

How to get and create RDF Data?

(Microformats, GRDDL, RDFa, POWDER)

Simple approach

- ▶ Write RDF/XML, RDFa, or Turtle “manually”
- ▶ In some cases that is necessary, but it really does not scale...

RDF with XHTML

- ▶ Obviously, a huge source of information
- ▶ By adding some “meta” information, the same source can be reused for, eg, data integration, better mashups, etc
 - typical example: your personal information, like address, should be readable for humans and processable by machines
- ▶ Two solutions have emerged:
 - use microformats and convert the content into RDF
 - add RDF statements directly into XHTML via RDFa

Microformats

- ▶ Not a Semantic Web specification, originally
 - there is a separate microformat community
- ▶ Approach: re-use (X)HTML attributes and elements to add “meta” information
 - typically @abbr, @class, @title, ...
 - different agreements for different applications

Microformat example: hCalendar

- ▶ Goal: “markup” calendaring information on your HTML page
 - use a community agreement using, eg, :
 - @class for event name
 - abbr element for dates
 - @title for date values
 - etc.
- ▶ Automatic procedures (ie, calendaring applications) may then get to the right data

Microformat example: hCalendar

Browser window: Dan Connolly, W3C
URL: <http://www.w3.org/People/Connolly/>

W3C Dan Connolly
Research Scientist, [MIT/CSAIL](#)

connolly@w3.org *
32 Vassar Street
Room 32-G506 *
Cambridge, MA, 02139 USA
+1-617-395-0241, DanC *

>>> **hCard**

standards: [HTML WG](#), [TAG](#), [GRDDL WG](#), [RDF Calendar](#),
[QA](#), [DAWG/SPARQL](#), [Semantic Web IG](#), [OWL](#), [HTML 2](#),
[ESW](#)

research: [breadcrumbs](#) journal/weblog, [cwm](#), [N3](#), [tabulator](#), [PAW](#), [TAMI](#),
[microformats](#) open source
life: [family](#), [volleyball](#), [guitar](#)

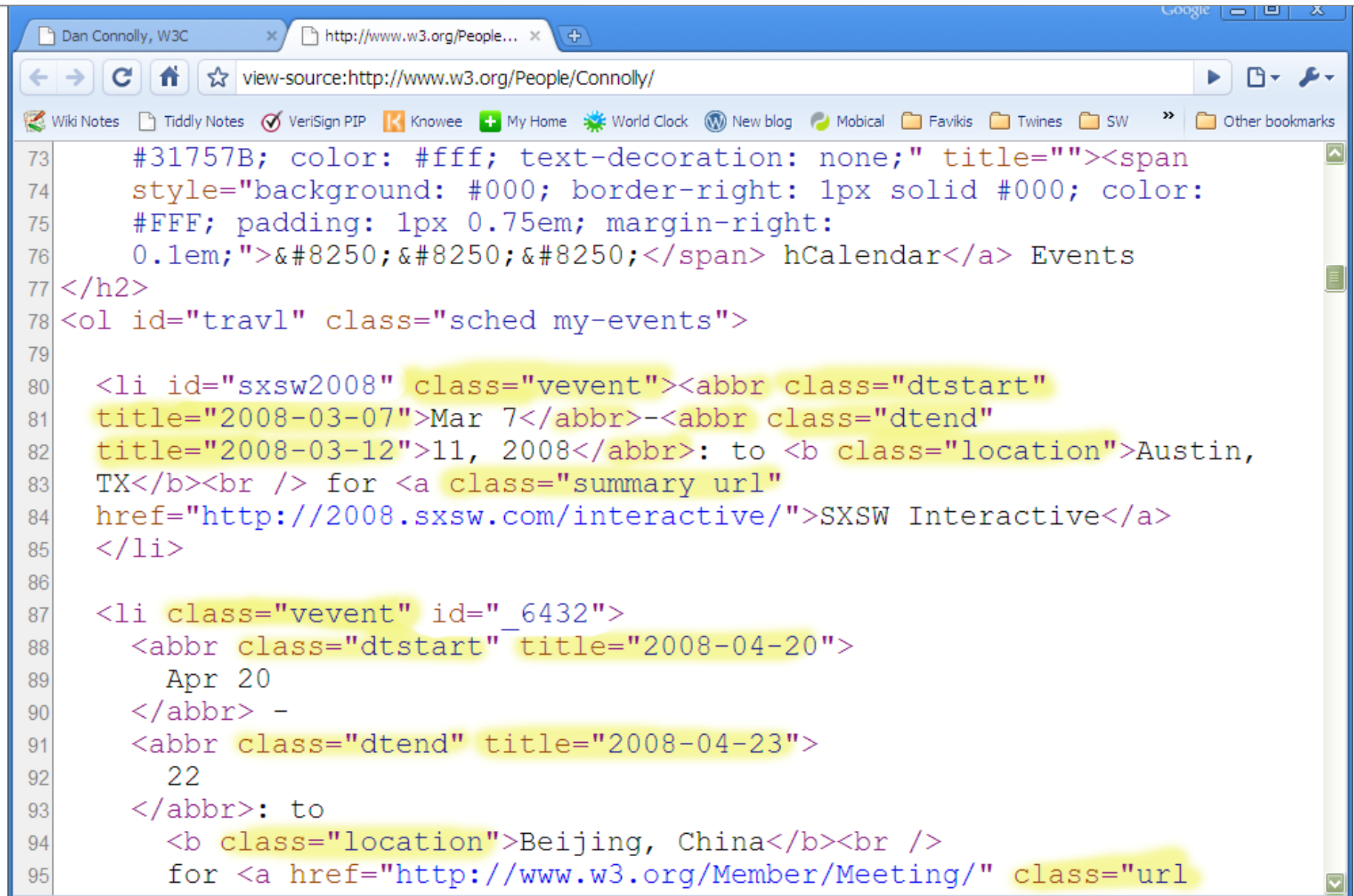


hCalendar Events

- Mar 7-11, 2008: to **Austin, TX**
for [SXSW Interactive](#)
- Apr 20 - 22 : to [Palo Alto, CA](#)
for [W3C AC meeting](#), [linked data workshop](#)
trip stuff
- May 19-May 22: to **Bristol**
for [TAG ftf](#)
trip stuff
- Sep 23 - 25 : in **Kansas City**
[TAG meeting](#)
- Oct 20 - 25 : to **NCE**
for [W3C TPAC](#)
- Nov 1 - 3 : to **BOS**
[TAMI meeting](#)
- Dec 08 - 11 : to **Cambridge, MA**
[TAG meeting](#)
itin
- Feb 2 to 7: to **Denver**
hoping to go for [Web Directions North](#)

Earlier travel/talks/events include [Tools of](#)

Behind the scenes...



```
73 #31757B; color: #fff; text-decoration: none;" title=""><span
74 style="background: #000; border-right: 1px solid #000; color:
75 #FFF; padding: 1px 0.75em; margin-right:
76 0.1em;">&#8250;&#8250;&#8250;</span> hCalendar</a> Events
77 </h2>
78 <ol id="travl" class="sched my-events">
79
80 <li id="sxsw2008" class="vevent"><abbr class="dtstart"
81 title="2008-03-07">Mar 7</abbr>-<abbr class="dtend"
82 title="2008-03-12">11, 2008</abbr>: to <b class="location">Austin,
83 TX</b><br /> for <a class="summary url"
84 href="http://2008.sxsw.com/interactive/">SXSW Interactive</a>
85 </li>
86
87 <li class="vevent" id="_6432">
88 <abbr class="dtstart" title="2008-04-20">
89 Apr 20
90 </abbr> -
91 <abbr class="dtend" title="2008-04-23">
92 22
93 </abbr>: to
94 <b class="location">Beijing, China</b><br />
95 for <a href="http://www.w3.org/Member/Meeting/" class="url
```

Microformat extraction

- ▶ To use it on the Semantic Web, microformat data should be converted to RDF
- ▶ A simple transformation (eg, in XSLT) can be defined, yielding:

```
<http://www.w3.org/People/Connolly/#sxsw2008>  
  a hcal:Vevent;  
  hcal:organizer <http://www.w3.org/People/Connolly/#me>;  
  hcal:summary "SXSW Interactive";  
  hcal:dtstart "2008-03-07"^^xsd:date;  
  hcal:dtend "2008-03-12"^^xsd:date;  
  hcal:url <http://2008.sxsw.com/interactive/>;  
  hcal:location "Austin, TX" .
```

