

Semantic Web Activities @ W3C

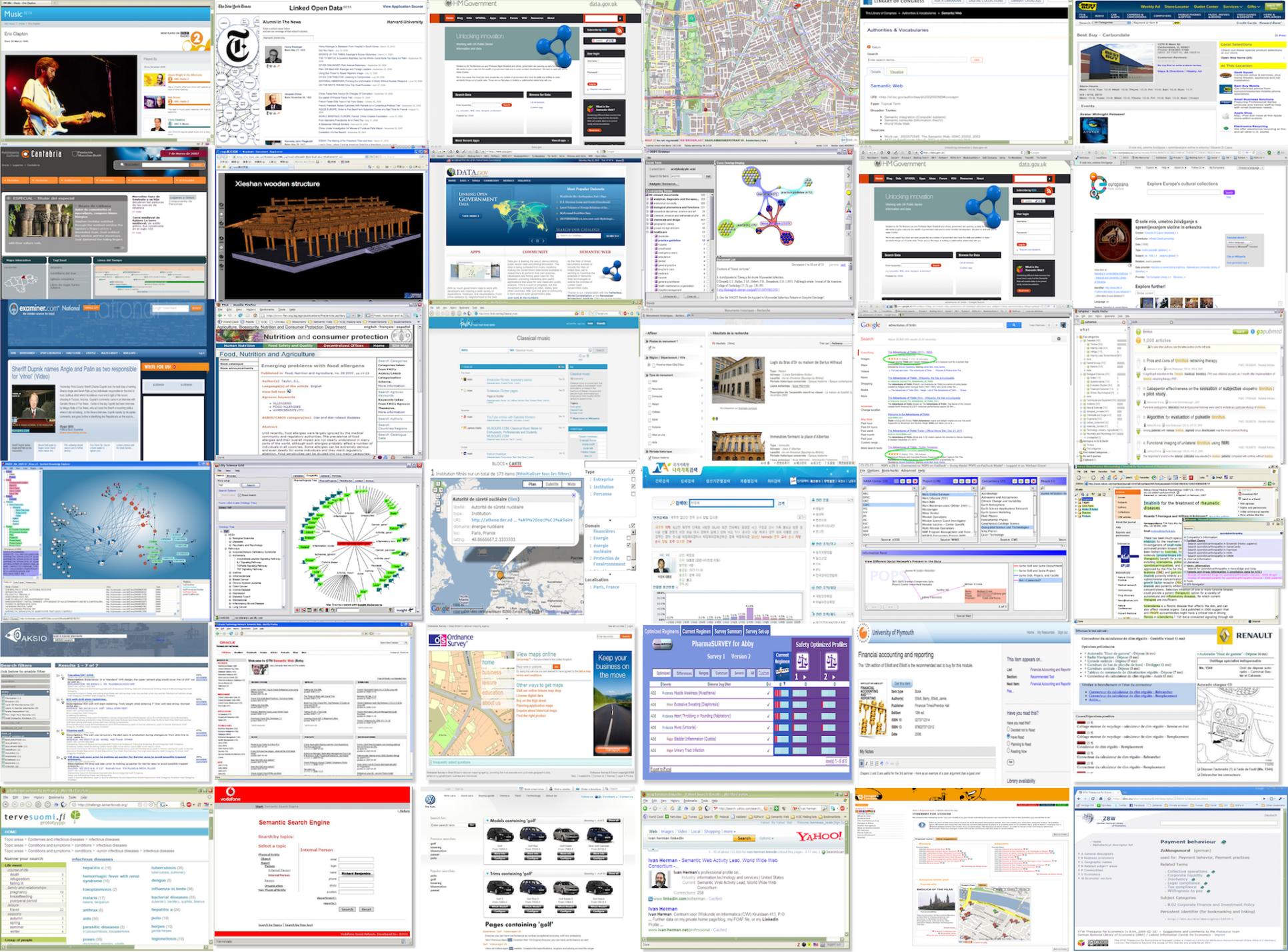
Ivan Herman, W3C

Semantic Technology & Business Conference

5th June, 2012

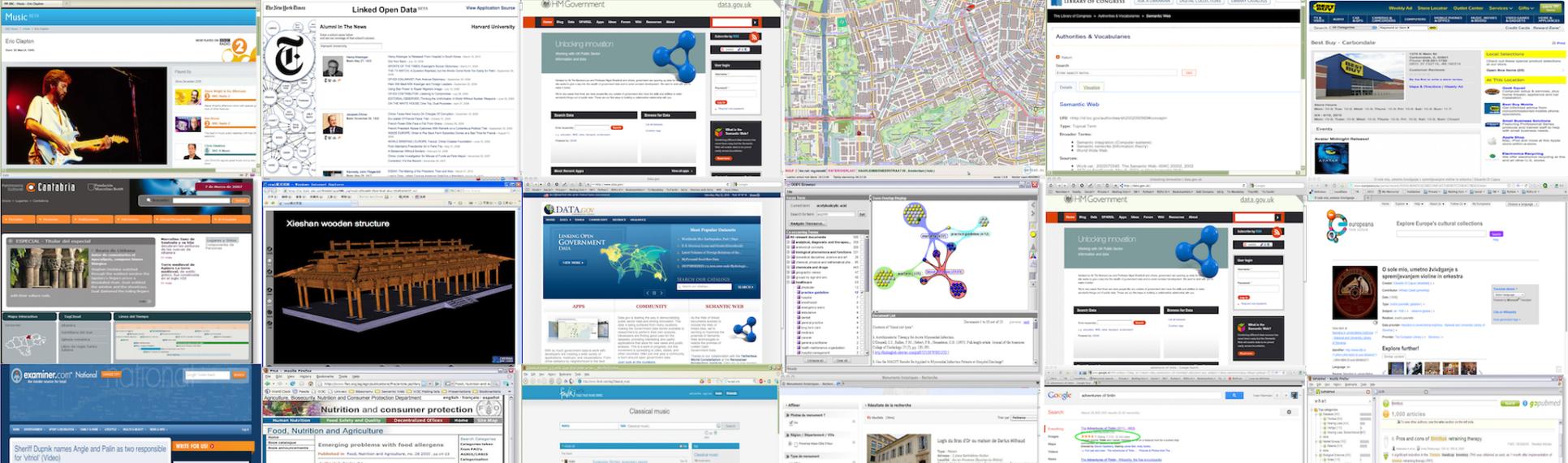
San Francisco, CA, USA



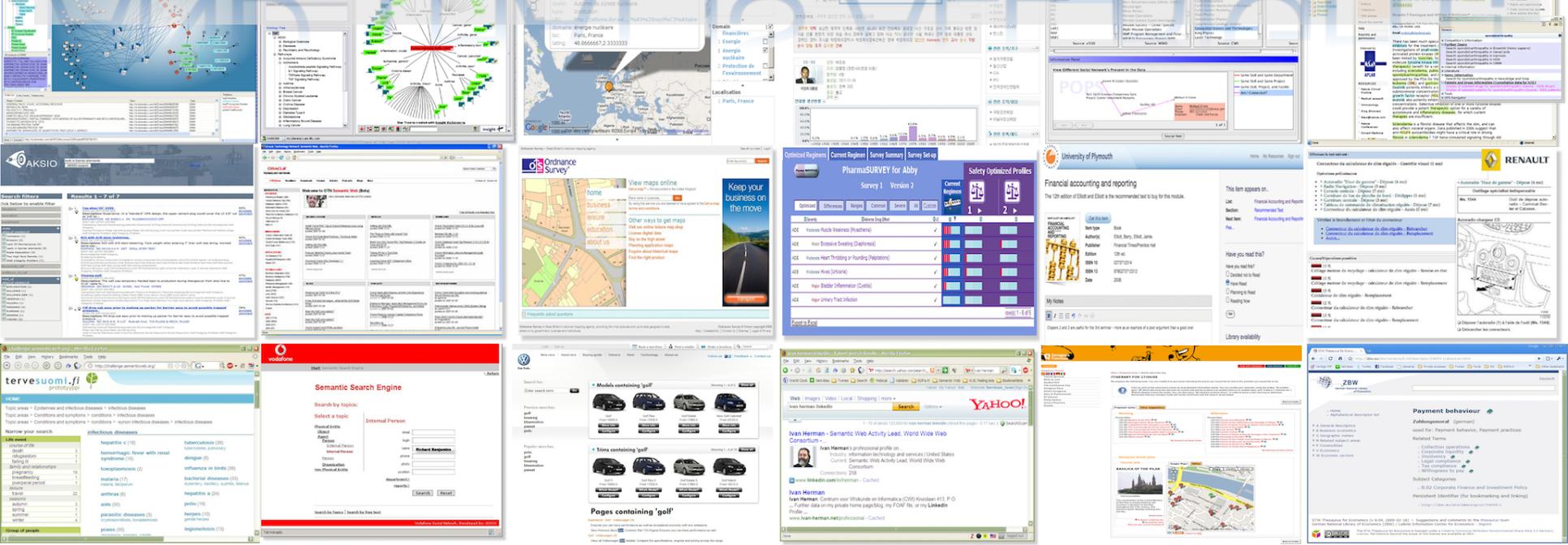


For some people, Semantic (Web) is...

- ▶ A system manipulating and analyzing knowledge
 - e.g., via big ontologies, vocabularies
 - Google's Knowledge Graph?
- ▶ Improve search by adding structure to embedded data
- ▶ A means to *integrate* many different pieces of data
- ▶ Integrate data-oriented applications
- ▶ And a mixture of all these...



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

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

Google™



W3C SEMANTIC WEB ACTIVITY

On this page → [publications, interviews](#) • [presentations](#) • [active groups](#) • [completed groups](#) • [past groups](#)

NEARBY

[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?



[What is the Semantic Web?](#)

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.

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, essentially, done:
 - Ontology for Media Resources
 - Media Fragments URI
 - 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 work
 - Update of RDF
 - Provenance
 - Linked Data Platform

The future

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

Link to specialized communities

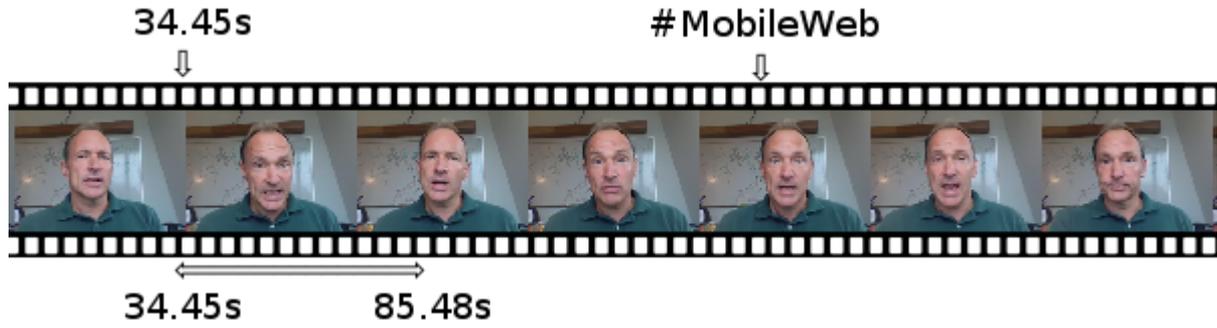
- ▶ Various communities have different emphasis on which part of the Semantic Web they want to use
- ▶ W3C has contacts with some of those
 - health care and life sciences (a separate IG is up and running)
 - libraries, publishing
 - financials
