

Semantic Web Activity @ W3C

Ivan Herman

W3C Workshop on Multilingual Web

Luxembourg, March 15, 2012



Before going into details...

What does the term “Semantic Web” mean to people?

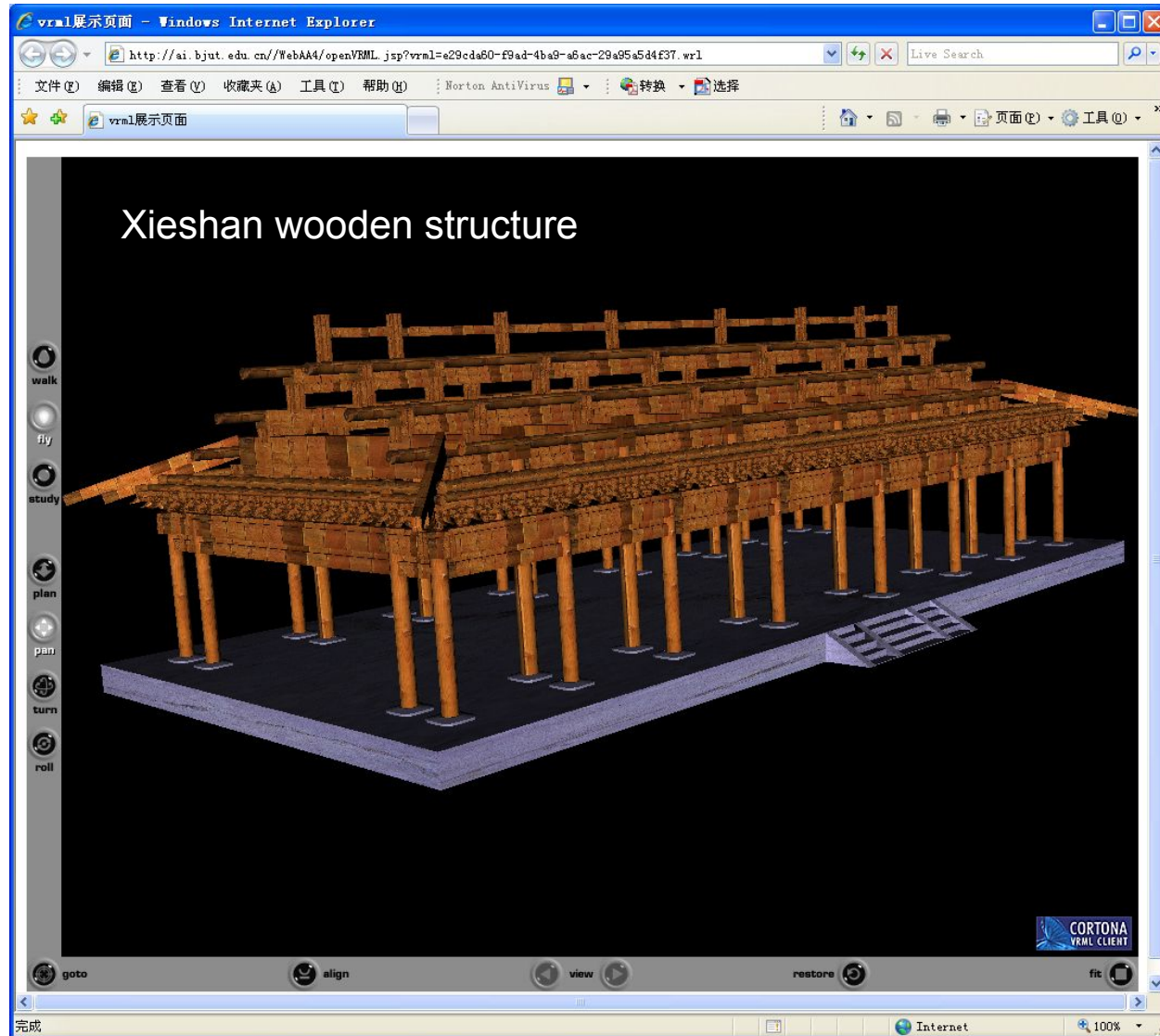
衆瞽
摸象之圖



For some people, Semantic (Web) is...

- ▶ An intelligent system manipulating and analyzing knowledge bases
 - e.g., via big ontologies, vocabularies
- ▶ A means to manage large amount of data
- ▶ Improve search by adding structure to embedded data
- ▶ A means to *integrate* many different pieces of data
- ▶ And a mixture of all these...

Example: OWL as a knowledge base



Example: making use of major ontologies

- ▶ Help in finding the best drug regimen for a specific patient

Optimized Regimens | Current Regimen | Survey Summary | Survey Set-up

PharmaSURVEYOR

PharmaSURVEY for Abby

Survey 1 Version 2

Optimized | Differences | Ranges | Common | Severe | All | Custom

Current Regimen

Safety Optimized Profiles

Severity	Adverse Drug Effect	v2	1	2
ADE	Moderate Muscle Weakness (Myasthenia)	✓		
ADE	Minor Excessive Sweating (Diaphoresis)	✓		
ADE	Moderate Heart Throbbing or Pounding (Palpitations)	✓		
ADE	Moderate Hives (Urticaria)	✓		
ADE	Major Bladder Inflammation (Cystitis)	✓		
ADE	Major Urinary Tract Infection	✓		

row(s) 1 - 6 of 6

[Export to Excel](#)

Example: making use of linked data

BBC - Music - Eric Clapton

http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

Netvibes Feedly Social Private Mailing lists SW Python RDFa it! Bookmarkslets Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG


QUICK FIND Enter an artist name ...

BBC Music > Artists > Eric Clapton

Eric Clapton

Born 30 March 1945.

MOST PLAYED ON BBC RADIO 2



David Redfern/Redferns


Biography

Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and #53 on their list of the Immortals: 100 Greatest Artists of All Time.

Latest Tracks Played On The BBC

- Promises**
BBC Radio 2 | [Ken Bruce](#) 22/02/2010
- Bad Love**
BBC Radio 2 | [Alex Lester](#) 22/02/2010
- Lay Down Sally**
BBC Radio 2 | [Chris Evans Breakfast](#) 18/02/2010
- I Ain't Gonna Stand For It**
BBC Radio 2 | [Alex Lester](#) 15/02/2010
- Wonderful Tonight**
BBC Radio 2 | [Ken Bruce](#) 10/02/2010

Audio Previews From Latest Album Review



Me And Mr Johnson

- 8 Milkcow's Calf Blues
- 10 Come on in My Kitchen

Example: making use of linked data

The screenshot shows a web browser window with the address bar displaying <http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294>. The page title is "BBC - Music - Eric Clapton".

The main content area features a Wikipedia entry for Eric Clapton, which is partially visible. Below the Wikipedia entry, there is a section titled "Links & Information". This section contains several links, some of which are circled in red:

- [Official homepage at ericclapton.com](#)
- [Fanpage at whereseric.com](#)
- [Wikipedia article on Eric Clapton](#)
- [MySpace at mspace.com/ericclapton](#)
- [Last.fm page on Eric Clapton](#)
- [MusicBrainz entry on Eric Clapton](#)

Below the links, there is a section titled "MEMBER OF" which lists the bands Eric Clapton has been a member of: [Derek and the Dominos](#), [Blind Faith](#) (1968-1969), [Cream](#) (1966-1968), [John Mayall & The Bluesbreakers](#) (1965-1966), and [The Yardbirds](#) (1963-1965).

Below this, there is a section titled "COLLABORATED ON" which lists various collaborations, including [J.J. Cale & Eric Clapton](#), [Eric Clapton & The Immediate All Stars](#), [Eric Clapton & The Impressions](#), [Eric Clapton & Jimmy Page](#), [Eric Clapton & David Sanborn](#), [Eric Clapton & Stan Webb's Chicken Shack](#), [Eric Clapton & The Powerhouse](#), [Eric Clapton & Stevie Ray Vaughan](#), [Eric Clapton & Marc Shaiman](#), [The Dirty Mac](#), [Bob Dylan, Roger McGuinn, Tom Petty, Neil Young, Eric Clapton & George Harrison](#), [Jimmie Vaughan, Eric Clapton, Bonnie Raitt, Robert Cray, B.B. King, Buddy Guy, Dr. John & Art Neville](#), [Elton John & Eric Clapton](#), [Michael Kamen, Eric Clapton and David Sanborn](#), [B.B. King & Eric Clapton](#), [Mark Knopfler & Eric Clapton](#), [Paul McCartney & Eric Clapton](#), [Sting with Eric Clapton](#), [Steve Winwood & Eric Clapton](#).

At the bottom of the "Links & Information" section, there is a note: "Links & information come from [MusicBrainz](#). You can add or edit information about [Eric Clapton at musicbrainz.org](#). Find out more about our use of this data. The BBC is not responsible for the content of external sites".

On the right side of the page, there is a section titled "Played By" which lists artists who have played on BBC Radio 2. The list includes:

- [Alex Lester](#) (BBC Radio 2)
- [Steve Wright in the Afternoon](#) (BBC Radio 2)
- [Sarah Kennedy](#) (BBC Radio 2)
- [Wake Up to Wogan](#) (BBC Radio 2)
- [Ken Bruce](#) (BBC Radio 2)
- [Steve Wright's Sunday Love Songs](#) (BBC Radio 2)
- [Jeremy Vine](#) (BBC Radio 2)

At the bottom of the page, there is a section titled "Latest News Stories" and a footer note: "Information displayed about artists played on BBC programmes is incomplete. [out more about this artist play count information.](#)"

Example: making use of structured data and search engine facilities

The Adventures of Tintin (2011) - IMDb

www.imdb.com/title/tt0983193/

2011 TR LocalData Mercurial reposi... Private Mailing lists Social SW Python RDFa it! Bookmarks To... Wolfram save on delicious

adventures of tintin - Google Search The Adventures of Tintin (2011) - IMDb

IMDb Find Movies, TV shows, Celebrities and more... All

Register | Login | Help

Movies TV News Videos Community IMDbPro Apps Your Watchlist

The Adventures of Tintin (2011)

PG 107 min - [Animation](#) | [Action](#) | [Adventure](#)
- 26 October 2011 (Netherlands)

Your rating: ★★★★★★ ★★ -/10
Ratings: **7.7/10** from 37,333 users Metascore: 68/100
Reviews: 304 user | 307 critic | 40 from Metacritic.com

Intrepid reporter Tintin and Captain Haddock set off on a treasure hunt for a sunken ship commanded by Haddock's ancestor.

Director: [Steven Spielberg](#)
Writers: [Steven Moffat](#) (screenplay), [Edgar Wright](#) (screenplay), [and 2 more credits](#) »
Stars: [Jamie Bell](#), [Andy Serkis](#) and [Daniel Craig](#)

[Watch Trailer](#) [+ Watchlist](#) [Check In](#)

[49 photos](#) | [25 videos](#) | [2035 news articles](#) | [full cast and crew](#) »

Nominated for Oscar. Another 8 wins & 15 nominations [See more awards](#) »

Watch it
 Buy it from [Amazon](#) »

Share this page:
 Like 39026 likes. Sign Up to see what your friends like.