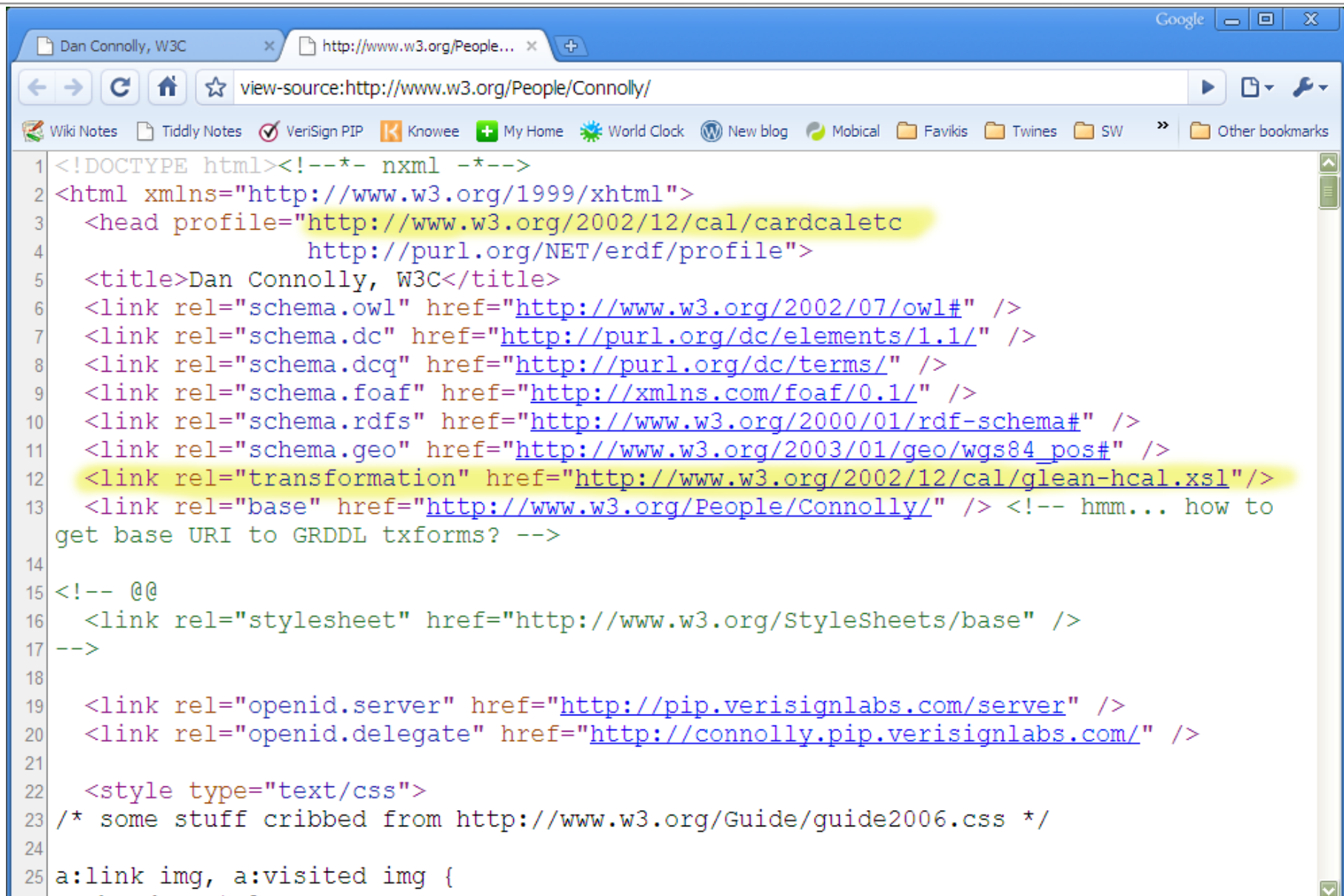

So far so good, but...

- ▶ The XSLT transformation is hCalendar specific
 - each microformat dialect needs its own
- ▶ How does a general processor find the right transformation?
- ▶ Enter GRDDL

GRDDL: find the right transformation

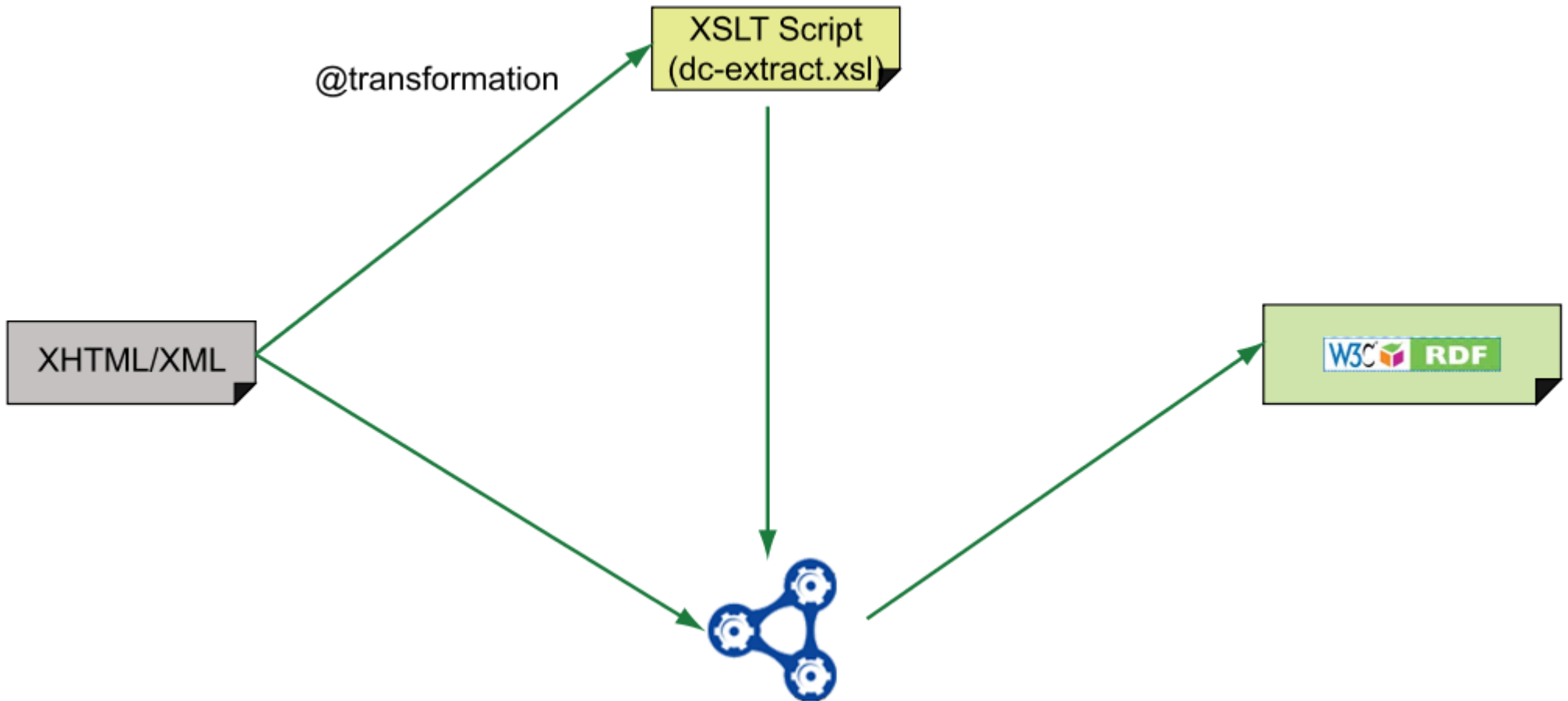
- ▶ GRDDL defines
 - a few extra attribute values to locate the right transformation
 - a precise processing model on how the transformation should be applied to generate RDF
- ▶ Note: we describe GRDDL in terms of XHTML (and microformats) but GRDDL can be used for any XML data