 - the oil, gas, and chemicals community
 - governments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

These community links are not one-way streets!

- ▶ The communities often contribute technologies that can be used in general
- ▶ For example:
 - New vocabularies may come to the fore: SKOS or FRBR (from libraries), annotations (originally from the HCLS work), Person vocabularies (from the eGov work)
 - Health Care had a major influence on the Provenance work
 - etc.



Audio, Video, and Semantics



Audio and Video are now first class entities on the Web

But... video and audio on the Web is not only what you see and hear

— it is also what you can search, discover, distribute, and manage!

Ontologies for Media Resources

- ▶ The “usual” Semantic Web problem: what vocabularies to use?
- ▶ The problem is not that there aren't any... but that there are too many!
 - EXIF, MPEG7, XMP, MRSS, ...
 - none of these cover *all* aspects
- ▶ The [Ontology for Media Resources](#) document
 - defines a core vocabulary
 - defines a set of mappings to other formats

Media Fragment URIs

- ▶ Questions:
 - what is the standard URI for, say, a temporal fragment of a video?
 - what should be the behavior of the user agents for these URIs?
- ▶ These are covered by the [Media Fragments URI](#) document, e.g.
 - `http://www.example.com/video.ogv#t=10,20`
 - `http://www.example.com/video.ogv#track=audio`
 - `http://www.example.com/video.ogv#xywh=160,120,320,240`

Media Work Status

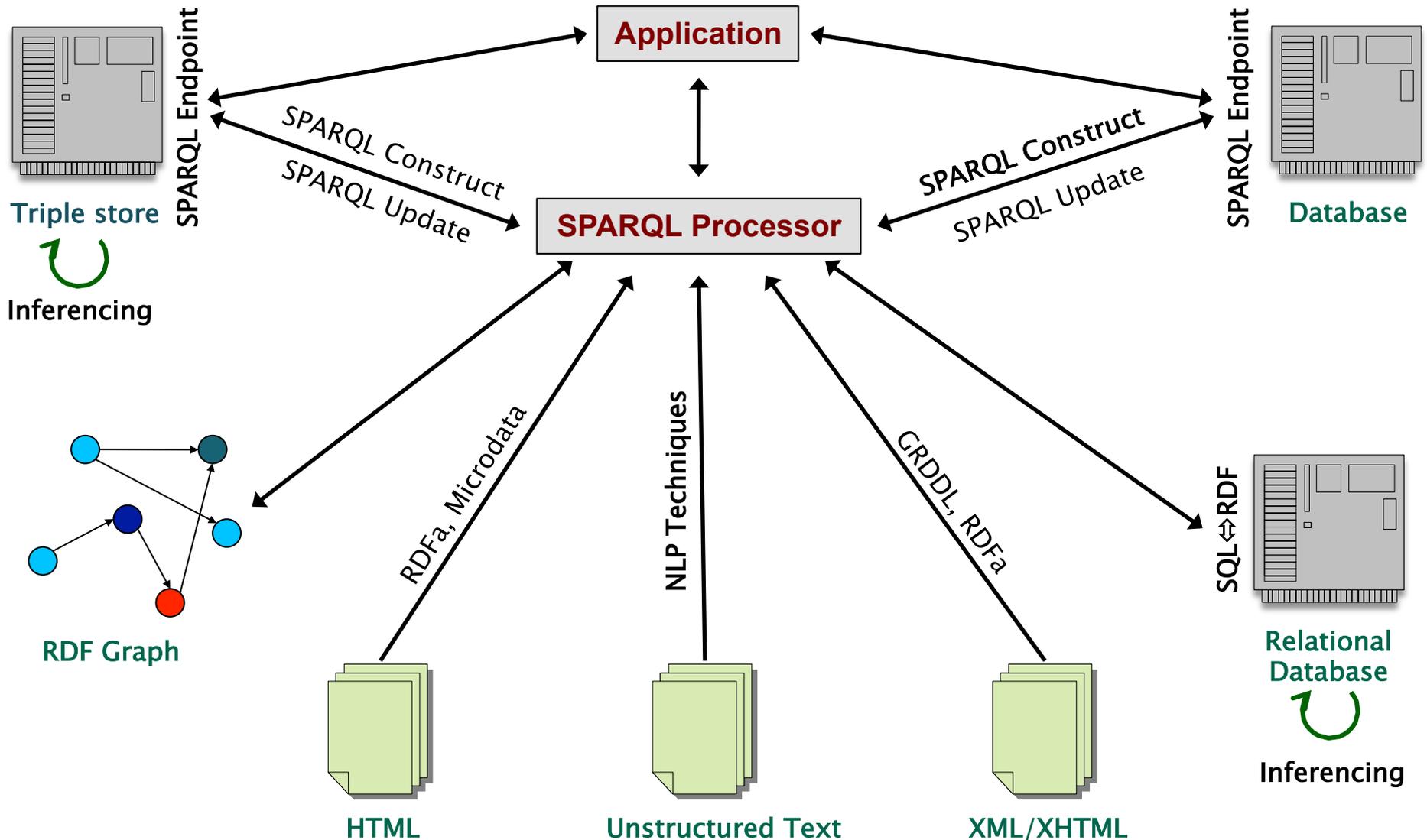
- ▶ Ontologies for Media Resources: published as a Recommendation in February 2012
- ▶ Media Fragments URI: should be published as a Recommendation very soon