Quick Links:
[overview](#)

Related News
[John Williams Receives 7 Nominations from the International Film Music Critics](#)
8 hours ago | [AwardsDaily](#)
[Details For The Adventures Of Tintin: The Secret Of The Unicorn Blu-Ray & DVD Release](#)
13 hours ago | [The Hollywood News](#)
[Outguess Ebert on the Oscars](#)
14 hours ago | [Roger Ebert's Blog](#)
[See all 2035 related articles](#) »

ROAD TO THE OSCARS [Learn more](#) »

Example: making use of structured data and search engine facilities

The Adventures of Tintin (2011) - IMDb

www.imdb.com/title/tt0983193/

Find Movies, TV shows, Celebrities and more... All

Register | Login | Help

2011 TR LocalData Mercurial reposi... Private Mailing lists Social SW Python RDFa it! Bookmarkslets To... Wolfram save on delicious

adventures of tintin - Google Search The Adventures of Tintin (2011) - IMDb

IMDb

Movies TV News Videos Community IMDbPro Apps

Your Watchlist

The Adventures of Tintin (2011)

Top 500

Watch it

Buy it from Amazon

Source of: http://www.imdb.com/title/tt0983193/

```
</div>

<div class="star-box-giga-star" itemprop="aggregateRating" itemscope itemType="http://schema.org/AggregateRating">
7.7
</div>
<div class="star-box-rating-widget">
<span class="star-box-rating-label">Your rating:</span>

<div class="rating-rating-list" data-auth="BCYoFakeBiePvOcomMkTrZnEx4FU8QWqY0zz0KrS5HEPMYjTJg8ZYOMLw694ZS5TMLQCqbuETbHv6sT4SAVPf
title="Users rated this 7.7/10 (37,333 votes) - click stars to rate">
<span class="rating-bg">&nbsp;</span>
<span class="rating-imdb" style="width: 0px">&nbsp;</span>
<span class="rating-stars">
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>1</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>2</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>3</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>4</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>5</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>6</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>7</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>8</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>9</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>10</span></a>
</span>
<span class="rating-rating"><span class="value"></span><span class="grey"></span><span class="grey">10</span></span>
<span class="rating-cancel"><a href="/title/tt0983193/vote?v=x&BCYoFakeBiePvOcomMkTrZnEx4FU8QWqY0zz0KrS5HEPMYjTJg8ZYOMLw694ZS5TMLQCqbuETbHv6sT4SAVPf"
Line 579, Col 58
```

Example: making use of structured data and search engine facilities

The screenshot shows a Google search for "adventures of tintin". The search bar is at the top with the text "adventures of tintin". Below the search bar, the results are displayed. On the left side, there is a sidebar with filters: "Everything", "Images", "Maps", "Videos", "News", "Shopping", "Books", "More", "Amsterdam", "Change location", "Any time", "Past hour", "Past 24 hours", "Past week", "Past month", "Past year", "Custom range...", and "More search tools".

The main results area shows several links. The first result is "The Adventures of Tintin (2011) - IMDb" with a URL www.imdb.com/title/tt0983193/. This result is circled in green. Below the link, it shows a star rating of 7.7/10 based on 37,333 votes, and a brief description: "Intrepid reporter Tintin and Captain Haddock set off on a treasure hunt for a sunken ship commanded by Haddock's ancestor. Directed by Steven Spielberg. Starring Jamie Bell, Andy Serkis." There are also links to "Full cast and crew" and "Pictures & Photos from The ...".

The second result is "The Adventures of Tintin - Wikipedia, the free encyclopedia" with a URL en.wikipedia.org/wiki/The_Adventures_of_Tintin. It provides a brief overview of the comic book series created by Georges Remi (1907-1983).

The third result is "The Adventures of Tintin (film) - Wikipedia, the free encyclopedia" with a URL [en.wikipedia.org/wiki/The_Adventures_of_Tintin_\(film\)](http://en.wikipedia.org/wiki/The_Adventures_of_Tintin_(film)). It describes the 2011 American performance capture 3D film.

The fourth result is "Welcome to the Adventures of Tintin" with a URL www.tintin.com/. It mentions that Tintin running with Snowy and Tintin Adventures inspire and delight readers all over the world.

The fifth result is "The Adventures of Tintin Trailer | Official Movie Site | Dec 21, 2011" with a URL www.us.movie.tintin.com/. It mentions that the Adventures of Tintin official site is a 3D motion capture film directed by Steven Spielberg, in theaters December 21, 2011.

The sixth result is "The Adventures of Tintin - Rotten Tomatoes" with a URL www.rottentomatoes.com/m/the_adventures_of_tintin/. This result is also circled in green. It shows a star rating of 75% based on 189 reviews, and a brief description: "Paramount Pictures and Columbia Pictures Present a 3D Motion Capture Film The Adventures of Tintin directed by Steven Spielberg from a screenplay by... Directed by Steven Spielberg. Starring Jamie Bell, Andy Serkis."

衆瞽
摸象之圖

And that is all right!



- ▶ We have to acknowledge that the field has grown and has become multi-faceted
- ▶ All different “views” have their success stories
- ▶ There are also no clear and water-proof boundaries between the different views
- ▶ The question is: where is the emphasis?

Data on the Web

- ▶ There are more and more data on the Web
 - government data, health related data, general knowledge, company information, flight information, restaurants,...
- ▶ More and more applications rely on the availability of that data



But: we do not want that!

Imagine...

- ▶ A “Web” where
 - documents are available for download on the Internet
 - but there would be no hyperlinks among them

CoCoDat - Collation of Cortical Data - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.cocomac.org

CoCoMac DATABASES ORT EXAMPLES

CoCoDat: Collation of Cortical [microcircuitry] Data

CoCoDat is a microcircuitry database that published experimental reports. The data and cellular compartment), as well as the

- Morphology
- Firing properties
- Ionic currents
- Ionic conductances
- Synaptic currents
- Connectivity

The database is available for download u data tables but also a Search Board with manual or automatic relaxation of the sea

- Brain region
- Layer
- Neuron type

http://www.cocomac.org/cocodat/catalyzer/ind

Cell Centered Database - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ccdb.ucsd.edu/sand/main?event=gallery&action=show&dpl=y

Cell Centered Database™ Gallery

National Center for Microscopy and Imaging Research

Data | Search | Gallery | Dictionary | Publications | MyCCDB | Data Download | Contact us | Help

2D image Reconstruction Segmentation Animation

NeuronDB = Thalamic relay neuron - Overview (A) () - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://senselab.med.yale.edu

NeuronDB

Thalamic relay neuron

Back

Mode: **Overview** Data/Search plus Connectivity plus Classical References/Notes Models

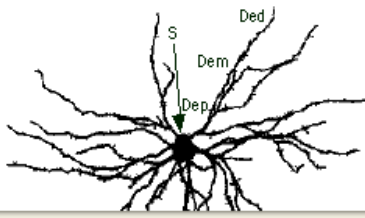
Region: Distal equivalent dendrite Middle equivalent dendrite Proximal equivalent dendrite Soma Axon hillock Axon fiber Axon terminal All Compartments

Properties: Receptors Channels Transmitters **All Properties**

Interoperation: Gene and Chromosome Experimental Data (neurodatabase.org) Microscopy Data (CCDB)

Neuron type: principal

Organism: Vertebrates



1. Equivalent dendrite
2. Distal equivalent dendrite
3. Middle equivalent dendrite
4. Proximal equivalent dendrite
5. Soma

Show other

Done

logged out

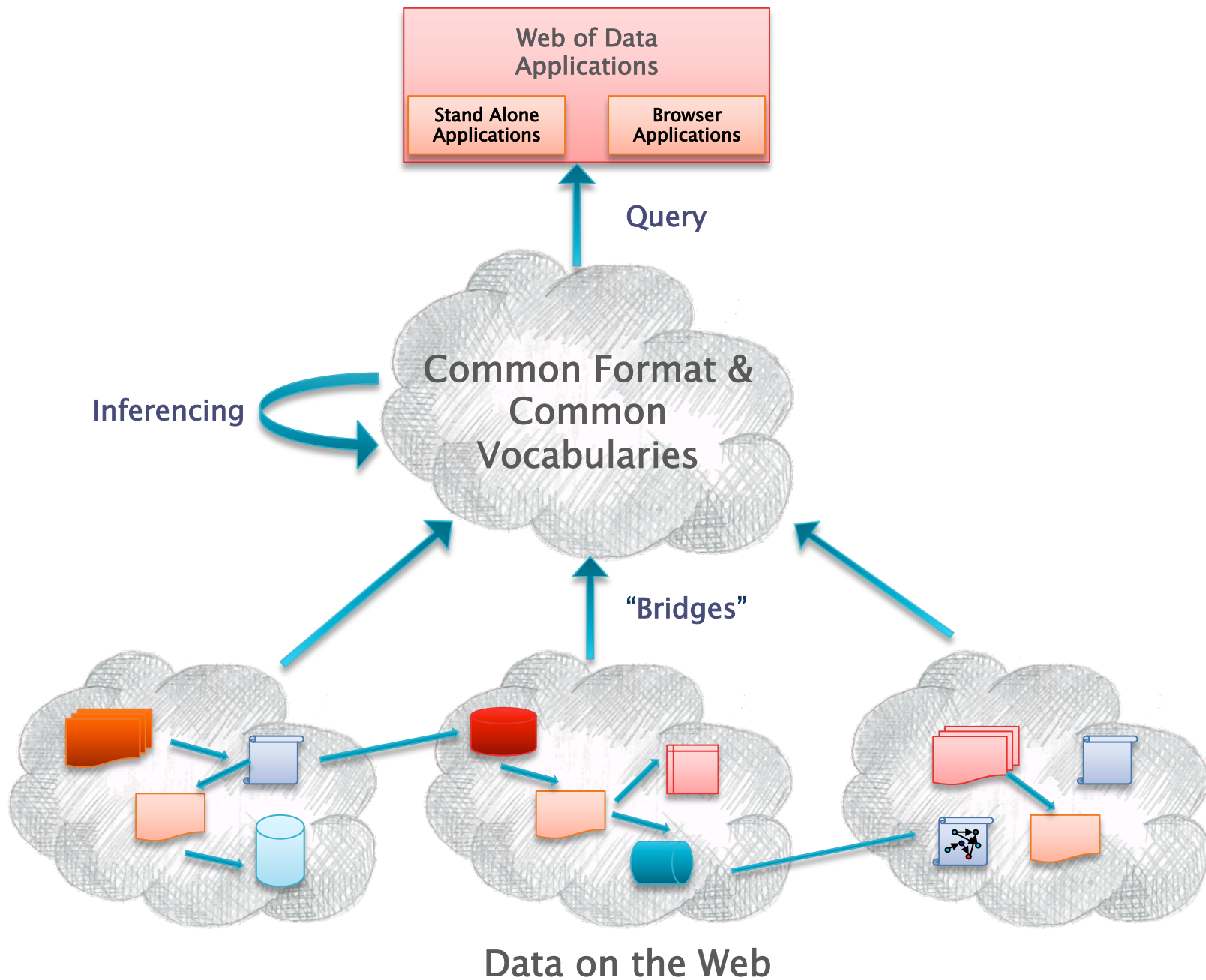
Data on the Web is not enough...