GRDDL: find the right transformation

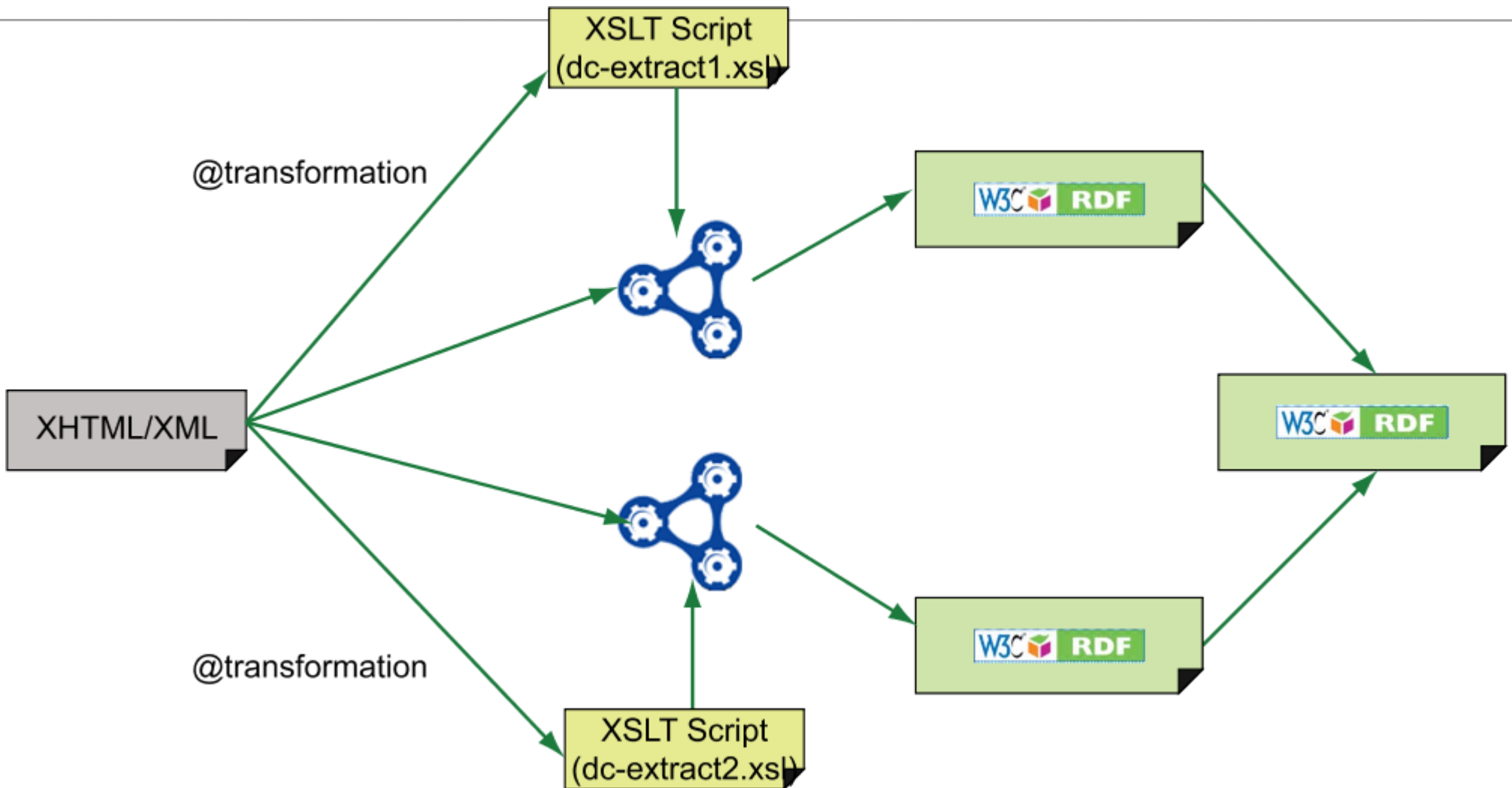


```
1 <!DOCTYPE html><!--*- nxml -*-->
2 <html xmlns="http://www.w3.org/1999/xhtml">
3   <head profile="http://www.w3.org/2002/12/cal/cardcaletc
4     http://purl.org/NET/erdf/profile">
5   <title>Dan Connolly, W3C</title>
6   <link rel="schema.owl" href="http://www.w3.org/2002/07/owl#" />
7   <link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
8   <link rel="schema.dcq" href="http://purl.org/dc/terms/" />
9   <link rel="schema.foaf" href="http://xmlns.com/foaf/0.1/" />
10  <link rel="schema.rdfs" href="http://www.w3.org/2000/01/rdf-schema#" />
11  <link rel="schema.geo" href="http://www.w3.org/2003/01/geo/wgs84_pos#" />
12  <link rel="transformation" href="http://www.w3.org/2002/12/cal/glean-hcal.xsl"/>
13  <link rel="base" href="http://www.w3.org/People/Connolly/" /> <!-- hmm... how to
  get base URI to GRDDL txforms? -->
14
15  <!-- @@
16    <link rel="stylesheet" href="http://www.w3.org/StyleSheets/base" />
17  -->
18
19  <link rel="openid.server" href="http://pip.verisignlabs.com/server" />
20  <link rel="openid.delegate" href="http://connolly.pip.verisignlabs.com/" />
21
22  <style type="text/css">
23  /* some stuff cribbed from http://www.w3.org/Guide/guide2006.css */
24
25  a:link img, a:visited img {
```

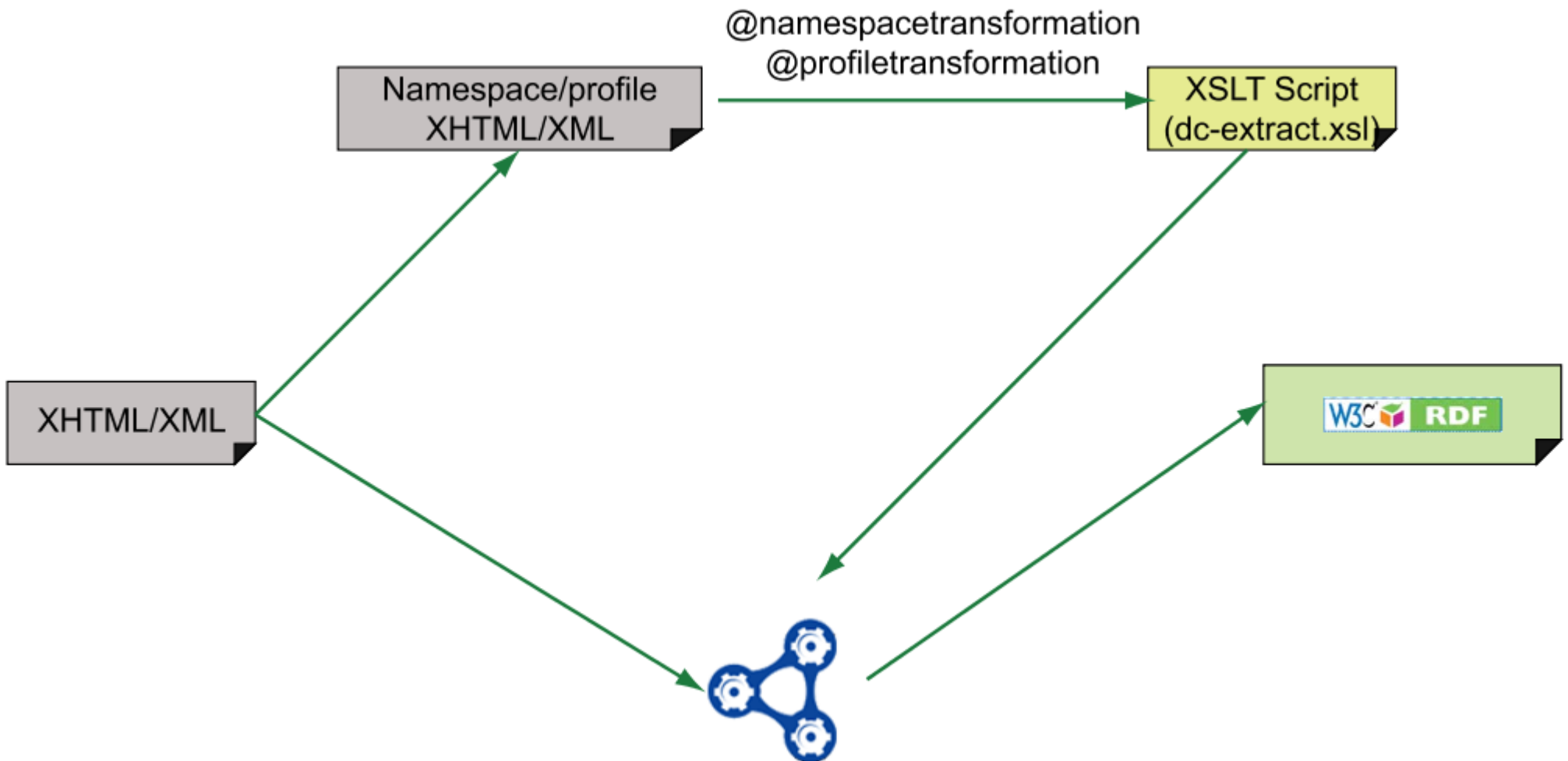
The GRDDL process: simple case



The GRDDL process: merging case



The GRDDL process: indirect case



Microformats & GRDDL: pros

▶ Pros:

- simple to define/add new vocabularies
 - there is a strong microformat community for this
- works with all current browsers, markup validators, etc
- fairly user friendly, easy to understand and use

Microformats & GRDDL: cons

▶ Cons:

- does not scale well for complex vocabularies
 - remember: needs a transformation per vocabulary
- difficult to mix vocabularies within one page
 - what if the usage of an attribute clashes among different vocabularies?
- some of the attributes are meant for other usage
 - eg, the abbr element, the @title attribute, ...

An alternative solution: XHTML+RDFa

- ▶ RDFa also uses (X)HTML attributes to add “meta” information
- ▶ However
 - it also uses proprietary attributes to avoid clashes with the intended usage of the (X)HTML ones
 - it includes a namespace+URI mechanism for disambiguation
 - it is one set of attributes for any vocabularies

XHTML+RDFa example

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 for 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, 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 ...)