Query RDF: SPARQL 1.1

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
- ▶ Return format definition in JSON and in CSV

SPARQL 1.1 as a unifying point



SPARQL 1.1 Status

- ▶ Technology has been finalized
- ▶ Goes to “Proposed Recommendation” soon
- ▶ Should be published as a standard by this fall

Access to Relational Databases

Many RDB systems can handle RDF

- ▶ Relational database vendors realize the importance of the Semantic Web market
- ▶ Many systems have a “hybrid” view:
 - traditional, relational storage, usually coupled with SQL
 - RDF storage, usually coupled with SPARQL
 - examples: Oracle 3g, IBM’s DB2, OpenLink Virtuoso,...

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

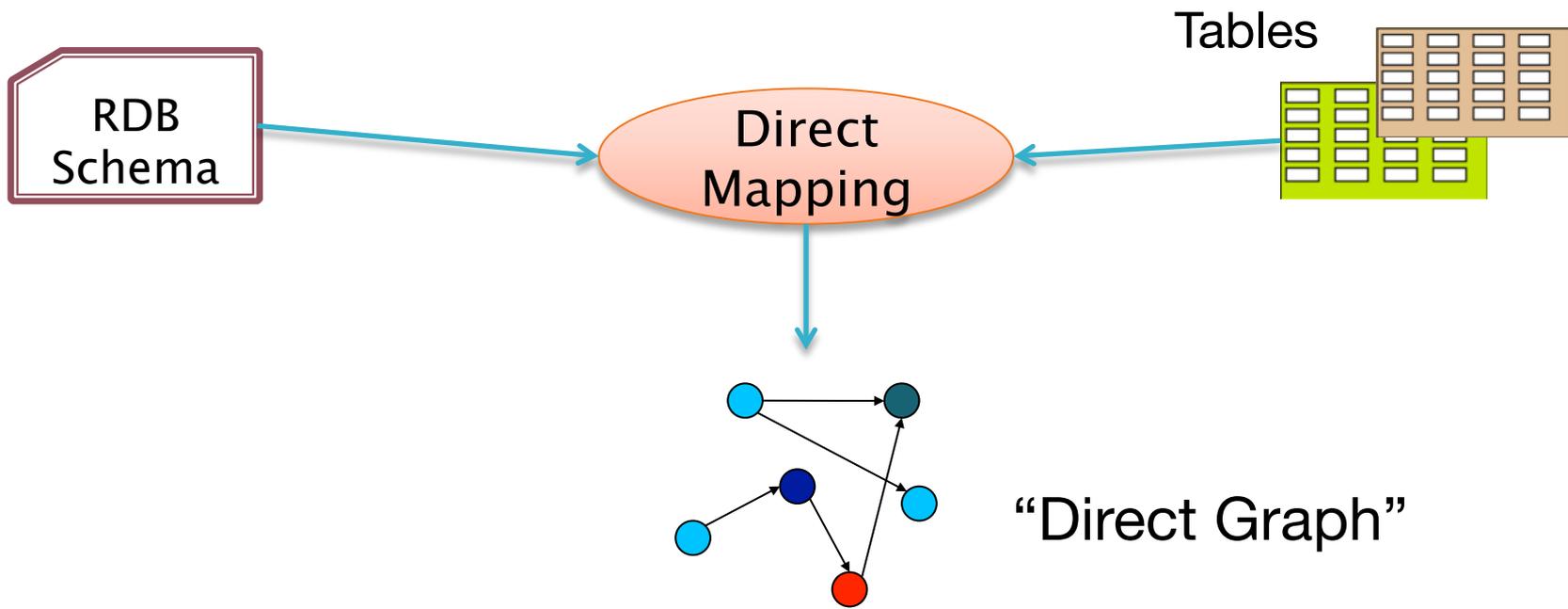
ISBN	Author	Title	Publisher	Year
0006511409X	id_xyz	The Glass Palace	id_qpr	2000
0007179871	id_xyz	The Hungry Tide	id_qpr	2004

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



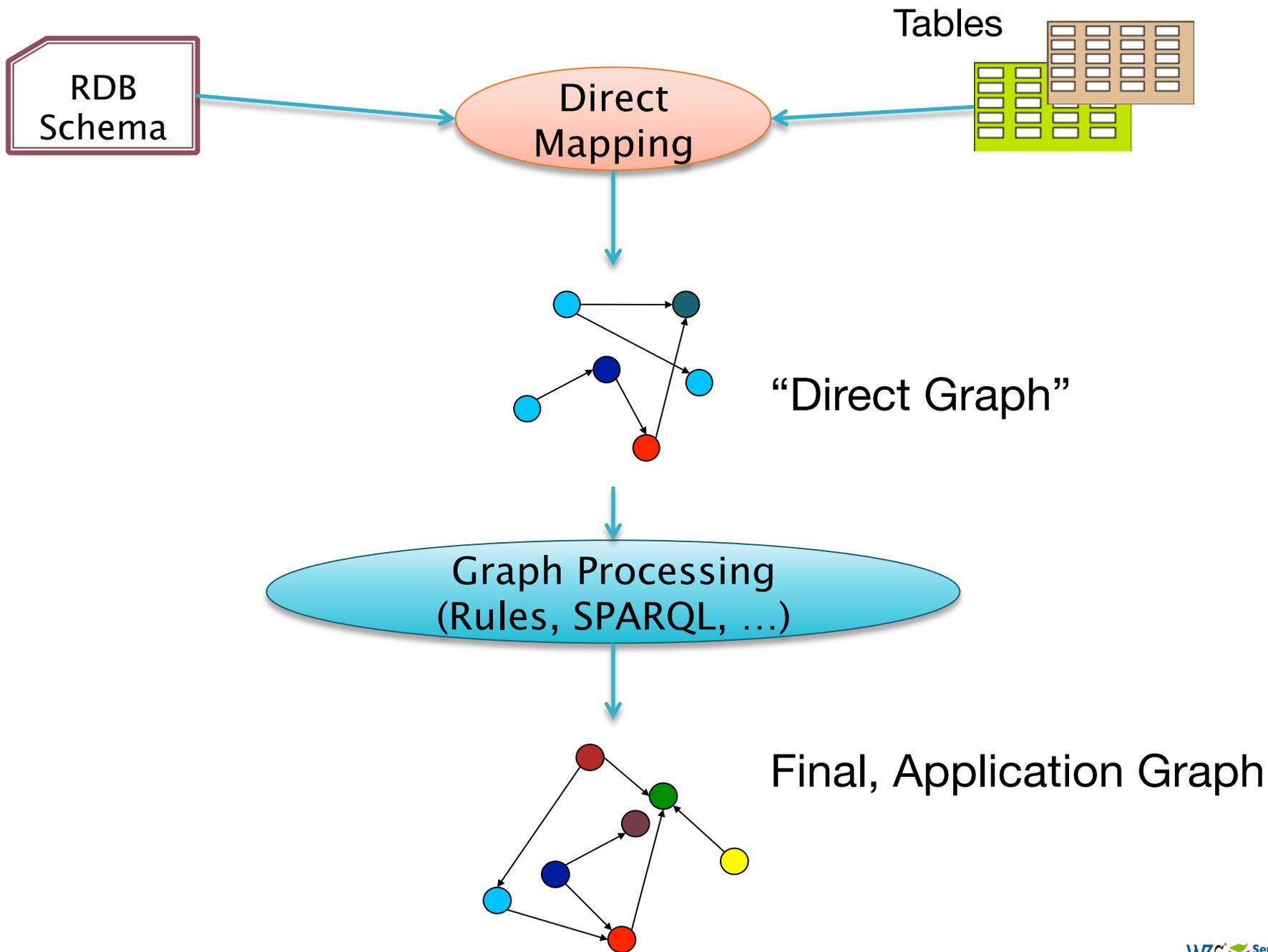
Pros and cons of Direct Mapping

▶ 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(!)