- ▶ We need a proper infrastructure for a real Web of Data
 - data is available on the Web
 - accessible via standard Web technologies
 - data are *interlinked over the Web*
 - the terms used for linkage are well defined
- ▶ I.e.: data can be integrated over the Web



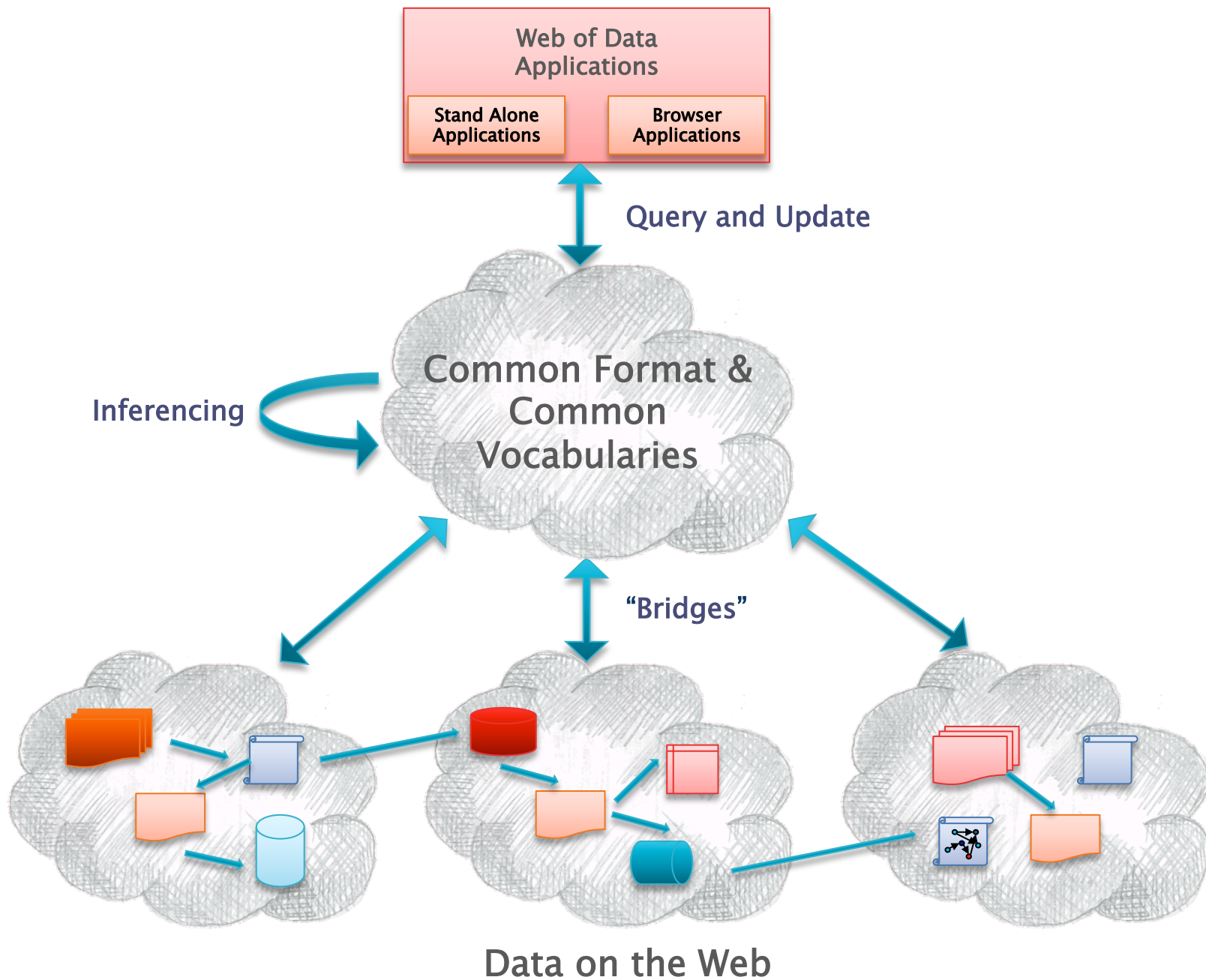
This is what we want!

Semantic Web technologies should be
at the service of such a Web of Data



On a longer term...

- ▶ Data should be easily manipulated from an application
- ▶ Some sort of a “globally linked database”
- ▶ But this is still further down the road...



Views: [desktop](#) [mobile](#) [print](#)[STANDARDS](#)[PARTICIPATE](#)[MEMBERSHIP](#)[ABOUT W3C](#)

Google™



W3C SEMANTIC WEB ACTIVITY

NEARBY

On this page → [publications, interviews](#) • [presentations](#) • [active groups](#) • [completed groups](#) • [past groups](#)[Activity news](#)[Specifications](#)[FAQ](#)[Use Cases and Case Studies](#)[List of Tools](#)[On-line validators](#)[List of Books](#)[Semantic Web Logos and Buttons](#)[SW Wiki](#)

ACTIVITY RSS FEEDS

[Activity news \(RSS 1.0\)](#)

So... what is happening at W3C?



The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

What is the Semantic Web?

The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework ([RDF](#)). See also the separate [FAQ](#) for further information.

The (almost) past

- ▶ Some technologies are in the process of finalization
 - SPARQL 1.1 (SPARQL Protocol and RDF Query Language)
 - RDB2RDF (Relational Databases to RDF)
 - RDFa 1.1 (RDF in attributes)

The present

- ▶ Some areas are subject of intensive work
 - RDF update (Resource Description Framework)
 - Provenance

The future

- ▶ We are discussing new works, new areas, e.g.,
 - Linked Data Platform
 - Access Control issues
 - Constraint checking on Semantic Web data
 - ...

Link to specialized communities

- ▶ Various communities have different emphasis on which part of the Semantic Web they want to use
- ▶ W3C has contacts with some of those
 - health care and life sciences (a separate IG is up and running)
 - libraries, publishing
 - financials
 - the oil, gas, and chemicals community

A large pile of M&M's candies in various colors (red, green, pink, orange) scattered on a light surface. Many of the candies have black text printed on them, including 'Linked Data', 'Semantic Web', 'MIT IAP 2010', and 'MIT IAP'. One red candy on the right side features the white M&M's logo. The candies are piled together, with some in sharp focus and others blurred in the background.

Query RDF: SPARQL 1.1

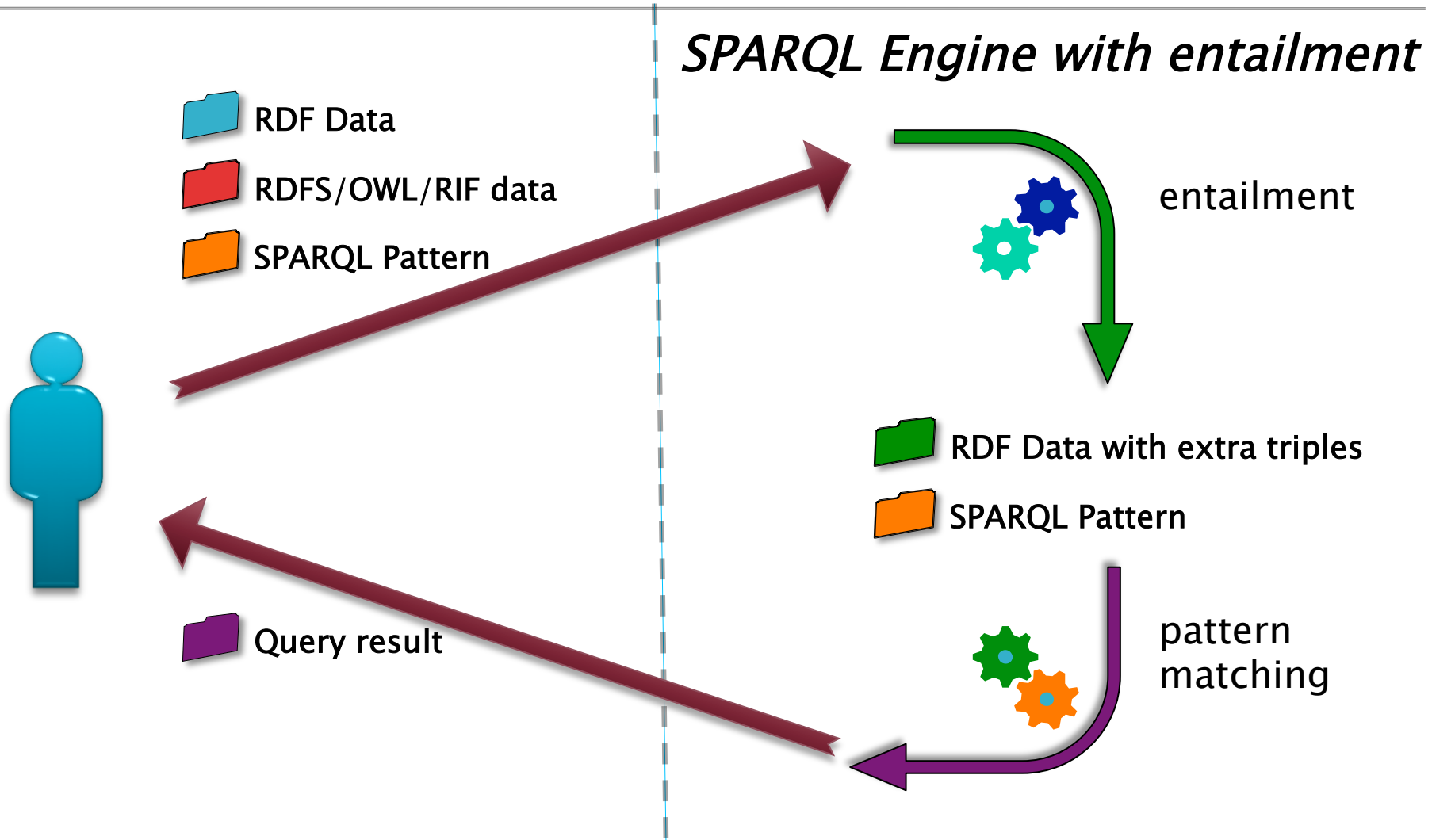
Reminder...