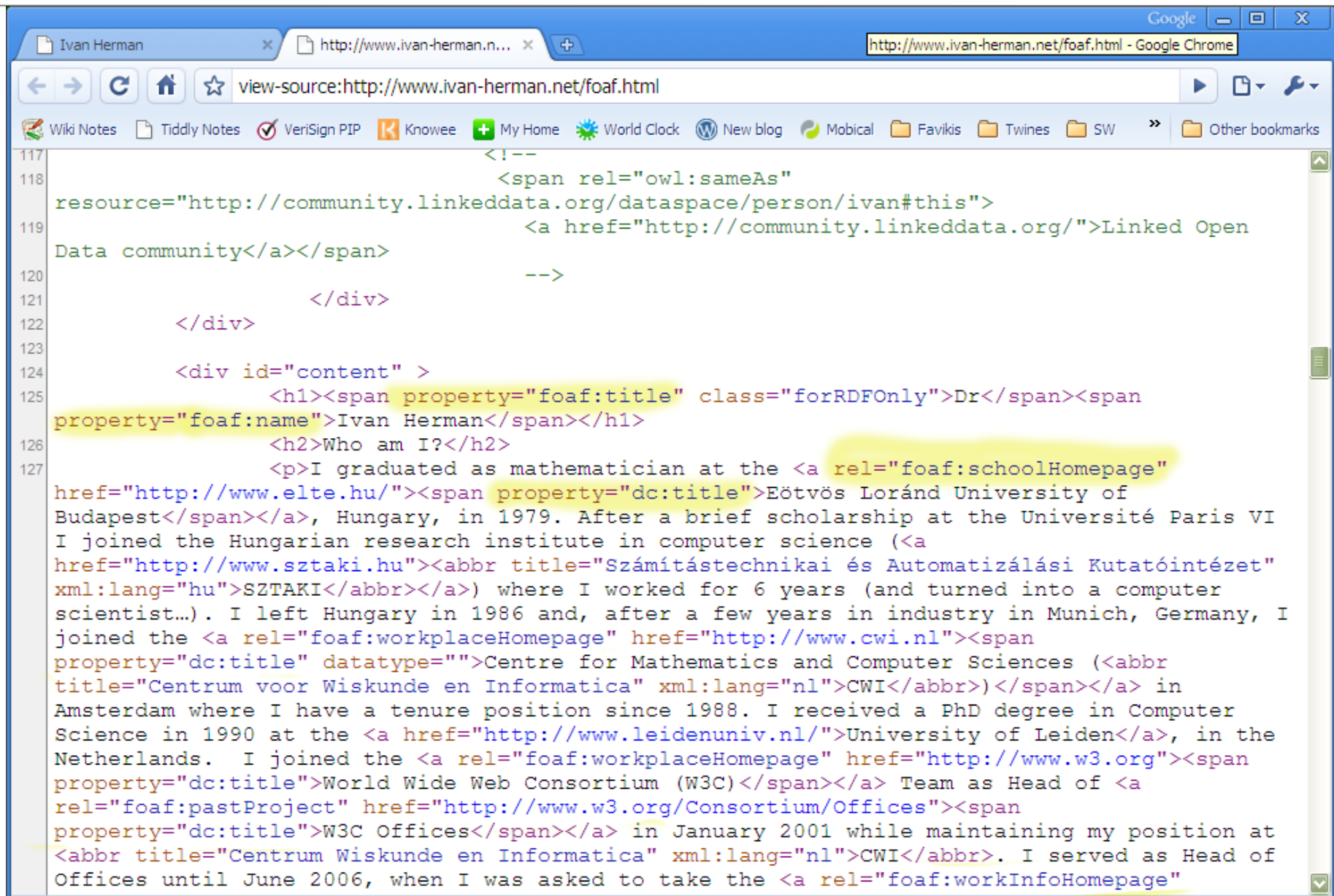
In my previous life (i.e., before joining W3C...) I was member of the Executive Committee of the Eurographics Association for 15 years, and I was vice-chair of the Association between 2000 and 2002. I was the co-chair of the 9th World Wide Web Conference, in Amsterdam, May 2000; since then, I have also been member of IW3C2 (International World Wide Web

some links

- personal homepage
- more data on me
- personal blog (RSS feed)
- a separate blog in Hungarian
- homepage at W3C
- "professional" homepage
- "official" CV
- more about me
- my photos

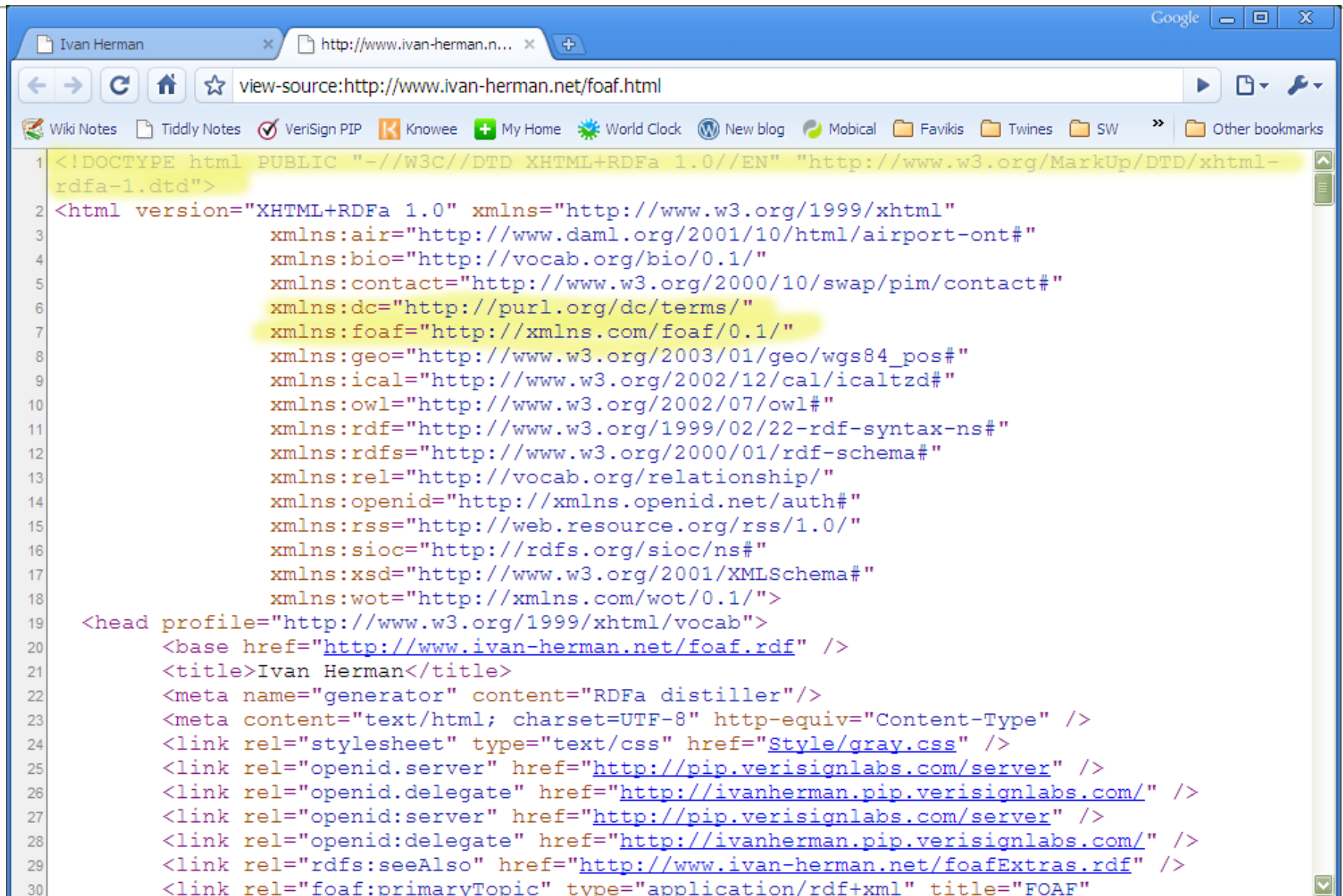
"social" links

Same example behind the scenes...



```
117         <!--
118         <span rel="owl:sameAs"
resource="http://community.linkeddata.org/dataspace/person/ivan#this">
119         <a href="http://community.linkeddata.org/">Linked Open
Data community</a></span>
120         -->
121     </div>
122 </div>
123
124 <div id="content" >
125     <h1><span property="foaf:title" class="forRDFOnly">Dr</span><span
property="foaf:name">Ivan Herman</span></h1>
126     <h2>Who am I?</h2>
127     <p>I graduated as mathematician at the <a rel="foaf:schoolHomepage"
href="http://www.elte.hu/"><span property="dc:title">Eötvös Loránd University of
Budapest</span></a>, Hungary, in 1979. After a brief scholarship at the Université Paris VI
I joined the Hungarian research institute in computer science (<a
href="http://www.sztaki.hu"><abbr title="Számítástechnikai és Automatizálási Kutatóintézet"
xml:lang="hu">SZTAKI</abbr></a>) 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 <a rel="foaf:workplaceHomepage" href="http://www.cwi.nl"><span
property="dc:title" datatype="">Centre for Mathematics and Computer Sciences (<abbr
title="Centrum voor Wiskunde en Informatica" xml:lang="nl">CWI</abbr></span></a> in
Amsterdam where I have a tenure position since 1988. I received a PhD degree in Computer
Science in 1990 at the <a href="http://www.leidenuniv.nl/">University of Leiden</a>, in the
Netherlands. I joined the <a rel="foaf:workplaceHomepage" href="http://www.w3.org"><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">W3C Offices</span></a> in January 2001 while maintaining my position at
<abbr title="Centrum Wiskunde en Informatica" xml:lang="nl">CWI</abbr>. I served as Head of
Offices until June 2006, when I was asked to take the <a rel="foaf:workInfoHomepage"
```