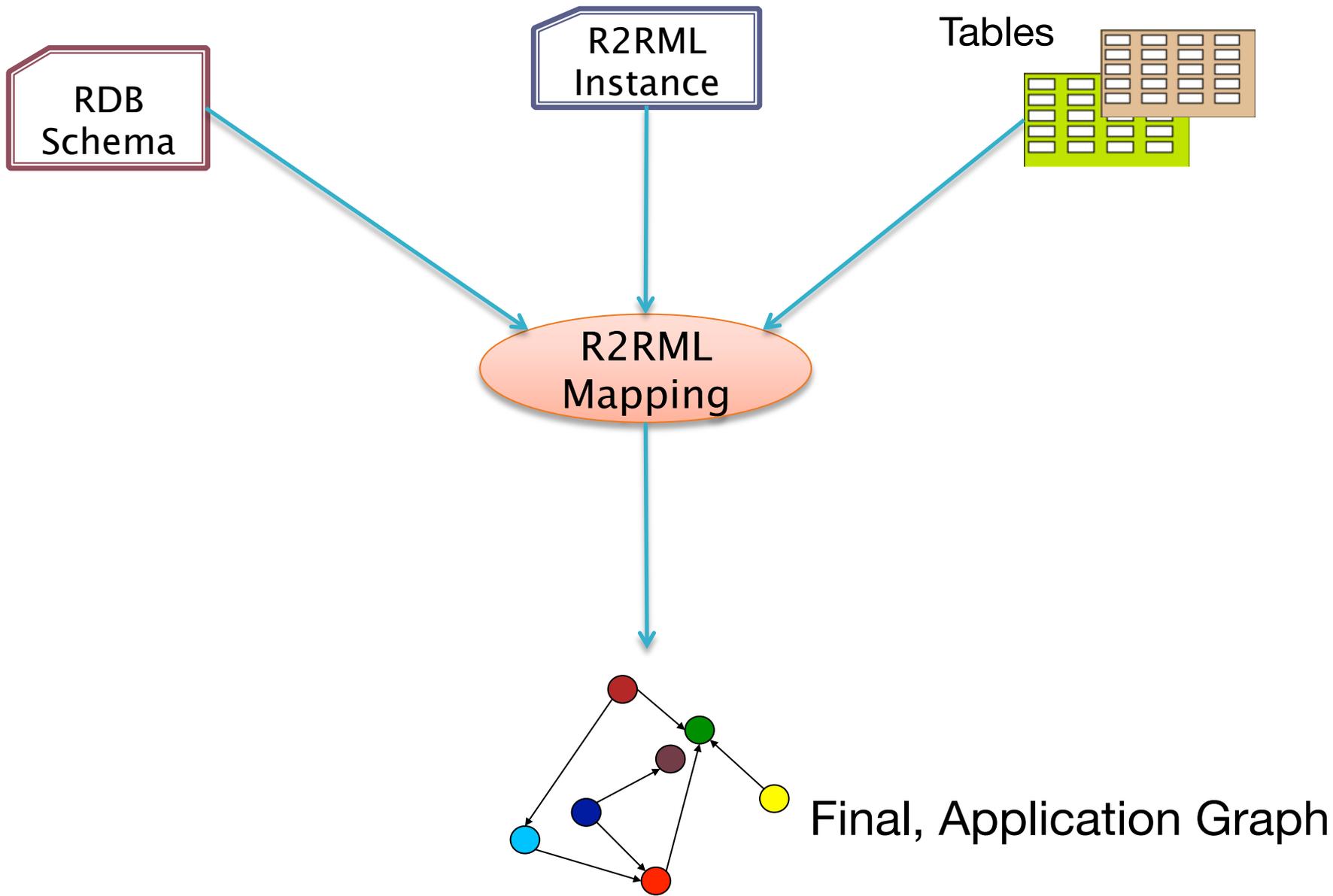
▶ Cons:

- 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
- ▶ Which of the two approaches is used depend on local tools, personal experiences and background,...
 - e.g., user can begin with a “default” R2RML, and gradually refine it

R2RML and Direct Mapping Status

- ▶ Technology has been finalized
- ▶ Implementations revealed some minor issues to fold into the specification
- ▶ Should be finished this summer

At the conference...



- ▶ “Implementations of R2RML”, Wednesday, 8:45am
 - Souripriya Das (Oracle), Jans Aasman (Franz Inc.), Juan Sequeda (Capsenta), and Tony Vachino (Spry, Inc.)

28 PAGINA'S MET ONMISBARE TUTORIALS

UITGAVE 35

GRATIS CD

145+ minuten videotutorials en podcasts, WordPress-thema van Themify.me, 50 textures van MediaMilitia, 16 vectors van WeGraphics

www.webdesignermagazine.nl

web designer

HTML / AJAX / CSS / PHP / MySQL / Silverlight / Expression

DW Dreamweaver FI Flash Ps Photoshop

HTML5

Structured data in HTML: RDFa & microdata

Een spannende toekomst voor webstandaarden

JQUERY GAAT MOBIEL

Multiplatformdevelopment wordt weer een stukje eenvoudiger

TUTORIALS

- Ontwerp een professioneel WordPress-thema
- Zet je eerste stappen met jQuery Mobile
- Maak een HTML5-pagina met nieuwe lay-out-tags
- Soepele afbeeldingseffecten met jQuery

these days

IE9-PROMOTIE EN HTML5-EXPERIMENT
These Days presenteert Microsofts designwedstrijd ExtremeRate Fest



IPAD-MAGAZINES MET INDESIGN
Alles over het maken van digitale magazines voor de Apple iPad

UITGAVE 35 € 8,95



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



Ivan Herman

Who am I?

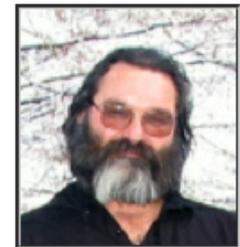
I graduated as mathematician at the [Eötvös Loránd University of Budapest](#), Hungary, in 1979. After a brief scholarship at the Université Paris VI I joined the Hungarian research institute in computer science ([SZTAKI](#)) where I worked for 6 years (and turned into a computer scientist...). I left Hungary in 1986 and, after a few years in industry in Munich, Germany, I joined the [Centre Mathematics and Computer Sciences \(CWI\)](#) in Amsterdam where I have a tenure position since 1988. I received a PhD degree in Computer Science in 1990 at the [University of Leiden](#), in the Netherlands. I joined the [World Wide Web Consortium \(W3C\)](#) Team as Head of [W3C Offices](#) in January 2001 while maintaining my position at CWI. I served as Head of Offices until June 2006, when I was asked to take the [Semantic Web Activity](#) Lead position, which is now my principal work at W3C.