- ▶ *SPARQL is a query language on RDF data*
- ▶ SPARQL is defined in terms of a protocol, to send query and results over the Web
- ▶ Is based on the idea of “graph pattern matching”:
 1. a graph pattern is described in the query, with real and unknown nodes (“variables”)
 2. if the pattern can match a portion of the graph, the unknown nodes are replaced by the “real” ones
 3. resulting information is returned
- ▶ First version of SPARQL was published in 2008

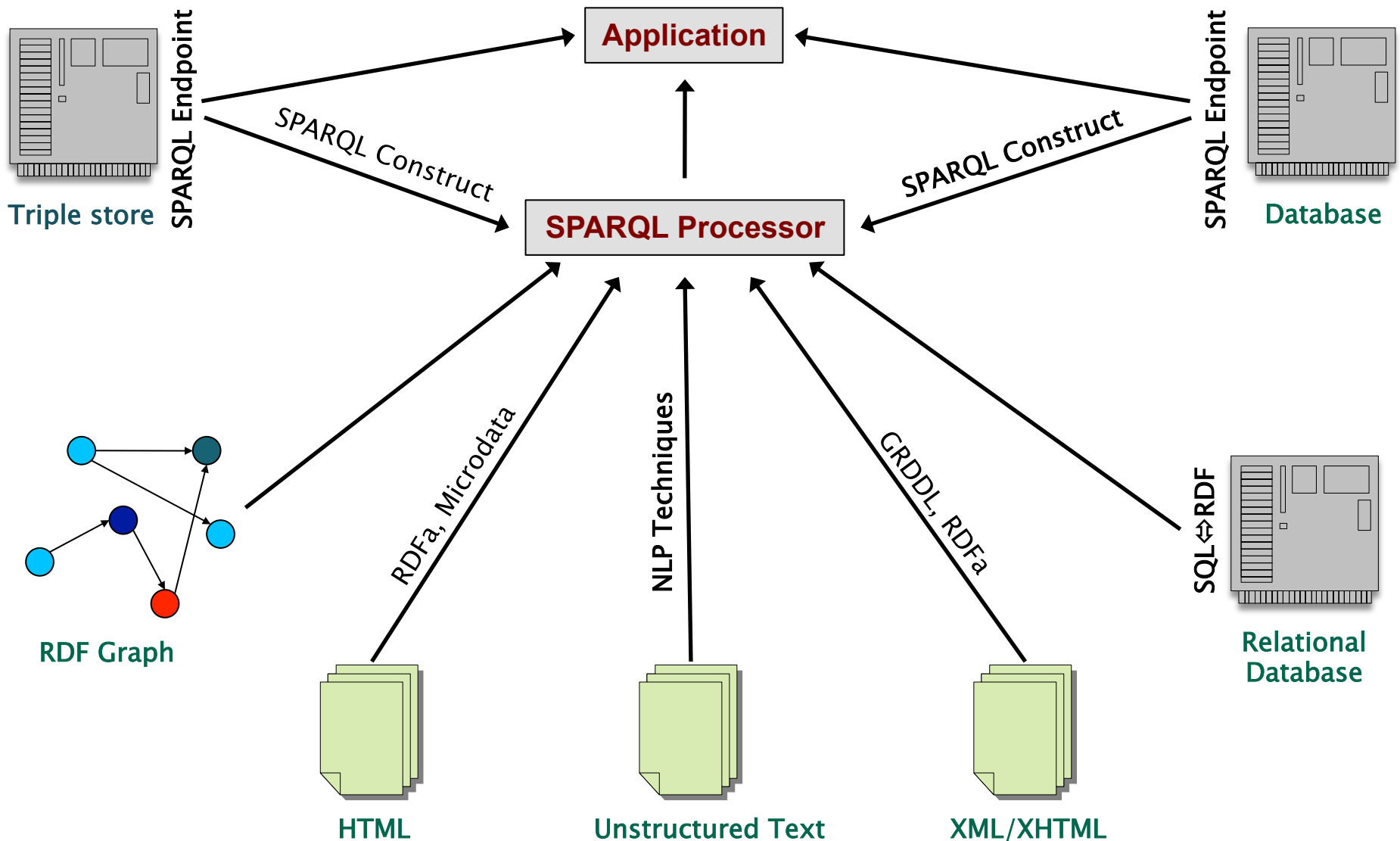
SPARQL 1.1: adding missing features to SPARQL

- ▶ Nested queries (i.e., **SELECT** within a **WHERE** clause)
- ▶ Negation (**MINUS**, and a **NOT EXIST** filter)
- ▶ Aggregate function on search results (**SUM**, **MIN**,...)
- ▶ Property path expression (**?x foaf:knows+ ?y**)
- ▶ SPARQL UPDATE facilities (**INSERT**, **DELETE**, **CREATE**)
- ▶ Combination with entailment regimes

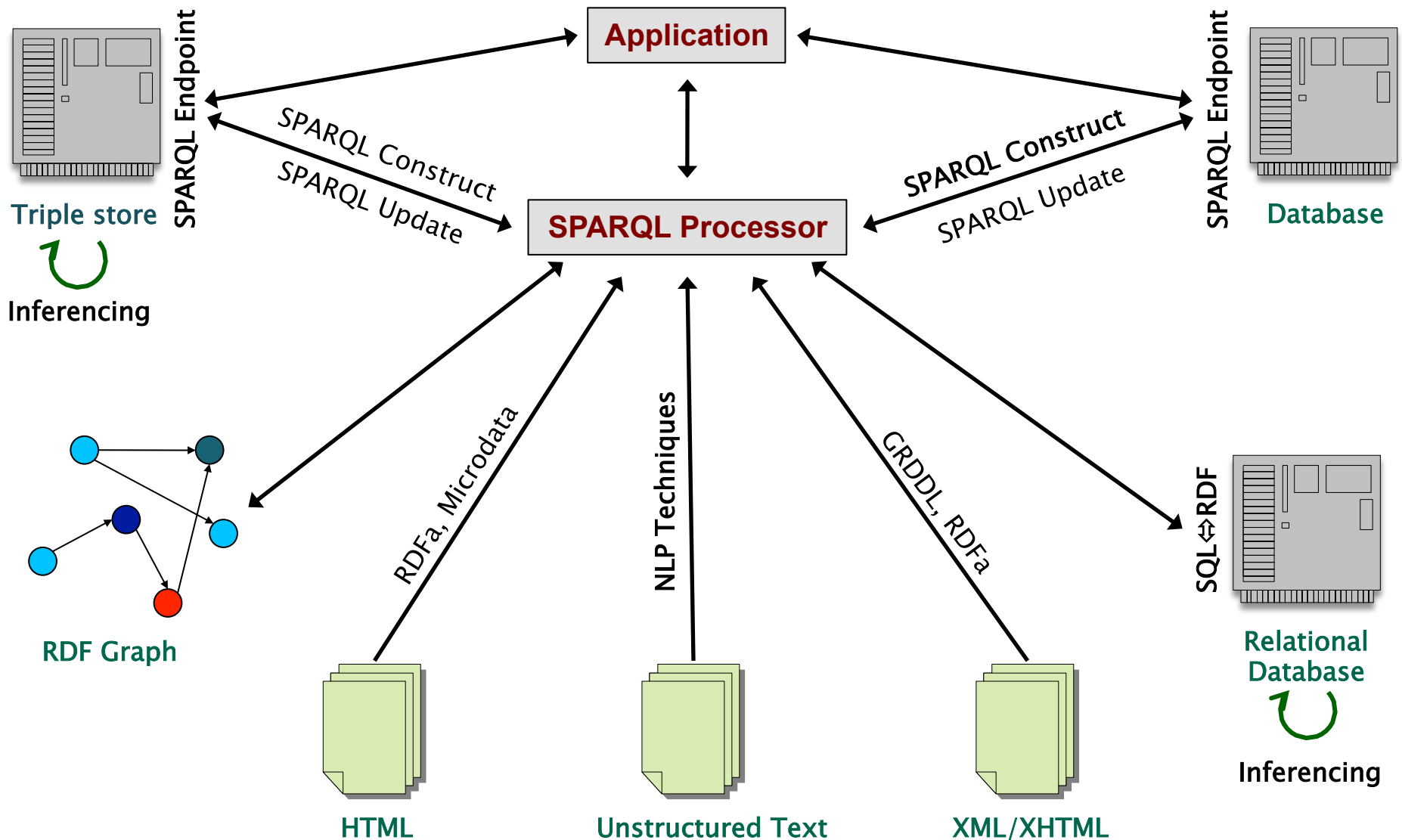
SPARQL 1.1 and RDFS/OWL/RIF



SPARQL as a unifying point



SPARQL 1.1 as a unifying point



SPARQL 1.1 Status

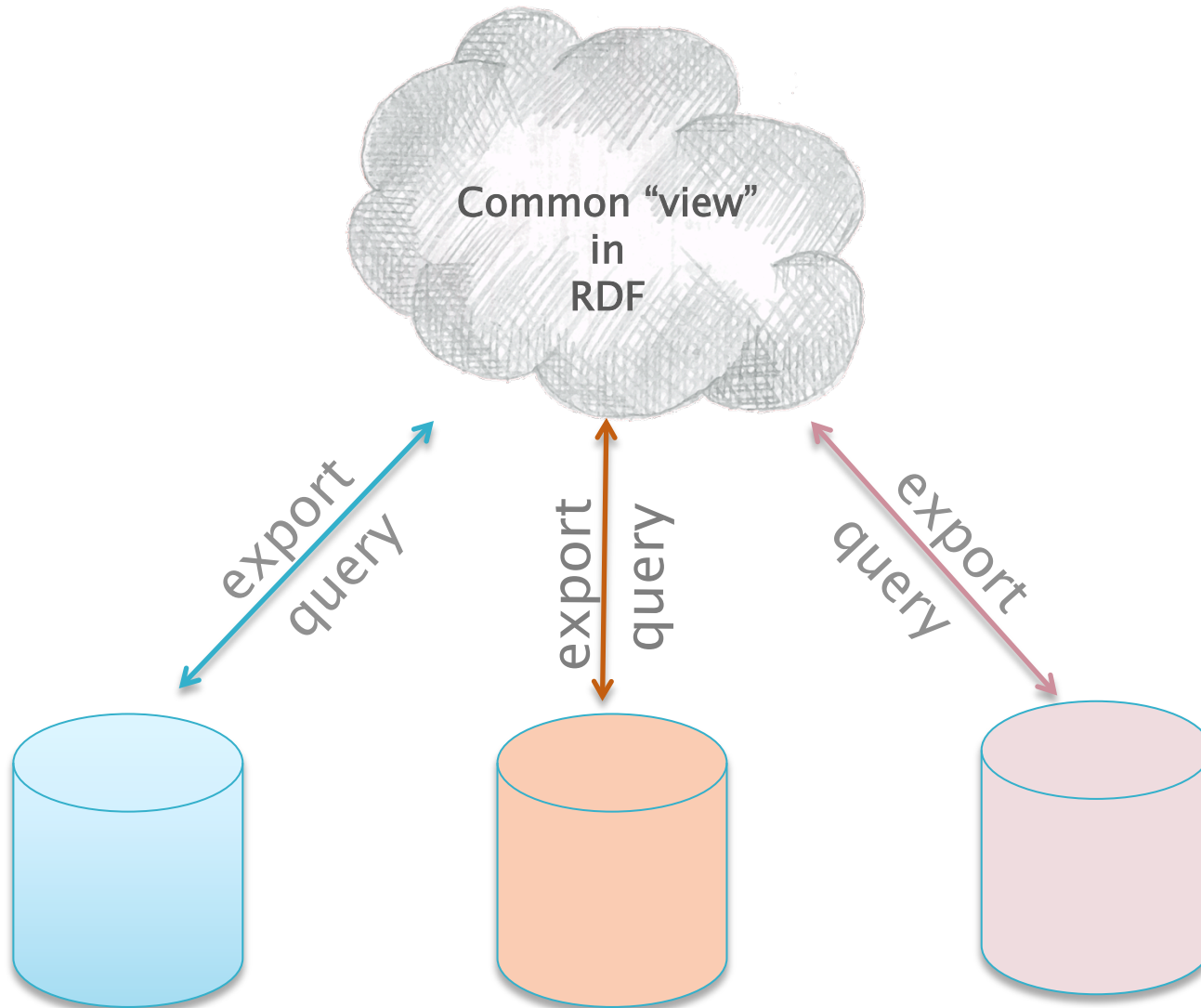
- ▶ Technology has been finalized
- ▶ Goes to “candidate recommendation” soon
- ▶ Should be finished this summer