Same example behind the scenes...



```
1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-
rdfa-1.dtd">
2 <html version="XHTML+RDFa 1.0" xmlns="http://www.w3.org/1999/xhtml"
3     xmlns:air="http://www.daml.org/2001/10/html/airport-ont#"
4     xmlns:bio="http://vocab.org/bio/0.1/"
5     xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#"
6     xmlns:dc="http://purl.org/dc/terms/"
7     xmlns:foaf="http://xmlns.com/foaf/0.1/"
8     xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
9     xmlns:ical="http://www.w3.org/2002/12/cal/icaltzd#"
10    xmlns:owl="http://www.w3.org/2002/07/owl#"
11    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
12    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
13    xmlns:rel="http://vocab.org/relationship/"
14    xmlns:openid="http://xmlns.openid.net/auth#"
15    xmlns:rss="http://web.resource.org/rss/1.0/"
16    xmlns:sioc="http://rdfs.org/sioc/ns#"
17    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
18    xmlns:wot="http://xmlns.com/wot/0.1/">
19 <head profile="http://www.w3.org/1999/xhtml/vocab">
20 <base href="http://www.ivan-herman.net/foaf.rdf" />
21 <title>Ivan Herman</title>
22 <meta name="generator" content="RDFa distiller"/>
23 <meta content="text/html; charset=UTF-8" http-equiv="Content-Type" />
24 <link rel="stylesheet" type="text/css" href="Style/gray.css" />
25 <link rel="openid.server" href="http://pip.verisignlabs.com/server" />
26 <link rel="openid.delegate" href="http://ivanherman.pip.verisignlabs.com/" />
27 <link rel="openid:server" href="http://pip.verisignlabs.com/server" />
28 <link rel="openid:delegate" href="http://ivanherman.pip.verisignlabs.com/" />
29 <link rel="rdfs:seeAlso" href="http://www.ivan-herman.net/foafExtras.rdf" />
30 <link rel="foaf:primaryTopic" type="application/rdf+xml" title="FOAF"
```

In a slightly more readable format...

```
<html xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/terms/"
  ...
>
....

<div about="http://www.ivan-herman.net/me" ... >
  ...
  <p>I graduated as mathematician at the
    <a rel="foaf:schoolHomepage" href="http://www.elte.hu/">
      <span property="dc:title">Eötvös Loránd University of
        Budapest</span>
    </a>, ...
  ...
```

In a slightly more readable format...

```
<html xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/terms/"
  ...
>
....
<div about="http://www.ivan-herman.net/me" ... >
  ...
  <p>I graduated as mathematician at the
    <a rel="foaf:schoolHomepage" href="http://www.elte.hu/">
      <span property="dc:title">Eötvös Loránd University of
        Budapest</span>
    </a>, ...
  ...
```

Triple

Triple

... yielding

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dc:   <http://purl.org/dc/terms/>.

<http://www.ivan-herman.net/me>
  foaf:schoolHomepage <http://www.elte.hu/>.

<http://www.elte.hu/>
  dc:title "Eötvös Loránd University of Budapest".
```

Microformats or RDFa?

- ▶ There has been many unnecessary controversies
- ▶ For simple applications microformats are enough
 - GRDDL bridges them to the rest of the Semantic Web
- ▶ For more complex documents RDFa is great
- ▶ It often boils down to matters of taste...

Bridge to relational databases: R2RML

- ▶ Data on the Web are mostly stored in databases
- ▶ “Bridges” are being defined:
 - a layer between RDF and the relational data
 - RDB tables are “mapped” to RDF graphs, possibly on the fly
 - different mapping approaches are being used
 - a number RDB systems offer this facility already (eg, Oracle, OpenLink, ...)
- ▶ R2RML is W3C’s evolving standard in this area

How to “assign” RDF data to resources?

- ▶ This is important when the RDF data is used as “metadata”
- ▶ Some examples:
 - copyright information for your photographs
 - is a Web page usable on a mobile phone and how?
 - bibliographical data for a publication
 - annotation of the data resulting from a scientific experiment
 - etc

If I know the URI of the resource (photograph, publication, etc), how do I find the relevant RDF data?

The data might be embedded

- ▶ Some data formats allow the direct inclusion of (RDF) metadata:
 - SVG (Scalable Vector Graphics)
 - direct inclusion of RDF/XML
 - via RDFa attributes
 - XHTML with RDFa or microformats+GRDDL
 - JPG files using the comment area and, eg, Adobe's XMP technology

Simple linkage

- ▶ Some formats have special link statements. Eg, in (X)HTML:

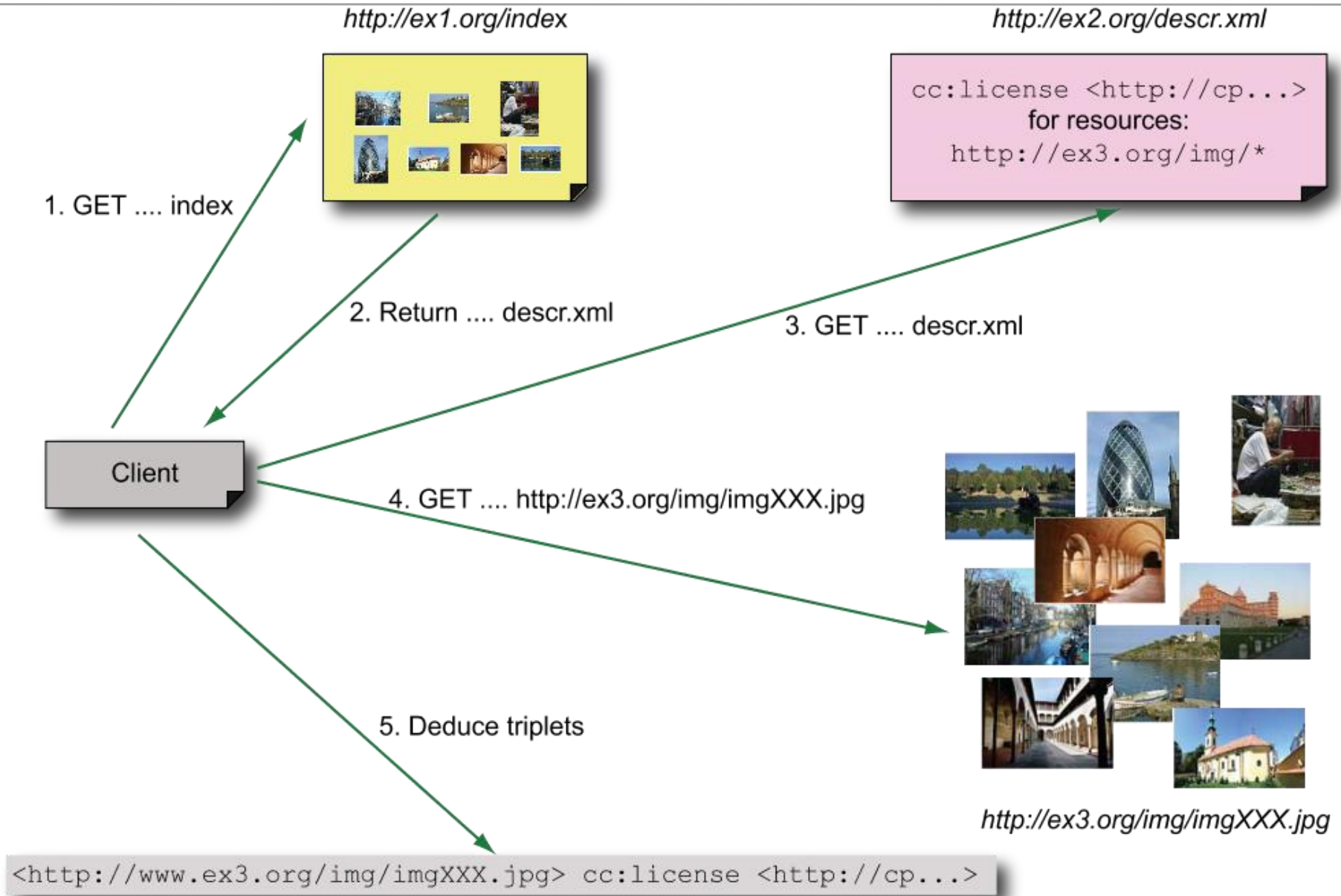
```
<html>
  <head>
    <link rel="meta" href="meta.rdf"/>
  ...
```

- ▶ Similar facilities might exist in other formats (eg, SMIL)

POWDER

- ▶ POWDER provides for a more elaborate scenarios:
 - define a set of resources by constraints on the URIs; eg
 - URIs must begin with `http://www.example.com/bla/`
 - the port number in the URI-s should be XYZW
 - define “description resources” that bind those resources to additional information
 - get such description resources, eg, via a link statements, via HTTP, via SPARQL, ...
- ▶ Use cases: licensing information, mobileOK (and other) trustmarks, finding accessible Web sites, content labeling (eg, child protection), ...

A POWDER scenario: copyright for photos



The gory details...

- ▶ The “description resource” is an XML file:

```
<powder xmlns="http://www.w3.org/2007/05/powder#"
        xmlns:cc="http://creativecommons.org/ns#">
  <attribution>
    <issuedby src="http://www.ivan-herman.net/me"/>
  </attribution>
  <dr>
    <iriset>
      <includehosts>www.ex2.org</includehost>
      <includepathstartswith>/img/</includepathstartswith>
    </iriset>
    <descriptorset>
      <cc:license rdf:resource="http://..." />
    </descriptorset>
  </dr>
```


The gory details...

- ▶ Powder processors may then return
 - direct RDF triples, eg:

```
<http://www.ex2.org/img/imgXXX.jpg> cc:license <http://...>.
```

- ▶ or can transform this XML file into an RDF (OWL) for more generic processors
 - ▶ a canonical processing of the XML file is defined by the POWDER specification

POWDER Service

- ▶ Online POWDER service can be set up:
 - a Web service with
 - submit a URI and a resource description file
 - return the RDF statements for that URI
 - such service should be set up, eg, at W3C
- ▶ A GRDDL transform is also defined

But there is a hidden “hiccup”

- ▶ RDF makes a strong separation between
 - URI as an ID for a resource
 - URI as a datatype (xsd:anyURI)
 - there is no “bridge” between the two
- ▶ POWDER includes a small extension to the formal semantics of RDF for two properties:
 - wdrs:matchregex and wdrs:notmatchregex such that
 - (R wdrs:matchregex Regex) holds iff the URI of R matches Regex

If you want the OWL version...

```
<> wdrs:issuedBy <http://www.ivan-herman.net/me> .

_:iriset_1 a owl:Class; owl:intersectionOf (
  [ a owl:Restriction;
    owl:onProperty wsd:matchregex ;
    owl:hasValue "..ugly regex for ex2.org"^^xsd:string ]
  [ a owl:Restriction;
    owl:onProperty wsd:matchregex ;
    owl:hasValue "..ugly regex for /img"^^xsd:string ]
) .

_:desc_1 a owl:Restriction;
  owl:onProperty cc:license;
  owl:hasValue <http://...>.

_:iriset_1 rdfs:subClassOf _:desc_1 .
```

Consequences of the “hiccup”

- ▶ In practice this means that only “POWDER aware” agents can fully handle the description files
 - note that the extension is fairly easy to add, so it is not a big implementation issue...
- ▶ Existence of the services to provide the triplets automatically relieve the pain...

Other POWDER features

- ▶ There are a number of additional features:
 - built in authentication: description resources must be attributed and this is open for authentication
 - assignments may carry validity dates
 - packaging several resource descriptions in one, possibly control their processing order
 - using tags to identify resources instead of URI patterns

Linking Open Data

Linking Open Data Project

- ▶ Goal: “expose” open datasets in RDF
- ▶ Set RDF links among the data items from different datasets
- ▶ Set up, if possible, query endpoints

Example data source: DBpedia

- ▶ DBpedia is a community effort to
 - extract structured (“infobox”) information from Wikipedia
 - provide a query endpoint to the dataset
 - interlink the DBpedia dataset with other datasets on the Web



UNIVERSITÄT LEIPZIG



Extracting structured data from Wikipedia

```
@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm  <http://dbpedia.org/property/>.
```

```
dbpedia:Amsterdam
```

```
dbterm:officialName "Amsterdam" ;
dbterm:longd "4" ;
dbterm:longm "53" ;
dbterm:longs "32" ;
dbterm:website <http://www.amsterdam.nl> ;
dbterm:populationUrban "1364422" ;
dbterm:areaTotalKm "219" ;
```

...