Before joining W3C I worked in quite different areas (distributed and dataflow programming, language design, system programming), but I spend most of my research years in computer graphics and information visualization. I also participated in various graphics related ISO standardization activities and software developments. My ["professional" home page](#) contains a list of [my publications](#) (see also [my Mendeley account](#)), [my public presentations](#), and details of the various projects I participated in the past. There is also a [dblp entry for my publications](#) generated automatically (although I am not sure it is complete...). (B.t.w., based on my publications, my [Erdős number](#) is ≤ 4 ...)

In my previous life (i.e., before joining W3C...) I was member of the Executive Committee of the [Eurographics Association](#) for 15 years, and I was vice-chair of the Association between 2000 and 2002. I was the co-chair of the [9th World Wide Web Conference](#), in Amsterdam, May 2000; since then, I have also been member of [IW3C2 \(International World Wide Web Conference Committee\)](#), responsible for the World Wide Web Conference series. Since autumn 2007 I am also member of [SWSA \(Semantic Web Science Association\)](#), the committee responsible for the International Semantic Web Conferences (better known as "ISWC") series.

Some personal data

- The Hungarian spelling of my [full name is Herman Iván](#). I.e, my name is Ivan (well, spelled properly: Iván) and my [surname is Herman](#) (many in the Netherlands and in Germany mix it up, and use "Herman" as my name... this is aggravated by the fact that, uniquely in Europe, the Hungarian custom is to put surname first).
- Nationalities: French and Hungarian
- Gender: male
- Family: I am married and have a son, David.
- Date and city of birth: 24th February, 1955, [Budapest](#), Hungary
- Email addresses: 'ivan' on my own ivan-herman.net domain, 'ivan' on the w3.org domain, or 'ivan.herman' on the cwi.nl domain
- (Mobile) Phone: +31-641044153
- Skype ID: ivan_herman
- I live in [Amstelveen](#) (see also [geonames](#)), the Netherlands (lat: 52.302063, long: 4.87397). This is a suburb of [Amsterdam](#). The closest airport is Amsterdam Schiphol
- I am the administrator of the [Semantic Web Activity Blog](#) at W3C which can either be [accessed directly](#) or via [its](#)



some links

- [personal homepage](#)
- [more data on me](#)
- [personal blog \(RSS feed\)](#)
- [homepage at W3C](#)
- ["professional" homepage](#)
- ["official" CV](#)
- [more about me](#)
- [my photos](#)

"social" links

- [facebook](#)
- [flickr](#)
- [picasa web](#)
- [linkedin](#)
- [IWIW](#)
- [tripit](#)
- [twitter](#)
- [Mendeley](#)
- [Google+](#)
- [freebase](#)


```

<meta property="foaf:accountName" content="ivan-herman" />
</a>
</li>
<li>
  <a href="https://plus.google.com/u/0/113268051484517627727" typeof="foaf:OnlineAccount">
    <span property="foaf:accountServiceHomepage" href="http://www.mendeley.com/">Google</span>
    <meta property="foaf:accountName" content="113268051484517627727" />
  </a>
</li>
<li>
  <a about="http://www.ivan-herman.net/foaf#me" rel="owl:sameAs" resource="http://rdf.freebase.com/ns/en.
</li>
</ul>
</div>
</div>
<div id="content" >
  <h1 property="schema:name foaf:name">Ivan Herman</h1>
  <meta property="foaf:title" content="Dr" />
  <h2>Who am I?</h2>
  <p>I graduated as mathematician at the <a rel="foaf:schoolHomepage schema:alumniOf" href="http://www.elte.hu/"><span
I joined the <a rel="schema:worksFor" href="http://www.w3.org" resource="http://www.w3.org/Data#W3C">
  <span property="dc:title">World Wide Web Consortium (W3C)</span>
</a> Team as Head of <a rel="foaf:pastProject" href="http://www.w3.org/Consortium/Offices"><span property="dc:title"
<link rel="owl:sameAs" href="http://www.ivan-herman.net/me" />
<link rel="owl:sameAs" href="http://www.ivan-herman.net/Ivan_Herman" />
<link rel="foaf:workplaceHomepage" href="http://www.w3.org"/>
<meta property="schema:jobTitle" content="Semantic Web Activity Lead" />

<p>Before joining W3C I worked in quite different areas (distributed and dataflow programming, language design, syst
</p>

<p>In my previous life (i.e., before joining W3C...) I was member of the Executive Committee of the <a rel="foaf:pastI
</p>
<h2>
Some personal data
</h2>
<ul>
  <li>The Hungarian spelling of my full name is <span property="foaf:name" lang="hu">Herman Iván</span>.
  Ie, my name is <span property="foaf:givenname schema:givenName">Ivan</span> (well, spelled properly:
  <span property="foaf:givenname schema:givenName" lang="hu">Iván</span>) and my surname is
  <span property="foaf:surname schema:familyName">Herman</span>
  (many in the Netherlands and in Germany mix it up, and use "Herman" as my name... this is aggravated by the fact t
</li>
  <li>Nationalities: <span property="schema:nationality">French</span> and <span property="schema:nationality">W

```

Yielding...

```
<http://www.ivan-herman.net/foaf#me>
  schema:alumniOf      <http://www.elte.hu> ;
  foaf:schoolHomePage <http://www.elte.hu> ;
  schema:worksFor      <http://www.w3.org/W3C#data> ;
  ...
<http://www.elte.hu>
  dc:title "Eötvös Loránd University of Budapest" .
  ...
<http://www.w3.org/W3C#data>
  dc:title "World Wide Web Consortium (W3C)"
  ...
```

The Telegraph

Search - enhanced by Google

Monday 09 April 2012

HOME NEWS SPORT FINANCE COMMENT BLOGS **CULTURE** TRAVEL LIFESTYLE FASHION TECH Dating Offers Jobs

Film Music Art Books TV and Radio Theatre Hay Festival Dance Opera Photography Comedy Video In the Know

Oscars Film Reviews Cinema Trailers Coming Soon Talking Movies Interviews DVDs Film Life Film Video

HOME » CULTURE » FILM » FILM REVIEWS

Oscars 2012: The Artist, review

The Artist, an utterly beguiling silent, black-and-white celebration of early Hollywood won Best Picture at the Oscars 2012.

★★★★★



To attend an Info Session or Masterclass Or download a brochure

[Click here](#)

INSEAD
The Business School for the World*

Global Executive MBA

+1

TELEGRAPH TICKETS »



The Telegraph

Search - enhanced by Google

Monday 09 April 2012

HOME NEWS FINANCE COMMENT BLOGS CULTURE TRAVEL LIFESTYLE FASHION TECH

Dating Offers Jobs

TV Radio TV and Radio Theatre Day Festival Dance Opera Photography Comedy Video In the Know

Technology Games Travel Cooking Sora Talking Movies Interviews DVDs Film Life Film Video

Oscars 2012: The Artist, review

The Artist, an utterly beguiling silent, black and white celebration of early Hollywood won Best Picture at the Oscars 2012

ratingValue="4.5" data-bbox="188 504 414 522">



To attend an Info Session or Masterclass Or download a brochure

[Click here](#)

INSEAD
The Business School for the World®

Global Executive MBA

TELEGRAPH TICKETS »



```

<li class="list-item">
</li><a href="#">
</a>
</div>
</div>
<!-- googleon: all -->
<div id="tmglBody" >
<div class="access"><a name="a">
</a>
</div>
<div class="twoThirdsThird2 gutter">
<div class="twoThirds gutter">
<div class="story">
<h1 itemprop="name">Oscars 2012: The Artist, review
<h2 itemprop="description">
The Artist, an utterly beguiling silent, black and white celebration of early Hollywood won Best Picture at the Oscars 2012
</h2>
<div class="rating" itemprop="reviewRating">
<meta itemprop="worstRating" value="1">
<meta itemprop="bestRating" value="5">
<span itemprop="ratingValue" value="4.5">

<div id="storyEm
<div class="slideshow ssIntro">
<div class="nextPrevLayer">
</div>
</div>
</div>
</div>
</div>
<div class="oneHalf gutter">
<div class="story">
<div class="cl"> </div>
<!-- remove the whitespace added by escenic before end of