Access to Relational Databases

Relational Databases and RDF

- ▶ Most of the data on the Web is, in fact, in RDB-s
- ▶ Proven technology, huge systems, many vendors...
- ▶ Data integration on the Web must provide access to RDB-s
- ▶ *RDB2RDF provides means to “see” relational data as RDF*

RDF provides a common “view”



What is “export”?

- ▶ “Export” does not *necessarily* mean physical conversion
 - for very large databases a “duplication” would not be an option
 - systems may provide SPARQL \Leftrightarrow SQL “bridges” to make queries on the fly
- ▶ Result of export is a “logical” view of the RDB content

Simple export: Direct Mapping

- ▶ A canonical RDF “view” of RDB tables
- ▶ Only needs the information in the RDB Schema

Fundamental approach

Each column name provides a predicate

ISBN	Author	Title	Publisher	Year
0006511409X	id_xyz	The Glass Palace	id_qpr	2000
0007179871	id_xyz	The Hungry Tide	id_qpr	2004

Each row is a subject

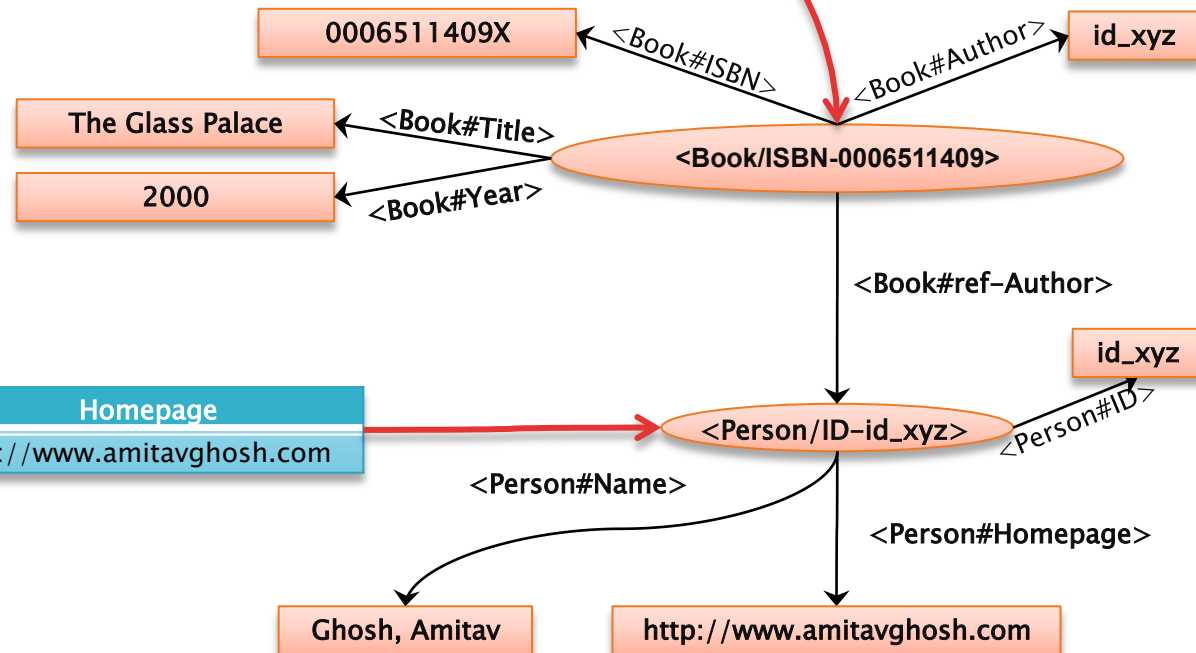
Table references are URI objects

Cells are Literal objects

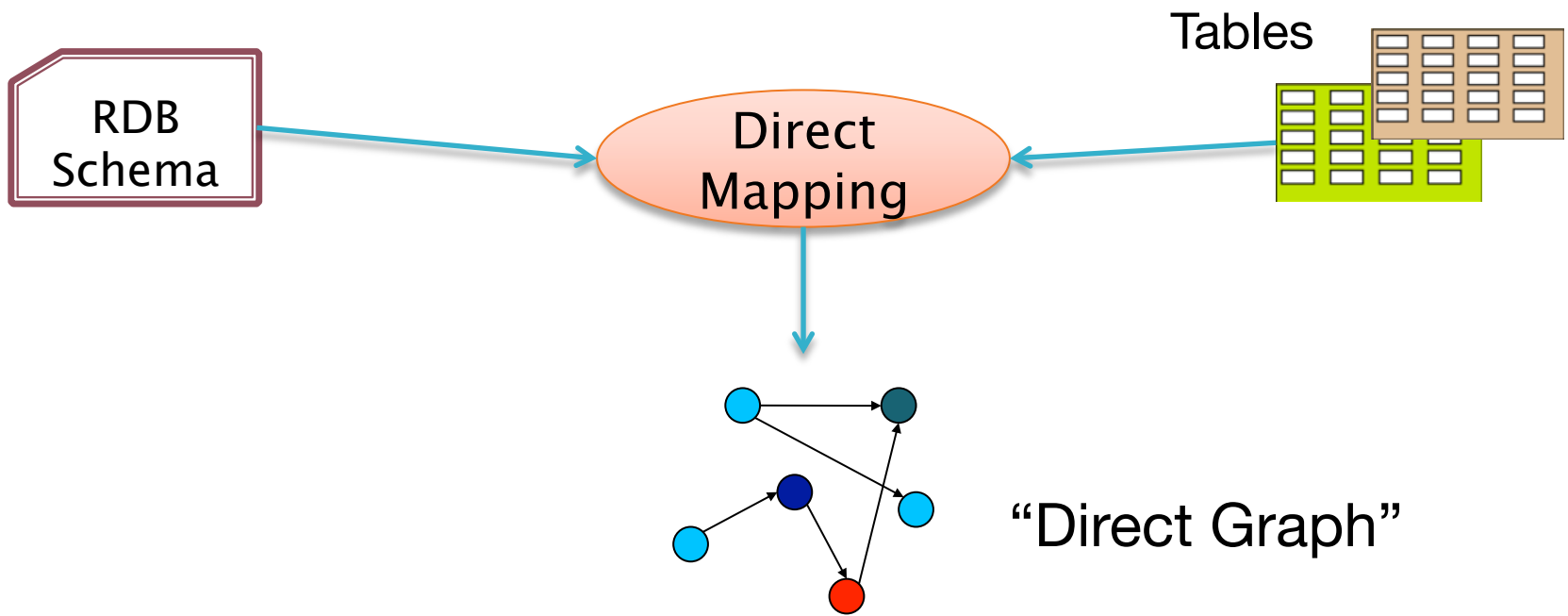
ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

Fundamental approach

ISBN	Author	Title	Publisher	Year
0006511409X	id_xyz	The Glass Palace	id_qpr	2000
0007179871	id_xyz	The Hungry Tide	id_qpr	2004



ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com



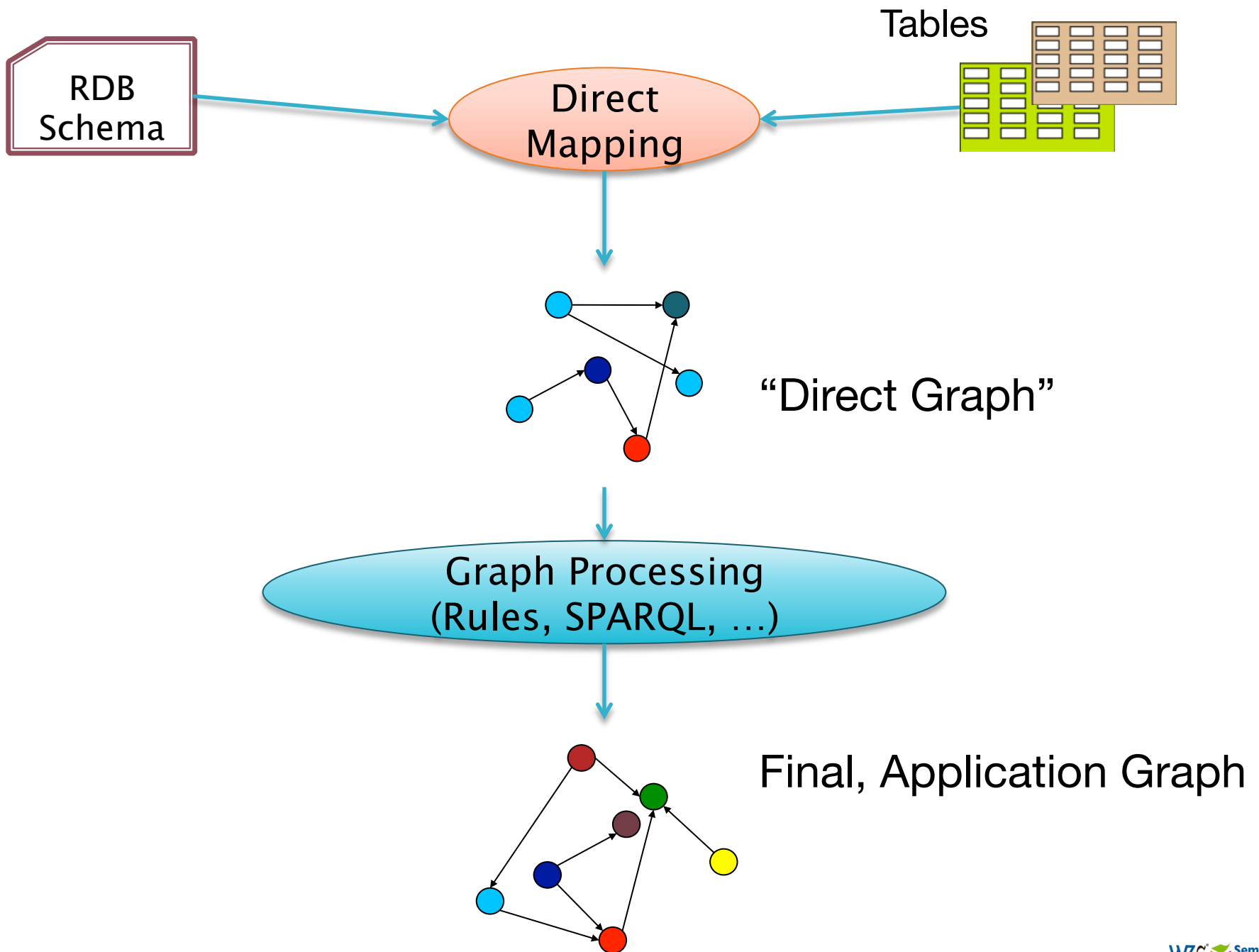
Pros and cons of Direct Mapping

► Pros:

- Direct Mapping is simple, does not require any other concepts
- know the Schema \Rightarrow know the RDF graph structure
- know the RDF graph structure \Rightarrow good idea of the Schema(!)