```
dbpedia:ABN_AMRO
```


```
dbterm:location dbpedia:Amsterdam ;
```

...

Amsterdam	
— Municipality / City —	
	
Coordinates:	52°22′23″N 4°53′32″E﻿ / ﻿52.37306°N 4.89222°E﻿ / 52.37306; 4.89222
Country	Netherlands
Province	North Holland
COROP	Amsterdam
Boroughs	Boroughs
Government	
 - Mayor	Eberhard van der Laan (PvdA)
 - Aldermen	Carolien Gijkiels Hans Gerson Maarten van Poelgeest Freek Ossel Manjke Vos
 - Secretary	Henk de Jong
Area ^{[1][2]}	
 - Municipality / City	219 km ² (84.6 sq mi)
 - Land	166 km ² (64.1 sq mi)
 - Water	53 km ² (20.5 sq mi)
 - Urban	1,003 km ² (387.3 sq mi)
 - Metro	1,815 km ² (700.8 sq mi)
Elevation ^[3]	2 m (7 ft)
Population (June 2009) ^{[4][5]}	
 - Municipality / City	762,057
 - Density	4,459/km ² (11,548.8/sq mi)
 - Urban	1,364,422
 - Metro	2,158,372
 - Demonym	Amsterdammer
Time zone	CET (UTC+01)
 - Summer (DST)	CEST (UTC+02) (UTC)
Postal codes	1011–1109
Area code(s)	020
Website	www.amsterdam.nl

Automatic links among open datasets

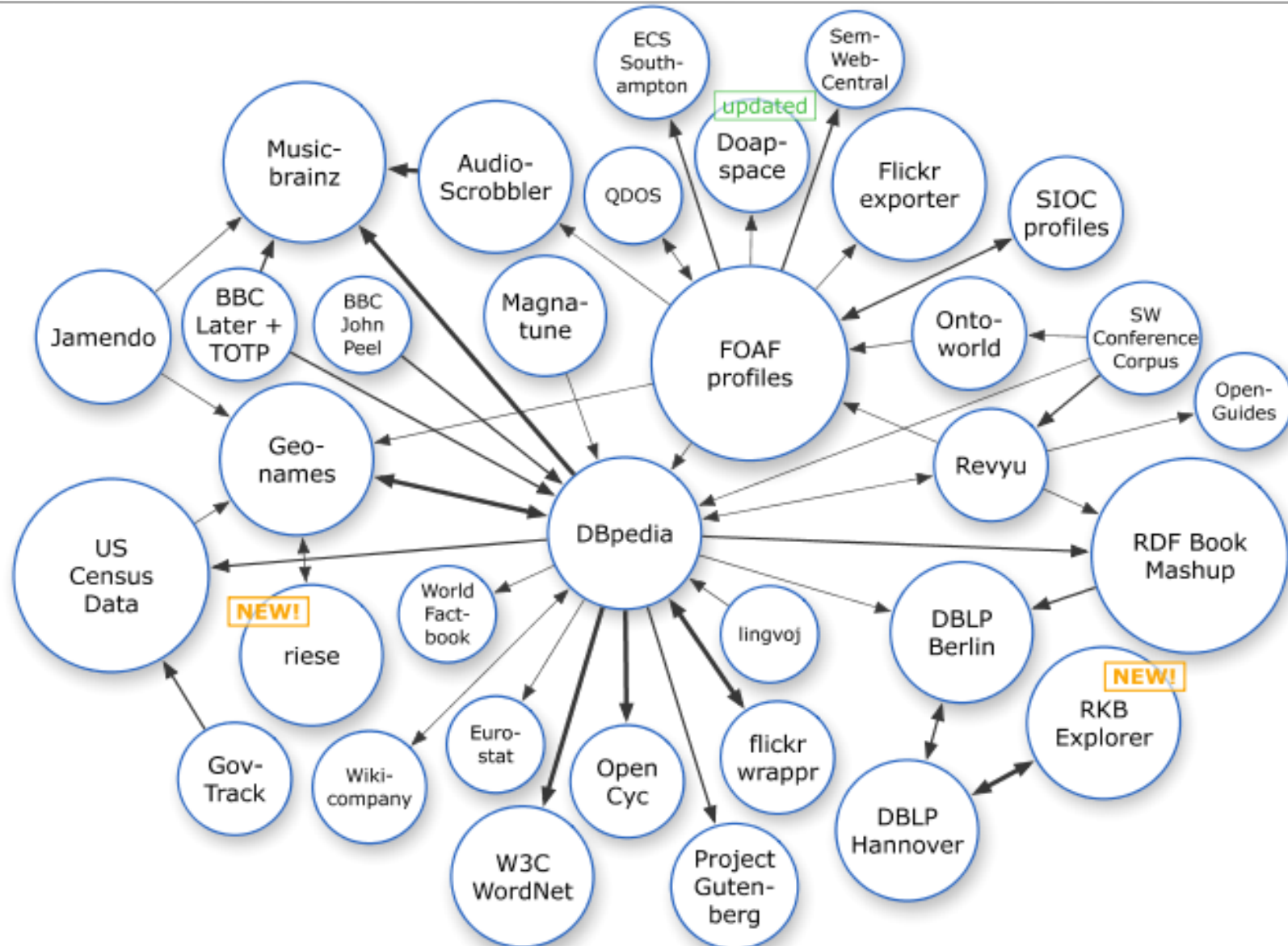
```
<http://dbpedia.org/resource/Amsterdam>  
  owl:sameAs <http://rdf.freebase.com/ns/...> ;  