```

```
<li class="first"><a href="/">Home</a><span>&raquo;</span></li>
<li><a href="http://www.telegraph.co.uk/culture/">Culture</a><span>&raquo;</span></li>
  <li><a href="http://www.telegraph.co.uk/culture/film/">Film</a><span>&raquo;</span></li>
  <li class="styleSix"><a href="http://www.telegraph.co.uk/culture/film/filmreviews/">Film R
</div>
</div>
<!-- googleon: all -->
<div id="tmglBody" >
  <div class="access"><a name="article"></a></div>

  <div class="twoThirdsThird2 gutterUnder">
    <div class="twoThirds gutter" itemscope itemtype="http://schema.org/Review">
      <div class="storyHead">

        <h1 itemprop="name">Oscars 2012: The Artist, review</h1>
        <h2 itemprop="description">
The Artist, an utterly beguiling silent, black-and-white celebration of early
Hollywood won Best Picture at the Oscars 2012.
        </h2>

        <div class="rating" itemprop="reviewRating" itemscope itemtype="http://schema.org/Rating">
          <meta itemprop="worstRating" content = "0.5">
          <meta itemprop="bestRating" content = "5">
          <span itemprop="ratingValue" class="hidden">5</span>
          
        </div>
        <div class="artIntro">
          <div id="storyEmbSlide">
            <div class="slideshow ssIntro">
              <div class="nextPrevLayer">
                <div class="ssImg">
                  

                    <div class="ingCaptionCredit">
                      <span class="caption">B er n ce Bejo as ris
                    </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
<div class="oneHalf gutter">
  <div class="story">
    <div class="cl"> </div>
    <!-- remove the whitespace added by escenic before end of </a> tag -->
```

Yielding...

```
[ rdf:type schema:Review ;  
  schema:name "Oscars 2012: The Artist, review" ;  
  schema:description "The Artist, an utterly beguiling..." ;  
  schema:ratingValue "5" ;  
  ...  
]
```

Everything

Images

Maps

Videos

News

Shopping

More

Amsterdam

Change location

Any time

Past hour

Past 24 hours

Past week

Past month

Past year

Custom range...

More search tools

The Artist showtimes for Amsterdam

[Pathe Tuschinski](#) - Reguliersbreestraat 26-34, Amsterdam - [Map](#)

11:50 - 14:05 - 19:10

[Filmtheater "De Uitkijk"](#) - Prinsengracht 452, Amsterdam - [Map](#)

12:15 - 19:00 - 21:15

[Filmtheater Rialto](#) - Ceintuurbaan 338, Amsterdam - [Map](#)

12:45

[+ Show more theaters](#)

The Artist (2011) - IMDb

www.imdb.com/title/tt1655442/

Silent **movie** star George Valentin bemoans the coming era of talking ... Still of Jean Dujardin and Missi Pyle in **The Artist** Still of Bérénice Bejo in **The Artist** Reem ...

[Full cast and crew](#) - [The Artist Trailer \(Official ...](#) - [Bérénice Bejo](#) - [Jean Dujardin](#)

The Artist (film) - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/The_Artist_\(film\)](http://en.wikipedia.org/wiki/The_Artist_(film))

The Artist is a 2011 French romantic comedy drama in the style of a black-and-white silent **film** written and directed by Michel Hazanavicius, starring Jean ...

[Jean Dujardin](#) - [Bérénice Bejo](#) - [Uggie](#) - [Diegesis](#)

The Artist Trailer 2011 HD - YouTube



www.youtube.com/watch?v=O8K9AZcSQJE

25 Aug 2011 - 3 min - Uploaded by TrailersApplecom

I love how George Clooney, and Brad Pitt, lost the Best actor category to this **film**. It just shows that there is ...

[More videos for the artist movie »](#)

Oscars 2012: The Artist, review - Telegraph

www.telegraph.co.uk/Culture/Film/Film_Reviews

★★★★★ Review by Robbie Collin

27 Feb 2012 – **The Artist**, the final **film** to be released in 2011 and also the most heart-swellingly joyful one, is a silent **movie**, screened in black and white and ...

[The Artist is the perfect film about Hollywood | Hadley Freeman](#)

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

RDFa and microdata: differences

- ▶ Microdata has been optimized for simpler use cases:
 - 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

Structured data in HTML is mainstream!

... 25% of webpages containing RDFa data [...] over 7% of web pages containing microdata.

[Mail from Peter Mika, Yahoo!](#)

*Based on a crawl evaluation by P. Mika and T. Potter
LDOW2012 Workshop, April 2012, Lyon, France*

... web pages that contain structured data has increased from 6% in 2010 to 12% in 2012.

[Hannes Mühleisen and Christian Bizer](#)

*Web Data Commons—Extracting Structured Data from Two Large Web Corpora,
LDOW2012 Workshop, April 2012, Lyon, France*

RDFa 1.1 and microdata status

- ▶ For RDFa 1.1
 - Technology has been finalized
 - Is in Proposed Recommendation
 - Should be published as a Recommendation any day now
- ▶ For microdata
 - Technology has been finalized
 - There is [microdata→RDF](#) mapping in a separate Note
 - Is part of HTML5, hence its formal advancement depends on other technologies

At the conference...



- ▶ “Schema.org panel”, Wednesday, 9:45am
 - Dan Brickley (schema.org), Ramanathan Guha (Google), Steve Macbeth (Microsoft), Peter Mika (Yahoo!), Jeff Preston (Disney Interactive), Evan Sandhaus (NYT), Alexander Shubin (Yandex); moderator: Ivan Herman (W3C)

Cleaning up RDF



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/plans

- ▶ Standardize Turtle as a serialization format
- ▶ Clean up some aspects of datatyping, e.g.:
 - plain vs. typed literals
 - introduction of an `rdf:HTML` datatype
 - better definition 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
- ▶ Standardize a JSON format for linked data (JSON-LD)

Editorial improvements

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

Status

- ▶ Turtle is almost finalized
- ▶ Literal cleanup is done
- ▶ JSON-LD is in a good shape
- ▶ Lots of discussion currently on named graphs...
- ▶ A new version of the RDF 1.1 concepts has just been published this morning!
 - <http://www.w3.org/TR/rdf11-concepts/>

At the conference...



- ▶ “Updates to the Core RDF Standards”, Wednesday, 3:30pm
 - David Wood (3 Round Stones, Inc., also co-chair of the W3C RDF WG)
- ▶ “JSON-LD: JSON for Linked Data”, Thursday, 9:45am
 - Gregg Kellogg (Kellogg Associates)