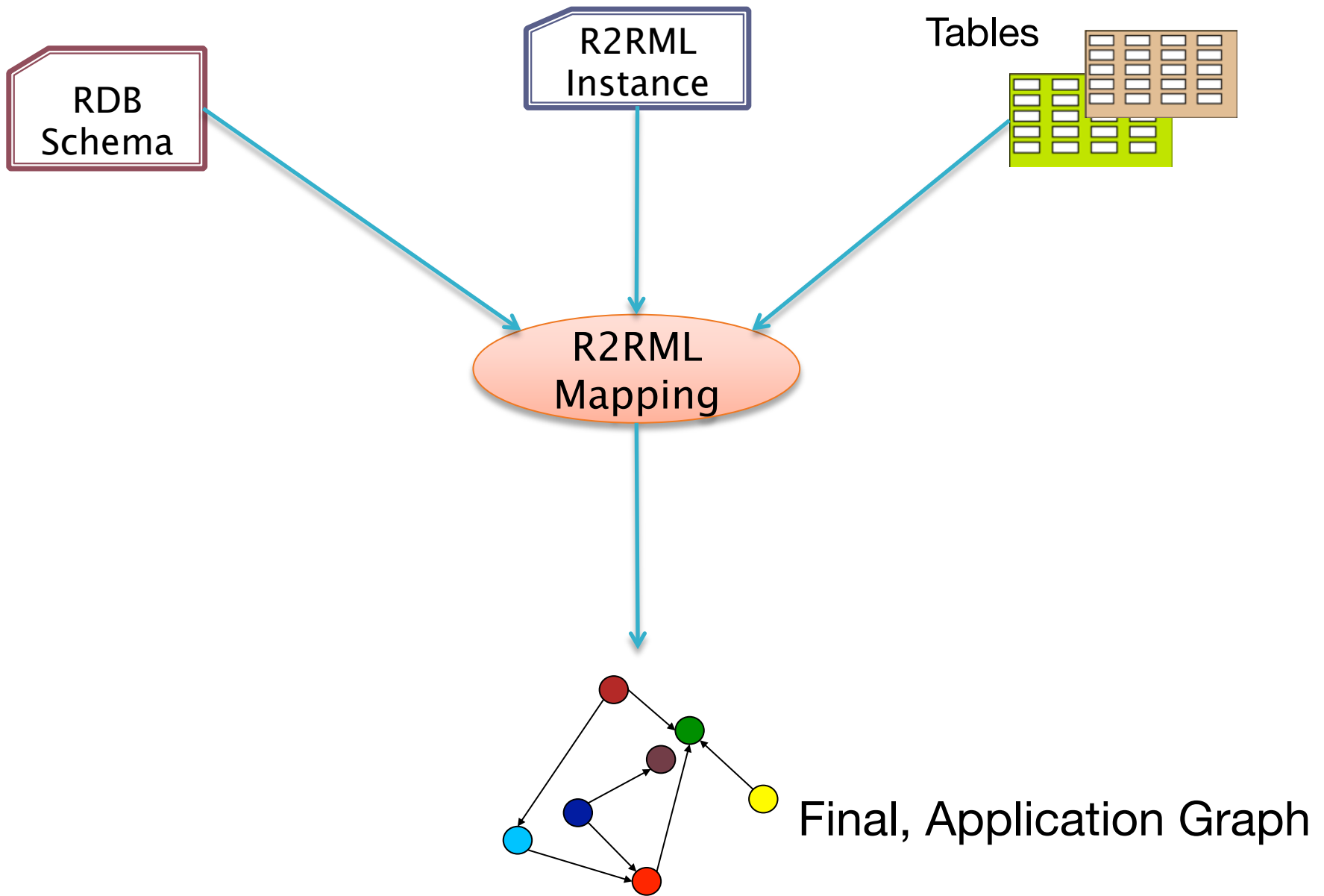
► Cons:

- the resulting graph is not what the application really wants



Beyond Direct Mapping: R2RML

- ▶ Separate vocabulary to control the details of the mapping, e.g.:
 - finer control over the choice of the subject
 - creation of URI references from cells
 - predicates may be chosen from a vocabulary
 - datatypes may be assigned
 - etc.
- ▶ Gets to the final RDF graph with one processing step



Relationships to the Direct Mapping

- ▶ Fundamentals are similar:
 - each row is turned into a series of triples with a common subject
- ▶ Direct mapping is a “default” R2RML mapping

R2RML and Direct Mapping Status

- ▶ Technology has been finalized
- ▶ Both documents are in “Candidate Recommendation” phase
- ▶ Should be finished this summer



By Ivan Herman

Ivan Herman

This photo was taken on November 20, 2011 in Amstelveen, North Holland, NL, using a Canon PowerShot G12.



822 views 11 likes

This photo belongs to

Ivan Herman's photostream

The Dutch Autumn...

The Dutch Autumn...

Comments and faves



koalie added this photo to her favorite



Add your comment here...

Structured data in HTML:
RDFa & microdata

```
@prefix cc: <http://creativecommons.org/ns#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix xhv: <http://www.w3.org/1999/xhtml/vocab#> .
```

```
<http://www.flickr.com/photos/ivan_herman/6370721871/in/photostream> cc:license <http://creativecommons.org/licenses/by-nc-nd/2.0/deed.en>;
dc:title "The Dutch Autumn..."@en-us;
xhv:license <http://creativecommons.org/licenses/by-nc-nd/2.0/deed.en> .
```

HTML pages are a huge source of structured data

- ▶ Not necessarily large amount of data per page, but lots of them...
- ▶ Have become very valuable to search engines
 - Google, Bing, Yahoo!, or Yandex (i.e., schema.org) all committed to use such data
- ▶ Two syntaxes have emerged at W3C:
 - microdata with HTML5
 - RDFa with HTML5, XHTML, and with XML languages in general

Example: making use of structured data and search engine facilities

The Adventures of Tintin (2011) - IMDb

www.imdb.com/title/tt0983193/

Find Movies, TV shows, Celebrities and more... All

Register | Login | Help

Movies TV News Videos Community IMDbPro Apps

Your Watchlist

The Adventures of Tintin (2011)

Top 500

Watch it

Buy it from Amazon

Source of: http://www.imdb.com/title/tt0983193/

```
</div>

<div class="star-box-giga-star" itemprop="aggregateRating" itemscope itemType="http://schema.org/AggregateRating">
  7.7
</div>
<div class="star-box-rating-widget">
<span class="star-box-rating-label">Your rating:</span>

<div class="rating-rating-list" data-auth="BCYoFakeBiePvOcomMkTrZnEx4FU8QWqY0zz0KrS5HEPMYjTJg8ZYOMLw694ZS5TMLQCqbuETbHv6sT4SAVPf"
  title="Users rated this 7.7/10 (37,333 votes) - click stars to rate">
<span class="rating-bg">&nbsp;</span>
<span class="rating-imdb" style="width: 0px">&nbsp;</span>
<span class="rating-stars">
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>1</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>2</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>3</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>4</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>5</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>6</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>7</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>8</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>9</span></a>
<a href="/register/login?why=vote" title="Register or login to rate this title" rel="nofollow"><span>10</span></a>
</span>
<span class="rating-rating"><span class="value"></span><span class="grey"></span><span class="grey">10</span></span>
<span class="rating-cancel"><a href="/title/tt0983193/vote?v=x&BCYoFakeBiePvOcomMkTrZnEx4FU8QWqY0zz0KrS5HEPMYjTJg8ZYOMLw694ZS5TMLQCqbuETbHv6sT4SAVPf"
  title="Click here to rate this title" rel="nofollow"></a>
</span>
</div>
Line 579, Col 58
```

RDFa and microdata: similarities

- ▶ Both have similar philosophies:
 - the structured data is expressed *via attributes only* (no specialized elements)
 - both define some special attributes
 - e.g., `itemscope` for microdata, `resource` for RDFa
 - both reuse *some* HTML core attributes (e.g., `href`)
 - both reuse the textual content of the HTML source, if needed
- ▶ RDF data can be extracted from both
 - i.e., HTML+RDFa and HTML+microdata have become an additional source of Linked Data

RDFa and microdata: differences

- ▶ Microdata has been optimized for simpler use cases, concentrating on
 - one vocabulary at a time
 - tree shaped data
 - no datatypes
- ▶ RDFa provides a full serialization of RDF in XML or HTML
 - the price is an extra complexity compared to microdata
- ▶ RDFa 1.1 Lite is a simplified authoring profile of RDFa, very similar to microdata

RDFa 1.1 and microdata status

- ▶ For RDFa 1.1
 - Technology has been finalized
 - Is in “Candidate Recommendation”
 - Should be finished this summer
- ▶ For microdata
 - Technology has been finalized
 - Is part of HTML5, hence its advancement depends on other technologies

Relevant W3C notes

- ▶ “HTML Data Guide”
 - help publishers to choose what to use
 - how to combine microdata and RDFa in one document
- ▶ Microdata to RDF
 - defines a mapping of an HTML5+Microdata document to RDF

Cleaning up RDF



Reminder...

- ▶ *Resource Description Framework*: a graph-based model for (Web) data and its relationships
 - has a simple (**subject**, **predicate**, **object**) model
 - makes use of URI-s for the naming of terms
 - objects can also be Literals
 - informally: defines named relationships (named links) among entities on the Web
 - has different serialization formats
- ▶ Latest version was published in 2004

RDF cleanup (a.k.a. RDF1.1)

- ▶ Many issues have come up since 2004:
 - deployment issues
 - new functionalities are needed
 - underlying technology may have moved on (e.g., datatypes)
- ▶ The goal of the RDF Working Group is to refresh RDF
- ▶ NOT a complete reshaping of the standard!

