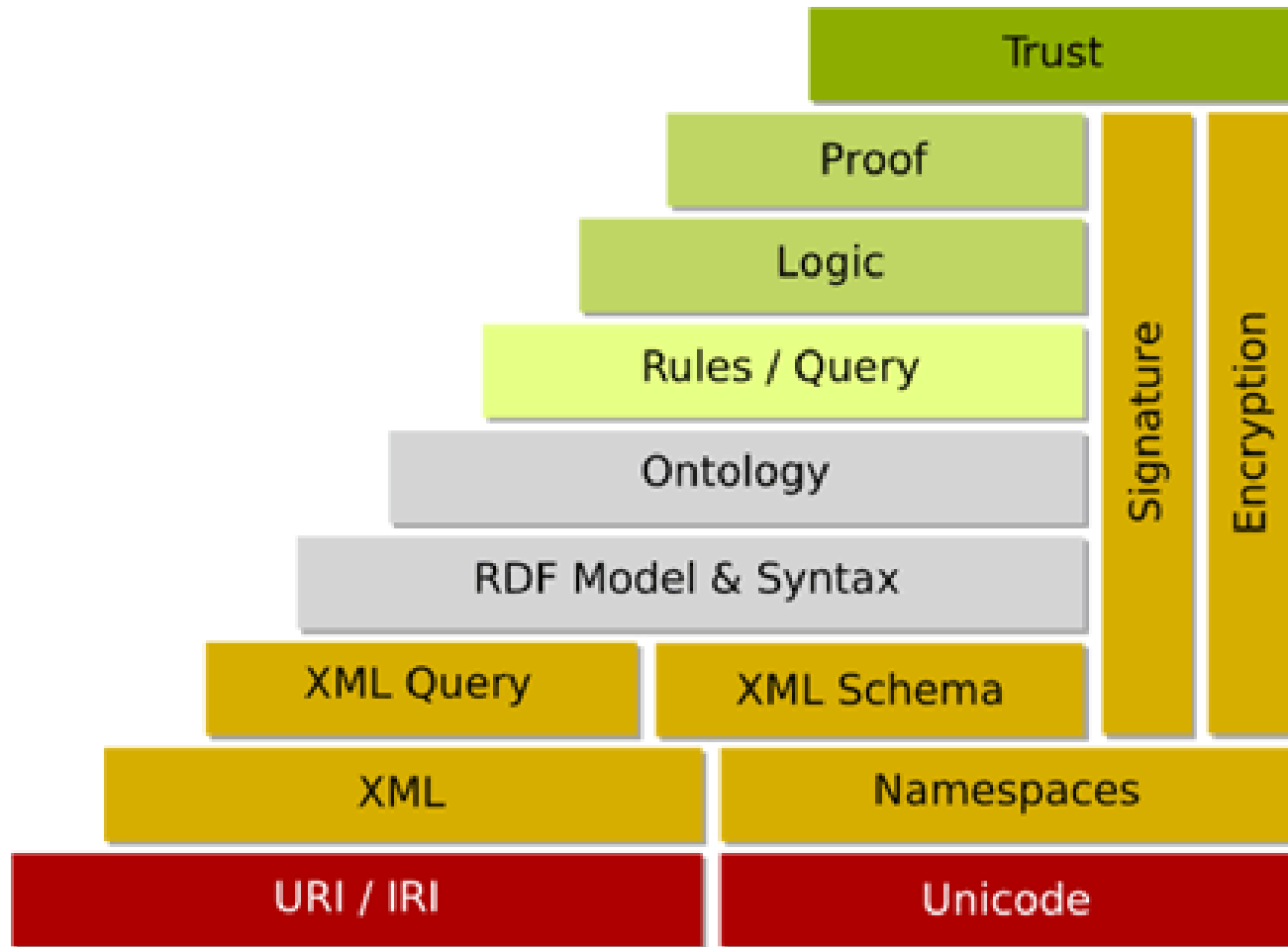
Linked Data Fusion - Requirements

1. Map data into a single schema
 - so that data can be rendered and queried properly.
1. Smush data from all sources about a single real-world entity
 - while keeping track of information provenance.
1. Resolve inconsistencies in the data
 - by applying different trust heuristics.

Data Quality and Trust

- There are no facts on the Web!
- The Web is a social thing and everything on the Web is a claim.
- Therefore we research more on quality assessment, trust, data-cleansing.
- Move the trust layer down in the Semantic Web Layer Cake
 - Right above RDF and below OWL, SPARQL and RIF?

Semantic web layer cake



User Interfaces and Interaction Paradigms

- How do we build interfaces that operate over such large amounts of data?
 - How to aggregate the data in a meaningful way?
- What will be their interaction paradigm?
 - Will the browser be something like a Web-Excel including drill-down?
 - Will end-users notice that they are using Linked Data?
- How to explain data provenance and data fusion?
- Will search engines turn into answer engines?

Licensing

- Framework for publishing information about permissions and restrictions
- Encourage data owners in their participation in the Web of Data
- Creative Commons provided a framework for open licensing of creative works, underpinned by the notion of copyright
- However, copyright law is not applicable to data

Thank you.



Questions?