Provenance

$$R_{ik} = 0$$

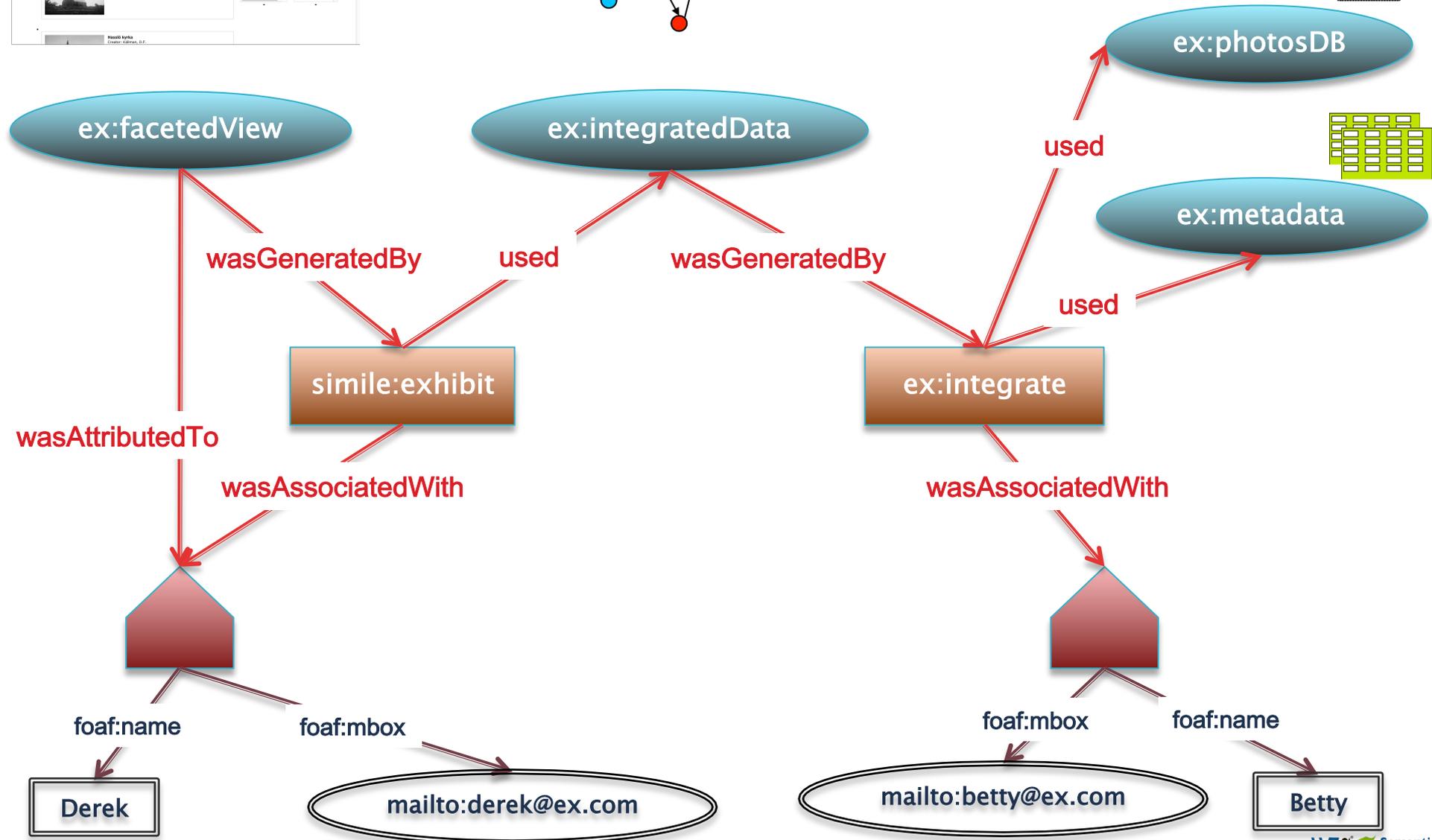
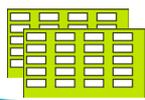
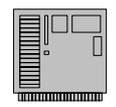
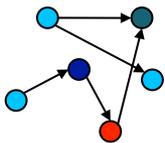
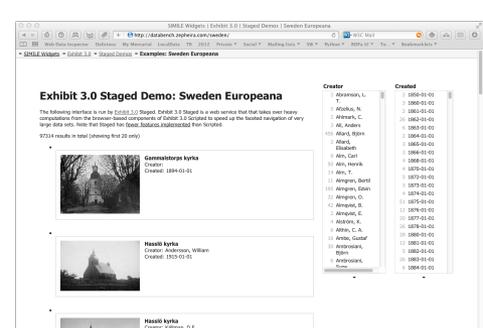


The goal is simple...

- ▶ We should be able to express all sorts of “meta” information on the data
 - who played what role in creating the data (author, reviewer, etc.)
 - view of the full revision chain of the data
 - in case of data integration: 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 provenance: easily usable and editable
 - complex 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
 - protocol (where to find provenance data)
 - primer
- ▶ Goal is to finalize the technical design in the fall of 2012

At the conference...



- ▶ “Benefits and Applications of W3C’s Provenance Standards in Enterprise Semantic Web Applications”, Thursday, 10:45am
 - Reza Bfar (Oracle)

Recherche Aix-en-Provence



Affiner

Photos du monument ?

- Non 4
- Oui 2

Région / Département / Ville

- Provence-Alpes-Côte D'Azur 6

Type de monument

- Eglise 3
- Monument 1
- Hôtel de ville 1
- Monastère 1

Période historique

- Moyen Âge 6
- Époque moderne 5
- Renaissance 2
- Époque contemporaine 1

Gares les plus proches

- Gardanne 6
- Aix-en-Provence 6
- Simiane 6

Type de propriété

Résultats de la recherche

6 résultats (105ms)

Trier par Pertinence Afficher 10 résultats par page



Hôtel de Ville

Type : Hôtel de ville
 Adresse : place de l' Hôtel-de-Ville
 Localité : Aix-en-Provence (Bouches-du-Rhône)
 Période historique concernée : Moyen Âge - Renaissance - Époque moderne -
 Liens externes : [Base Mérimée](#) - [Wikimedia Commons](#)

Description de l'ensemble inscrit ou classé : Hôtel de Ville (cad. AS 63) : classement par arrêté du 12 octobre 1995

Mis à disposition sur [Wikimedia Commons](#)

Eglise Saint-Jean-de-Malte

Type : Eglise
 Localité : Aix-en-Provence (Bouches-du-Rhône)
 Période historique concernée : Moyen Âge - Époque moderne -
 Liens externes : [Base Mérimée](#) - [Wikipedia francophone](#) - [Wikimedia Commons](#)

Description de l'ensemble inscrit ou classé : Eglise Saint-Jean-de-Malte : classement par liste de 1840

Mis à disposition sur [Wikimedia Commons](#)

Some characteristics of Linked Data and its Applications

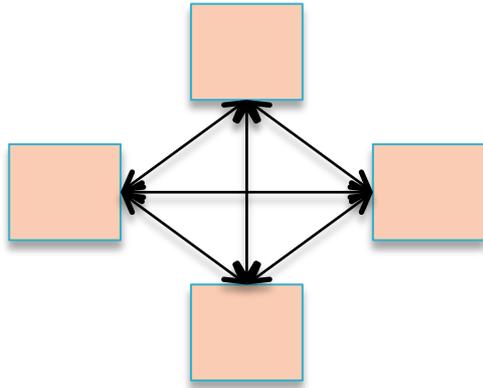
- ▶ The datasets are essentially read-only
 - they are curated “out of band”: regularly extracted from other databases, changed manually by data owners, etc
- ▶ The dominating paradigm is to extract data via SPARQL queries
- ▶ Applications use (very) large datasets via (RDF based) integration

However... Linked Data Has Potentials
for “Simpler” Applications

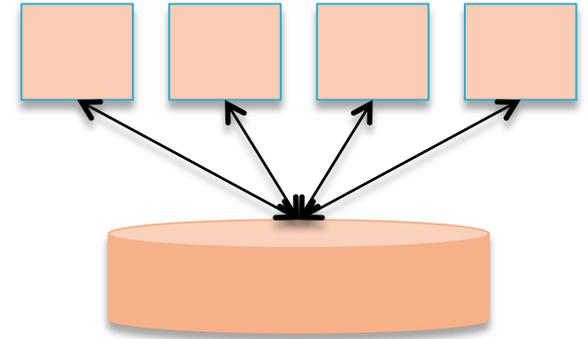
Example: Data-intensive application integration

- ▶ Application Lifecycle Management
 - integration of development teams around the globe
 - management of bug report, user requirements
 - versioning
- ▶ Distributed access to, and management of Library Catalogue data
- ▶ Integrated view of corporate and private address book data