Some new features

- ▶ Standardize Turtle as a serialization format
- ▶ Clean up some aspects of datatyping, e.g.:
 - plain vs. typed literals
 - details and role of `rdf:XMLLiteral`
- ▶ Proper definition for “named graphs”
 - including concepts, semantics, syntax, ...
 - obviously important for linked data access
 - but generates quite some discussions on the details
- ▶ etc.

Editorial improvements

- ▶ Cleanup the documents, make them more readable
 - possibly rewrite all documents
 - maybe a completely new primer
 - new structure for the Semantics document

Status

- ▶ Work has begun a bit less than a year ago
- ▶ Turtle is almost finalized
- ▶ Agreement on most of the literal cleanup
- ▶ Lots of discussion currently on named graphs...

Provenance

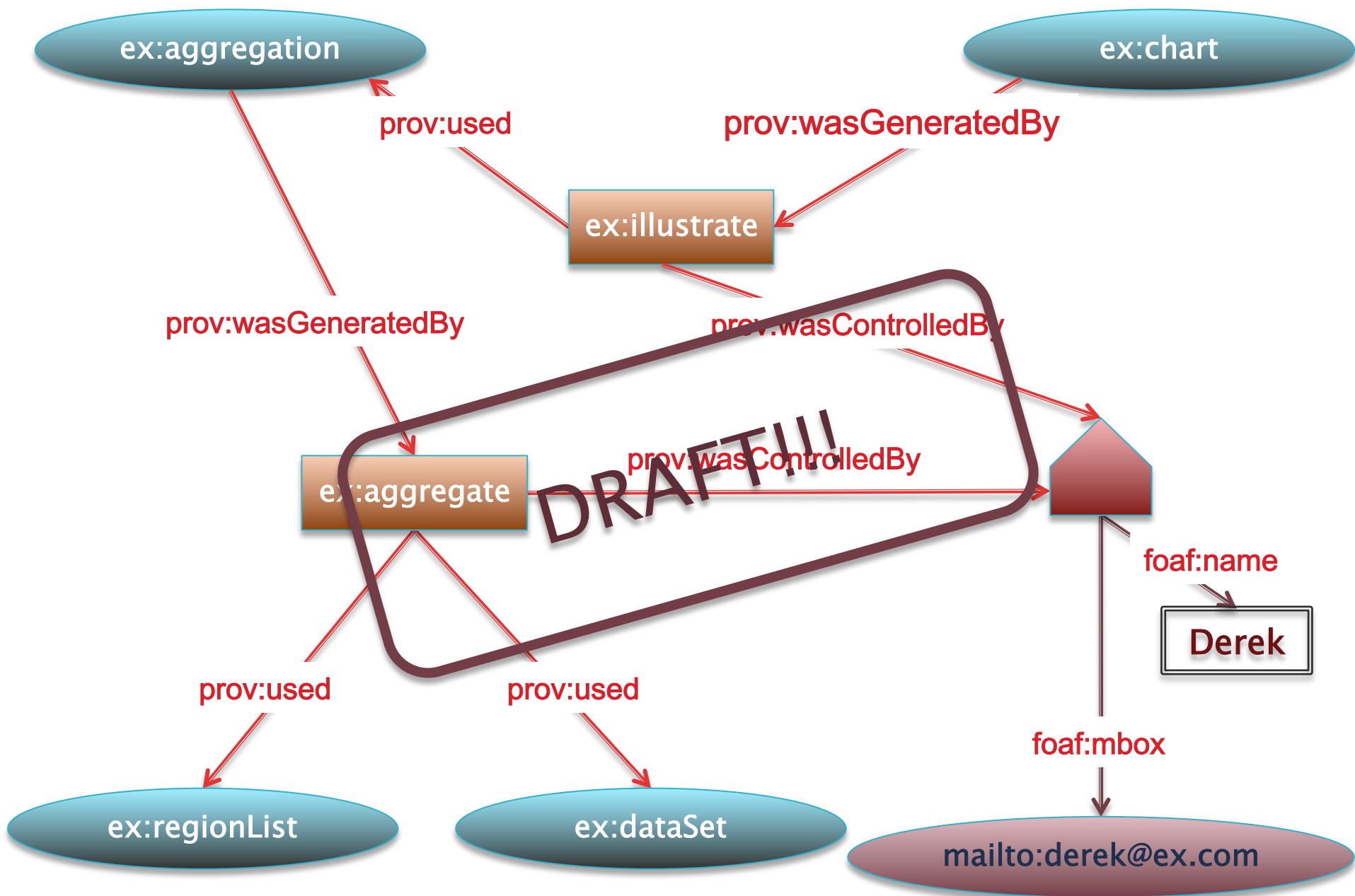
$$R_{ik} = 0$$

The goal is simple...

- ▶ We should be able to express all sorts of “meta” information on the data
 - creator: who played what role in creating the data (author, reviewer, etc.)
 - view of the full revision chain of the data
 - in case of a integrated data: which part comes from which original data and under what process
 - what vocabularies/ontologies/rules were used to generate some portions of the data
 - etc.

...the solution is more complicated

- ▶ Requires a complete model describing the various constituents (actors, revisions, etc.)
- ▶ The model should be usable with RDF
- ▶ Has to find a balance between
 - simple (“scruffy”) provenance: easily usable and editable
 - complex (“complete”) provenance: allows for a detailed reporting of origins, versions, etc.
- ▶ That is the role of the Provenance Working Group (started in 2011)



Status

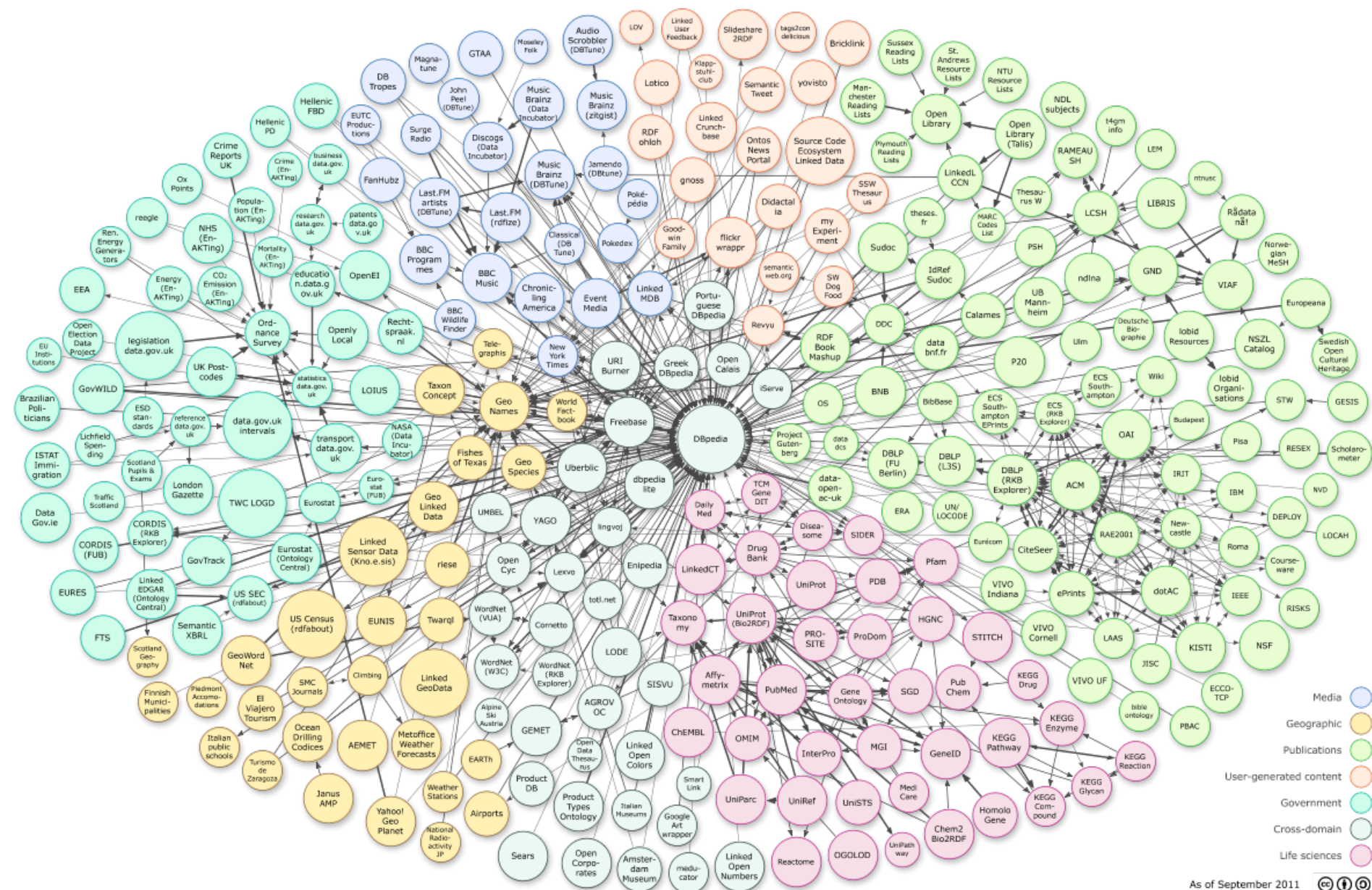
- ▶ Drafts have been published
 - abstract data model, OWL version
 - primer
- ▶ The current effort is on simplifying the first versions

[illegible]

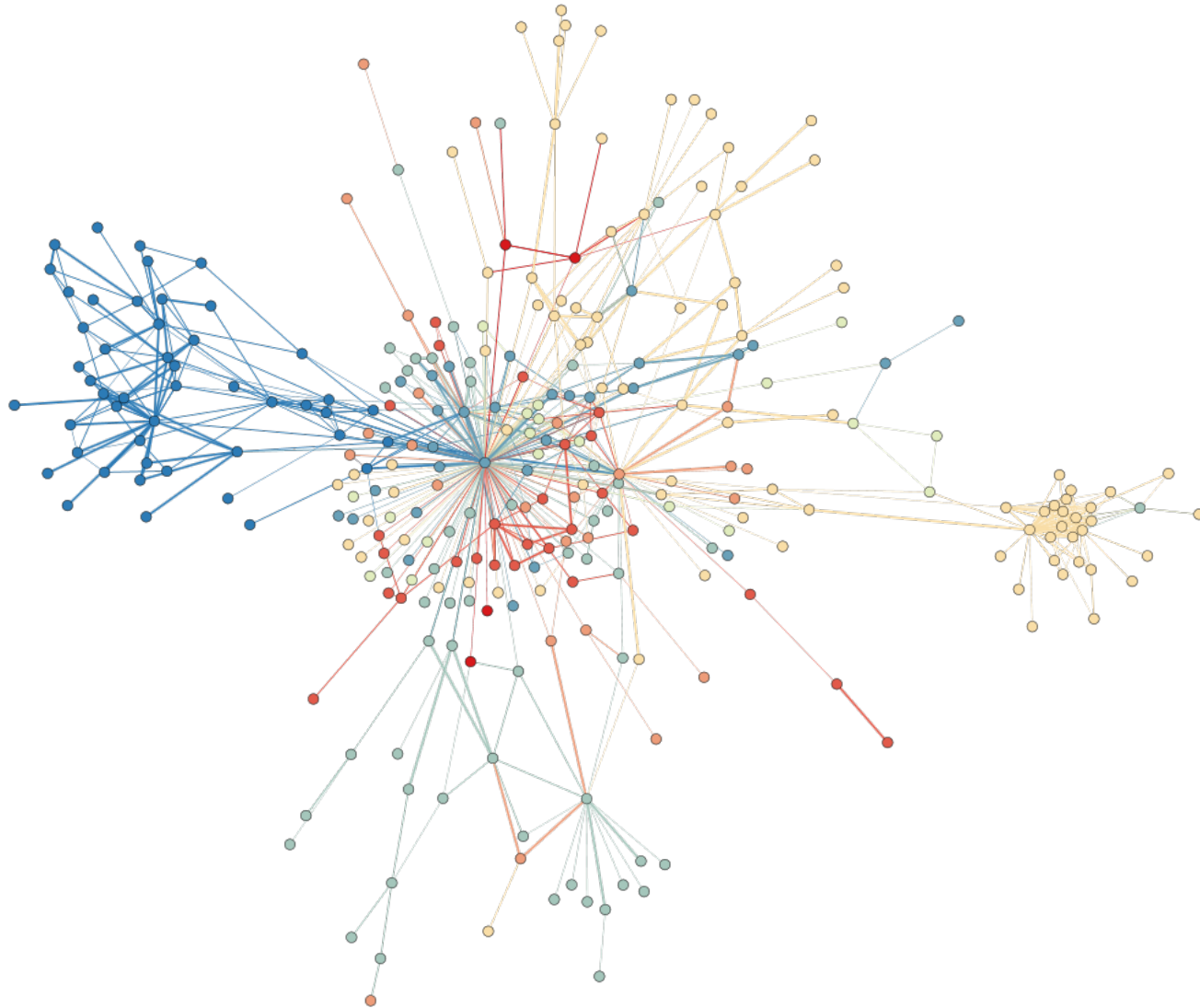
Linked Data: a seed for a Web of Data



- ▶ “Linked Data” is also a set of principles:
 - put things on the Web through URI-s
 - use HTTP URI-s so that things could be dereferenced
 - provide useful information (using standards) when a URI is dereferenced
 - *include links* to other URI-s
- ▶ RDF is an ideal vehicle to realize these principles



But: the number of links among datasets is still small



Linked Data offers major challenges for Semantic Web

- ▶ Scale: we are talking about billions of triples, increasing every day
- ▶ Highly distributed: data spread over the Web, connected via http links
- ▶ Very heterogeneous data of different origins
- ▶ Need not only to read but also to write the data (using the same concepts)
- ▶ Setting a SPARQL endpoint everywhere may not be realistic

Read only vs. Read/Write

- ▶ The current Linked Data work concentrates on *publishing* data for read-only usage
- ▶ The future requires an easy way to read and write data
 - SPARQL UPDATE is a step in this direction
 - but requires an additional SPARQL processor

Planned: Linked Data Platform WG

- ▶ General goal: to provide a “lower” level, HTTP based infrastructure to publish, read, write, or modify linked data
 - typical usage: data intensive application in a browser, application integration using shared data...
- ▶ The infrastructure should be “low-cost”, easy to implement and install

Planned: Linked Data Platform WG

► Two major work areas:

1. Define an HTTP protocol to
 - access and update RDF data through standard HTTP terms
 - define patch, paginated access
 - define a RESTful API
2. Linked Data Profiles: subsets of existing Semantic Web standards to be used for such HTTP based access
 - use only a subset of datatypes
 - use HTTP URI-s only
 - etc.

Planned: Linked Data Platform WG

► Two major work areas:

1. Define an HTTP protocol to

- access and update RDF data through standard HTTP terms
- (possibly) define patch, paginated access
- define a RESTful API

2. Linked Data Profiles: subsets of existing Semantic Web standards to be used for such HTTP based access

- use only a subset of datatypes
- use HTTP URI-s only
- etc.

Planned!!!

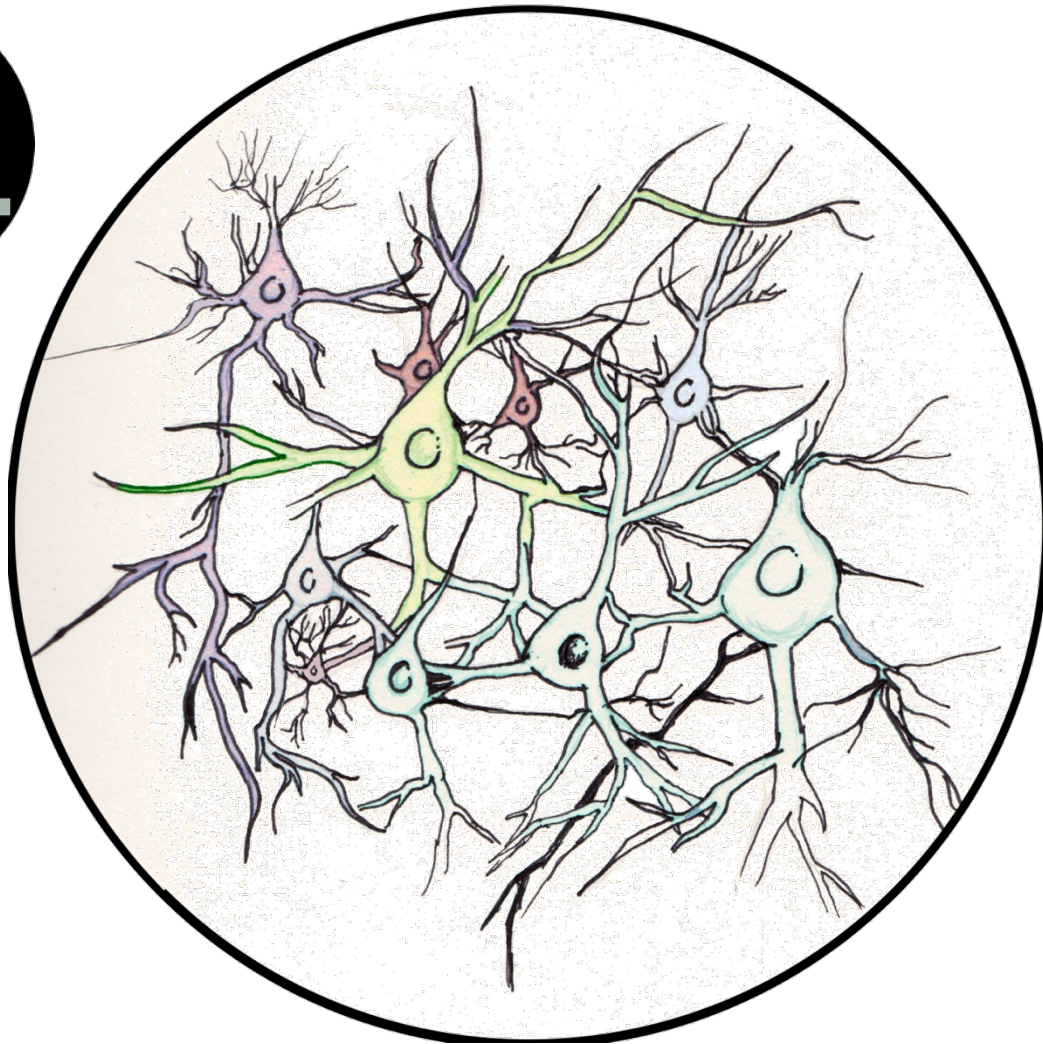
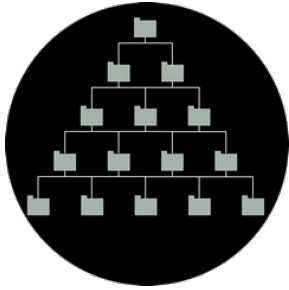


What else is on the horizon?

Further challenges raised by Linked Data

- ▶ Knowledge structures vs. data is very different: very shallow, simple vocabularies for huge sets of data
 - The role of reasoning is different (vocabularies, OWL DL, etc., may not be feasible)
- ▶ Highly distributed SPARQL implementations are necessary
- ▶ etc.

For example: data vs. vocabularies



Other work areas in activity that are explored

- ▶ Profiles for the publication of Linked Data, e.g.,
 - further profiles of OWL
 - URI patterns
 - datatypes
 - usage of Bnodes
 - etc.
- ▶ JSON serialization of RDF

Other work areas in activity that are explored

- ▶ Standardized approaches for Access Control to data
- ▶ Reconsider rule languages for (e.g., for Linked Data applications)
- ▶ Constraint checking of Data
- ▶ API-s for client-side Web Application Developers
- ▶ ...

To remember...

- ▶ *Data* on the Web is a major challenge
 - technologies are needed to use them, to interact with them, to integrate them
- ▶ Semantic Web technologies (RDF(S), vocabularies, SPARQL, etc.) can play a major role in publishing and using Data on the Web

Semantic and Multilingual Web?

(leaving my domain of expertise 😊)

- ▶ Relationships between Semantic Web and multilingualism is twofold:
 - can SW help in achieving MLW?
 - MLW challenges for SW?

SW helping MLW?

- ▶ Semantic Web has powerful technologies to *categorize* knowledge (e.g., SKOS and other vocabulary standards)
 - thesauri can be created with labels in different languages
 - some level of knowledge extraction and analysis could be done on those
- ▶ Via the Linked Data it is possible to interlink *information* different languages
 - DBpedia integrates the various Wikipedia instances
- ▶ It is possible to tag texts using the same terms (e.g., via stable URI-s)
- ▶ etc.

These may help in “binding”, translating, etc., information in different languages

MLW challenges for SW?

SW technologies and practice has to consider the challenges of MLW

- ▶ RDF has a very simple way of representing literals (copied from XML): single language tag. Is it enough?
- ▶ Ontologies/vocabularies are typically monolingual, terms are mostly English...
- ▶ Practice of vocabulary design very often forgets about MLW issues (first name, last name...)
- ▶ IRI equivalence is a major headache in practice
- ▶ etc.

I am looking forward to the discussions!

Thank you for your attention

These slides are also available on the Web:



<http://www.w3.org/2012/Talks/0315-Luxembourg-IH/>