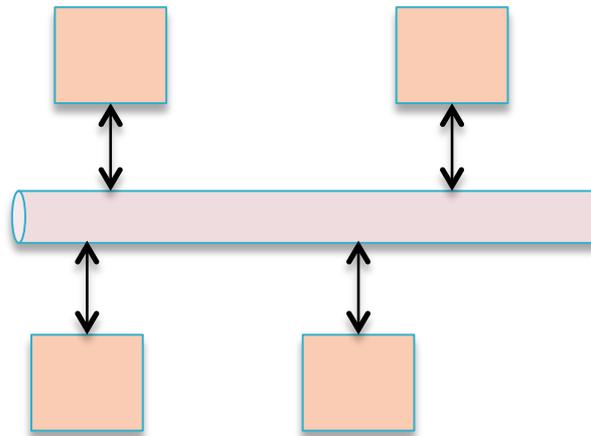
There has been approaches in the past



Point-to-point via API



Centralized repository



Central Hub/Bus

None of these are really satisfactory 😞

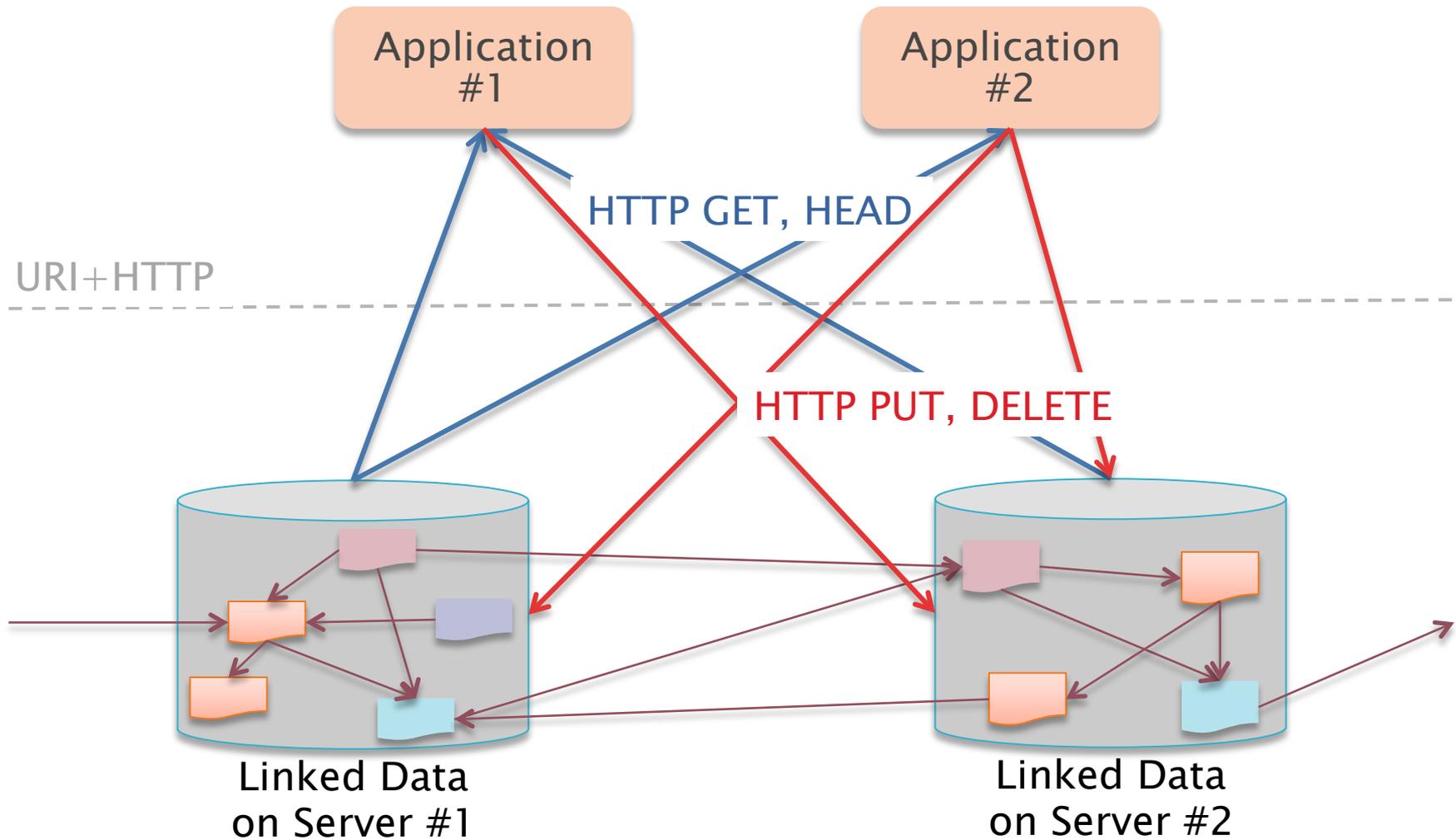
- ▶ They force to re-invent the wheel on many fronts
 - distribution of data around the Internet
 - access control issues
 - definition and implementation of new API-s, Protocols, data formats
 - etc.

Why not make use of an architecture that is...

- ▶ Distributed at its core
- ▶ Scalable in terms of users, of data, of hardware and software architecture
- ▶ Open to anyone
- ▶ Has a wide variety of available tools
- ▶ Widely known and deployed

Sounds Familiar? 😊

Linked Data Provides a Better Paradigm: Use the Existing Web Architecture!



Linked Data Platform WG

- ▶ Provide a simple, HTTP based infrastructure to publish, read, write, or modify linked data
- ▶ The infrastructure should be easy to implement and install
 - more complex applications may require more sophisticated tools like SPARQL, Provenance, OWL,...
 - provides an “entry point” for Linked Data applications!

Linked Data Platform WG

▶ Main Work items:

- define a RESTful way to access/update RDF data via HTTP
 - what does HTTP GET/PUT/DELETE/POST/... mean for Linked Data?
- define a “profile” of minimal requirements for applications:
 - what RDF datatypes are used
 - what serialization syntax(es) must be supported
 - how to access reasonable chunks of information (paging)
 - how to manage collections of RDF data
 - what vocabulary items to use for metadata
 - etc.

Status

- ▶ Has just started a few days ago!

At the conference...



- ▶ “Linked Enterprise Data Patterns”, Tuesday, 11:30am
 - David Wood (3 Round Stones, Inc.), Arnaud Le Hors (IBM), Ashok Malhotra (Oracle)
- ▶ LDP Working Group BOF: Tuesday, 12:30pm, Franciscan “C”



What else may be on the horizon?

Some challenges raised by Linked Data

- ▶ Knowledge vs. data ratio is different: very shallow, simple vocabularies for huge sets of data
 - the role of reasoning is different (vocabularies, OWL DL, etc., may not be feasible)
- ▶ Not enough links among datasets
 - lots of work on “creating” further links
- ▶ Scale: billions of triples, increasing every day
 - setting a SPARQL endpoint everywhere may not be realistic
- ▶ Highly distributed
 - data may not be in one single database, even within the same organization

Possible future works in the activity

- ▶ Check on the quality of the published data
- ▶ Reconsider rule languages for (e.g., for Linked Data applications)
- ▶ Relationship to JSON
- ▶ Constraint checking of Data
- ▶ API-s for client-side Web Application Developers

Possible future works in the activity

- ▶ Issues around internationalization of Semantic Web technologies
- ▶ Relationship between Semantic Web technologies and Big Data, Cloud Storage and Computing,...
- ▶ Specific standard vocabularies (e.g., data annotation, governmental vocabularies)
 - some of these may be defined at W3C, some elsewhere

Lot remain to be done...

- ▶ Lots of issues to be solved
- ▶ But... W3C needs experts!
 - consider joining W3C, as well as the work done there!

Enjoy The Conference!



Thank you for your attention

These slides are also available on the Web:

<http://www.w3.org/2012/Talks/0605-SemTech-IH/>