  owl:sameAs <http://sws.geonames.org/2759793> ;  
  ...
```



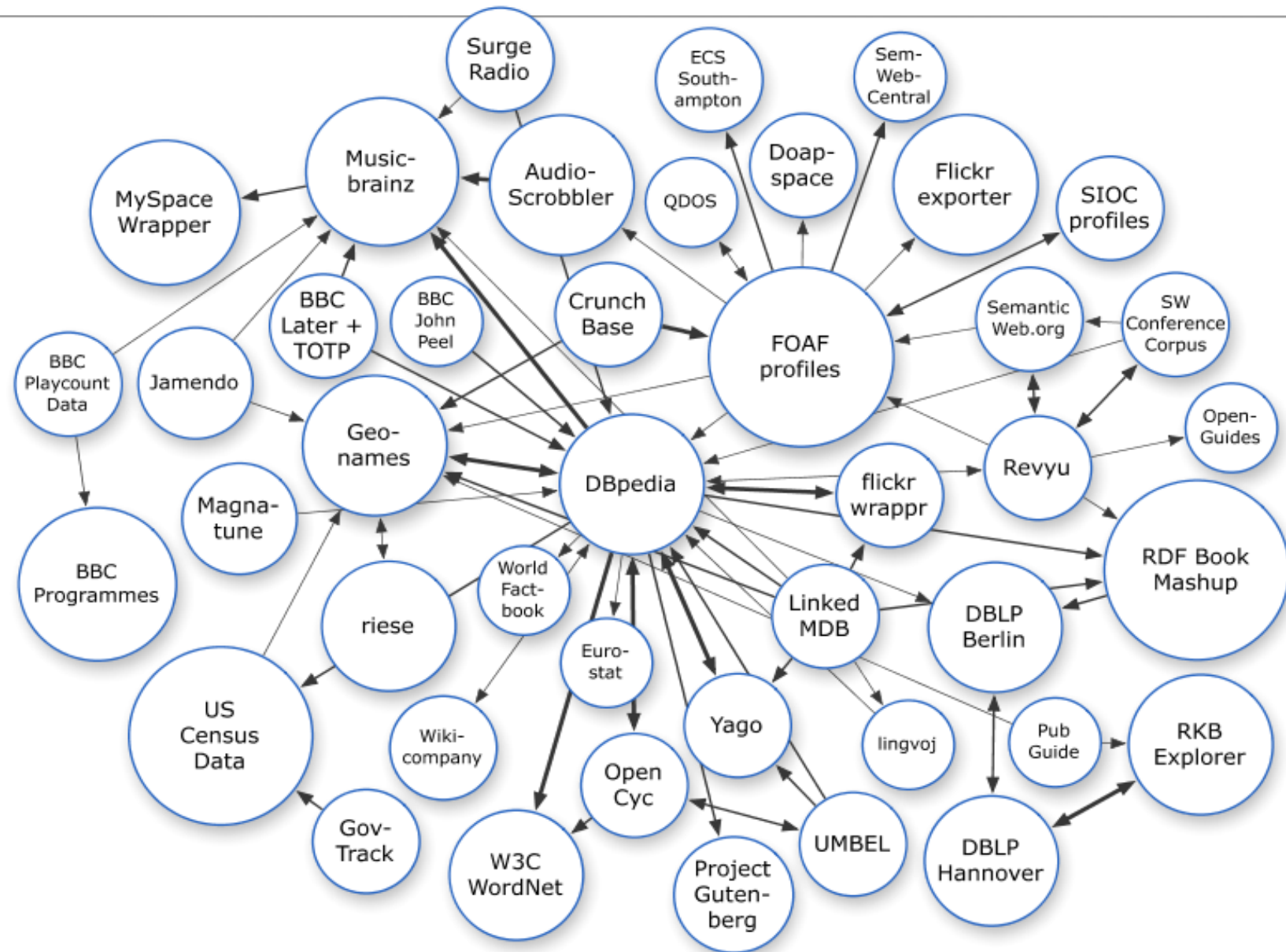
```
<http://sws.geonames.org/2759793>  
  owl:sameAs <http://dbpedia.org/resource/Amsterdam>  
  wgs84_pos:lat "52.3666667" ;  
  wgs84_pos:long "4.8833333" ;  
  geo:inCountry <http://www.geonames.org/countries/#NL> ;  
  ...
```

Processors can switch automatically from one to the other...

The LOD "cloud", March 2008

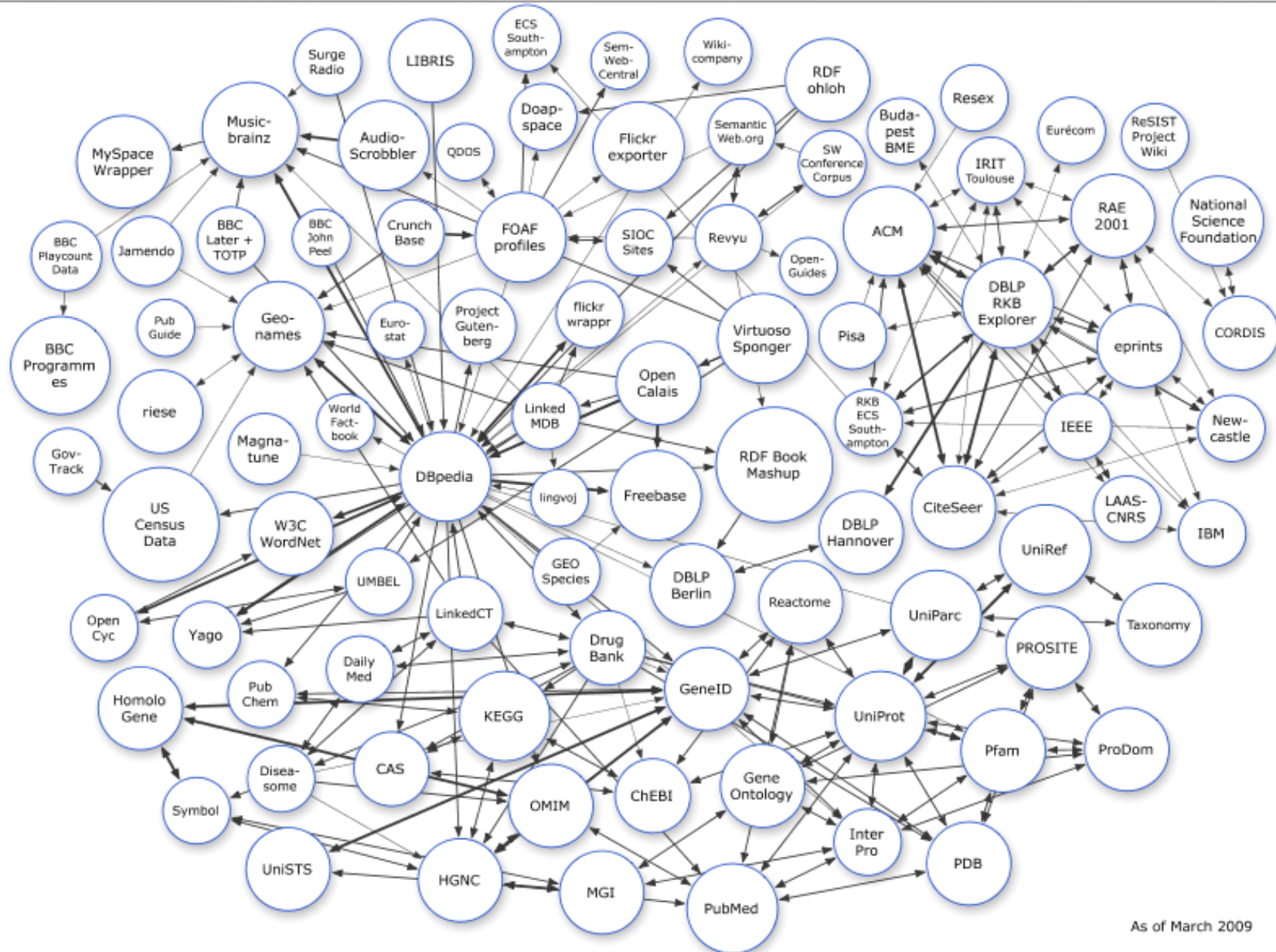


The LOD “cloud”, September 2008



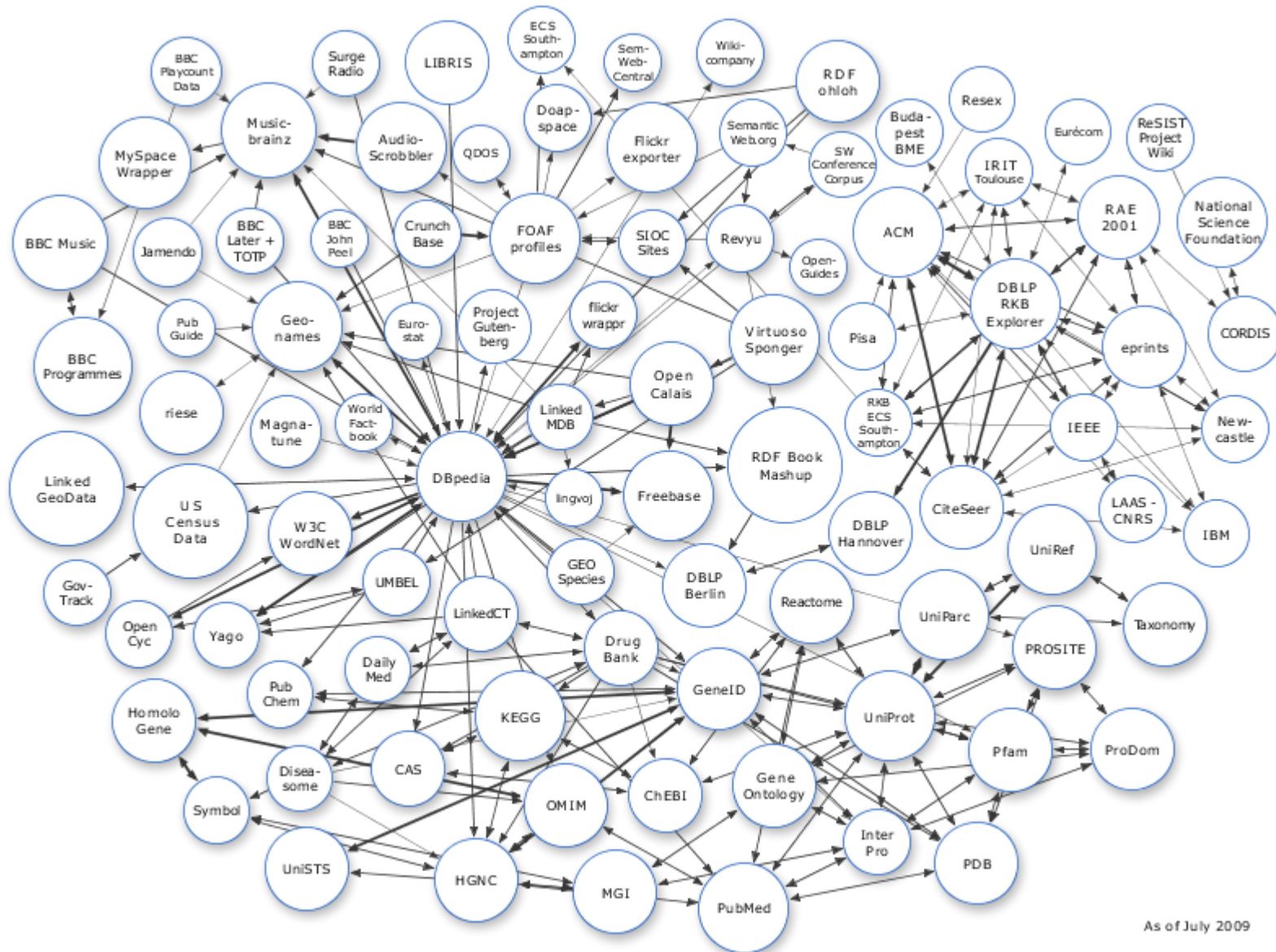
As of September 2008

The LOD "cloud", March 2009



As of March 2009

The LOD "cloud", June 2009

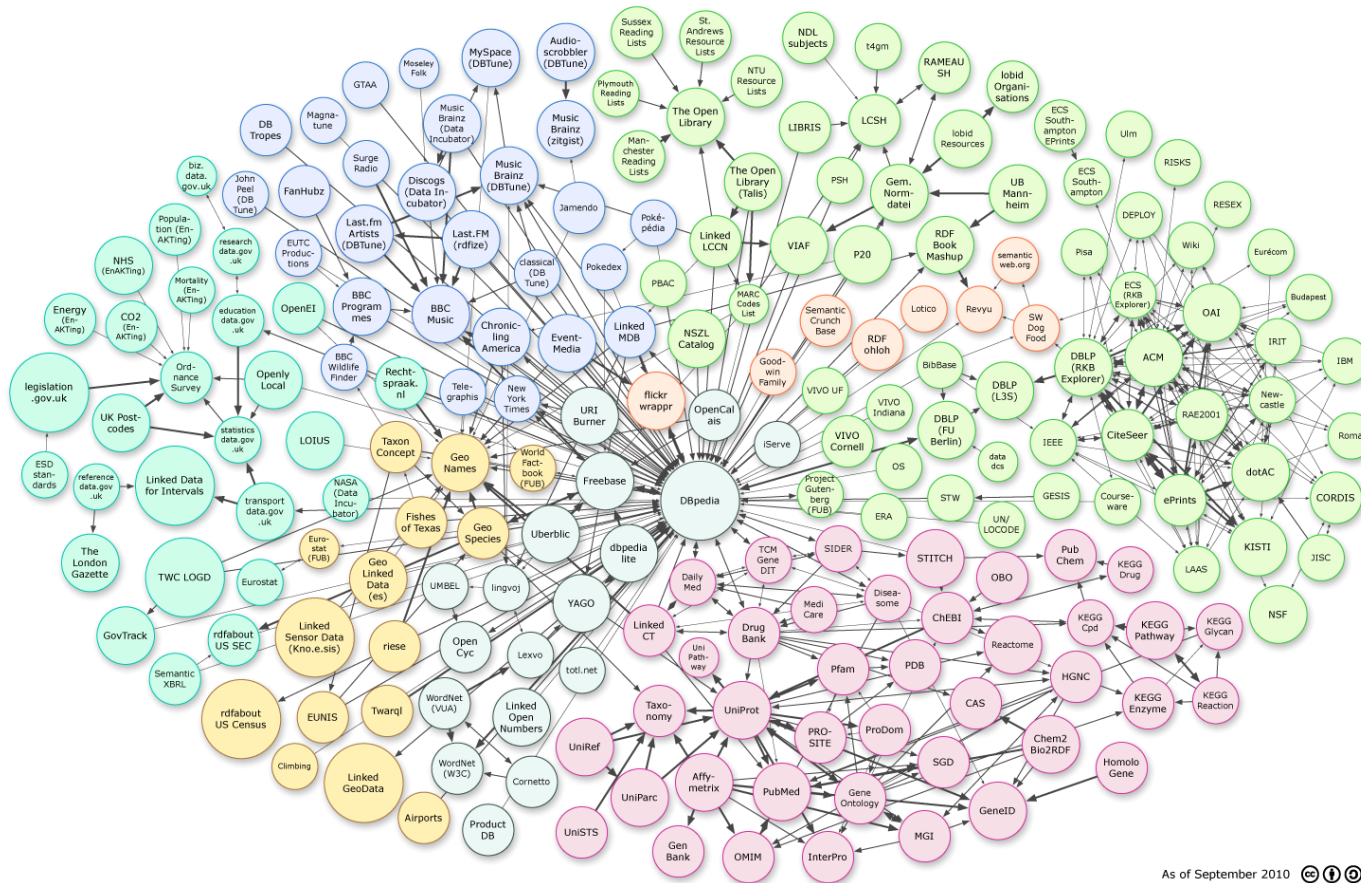


As of July 2009

Application specific portions of the cloud

► Eg, “bio” related datasets

- done, partially, by the “Linking Open Drug Data” task force of the HCLS IG at W3C



As of September 2010 © ⓘ ⓘ

Three technologies have emerged

- ▶ To re-use thesauri, glossaries, etc: **SKOS**
- ▶ To define more complex vocabularies with a strong logical underpinning: **OWL**
- ▶ Generic framework to define rules on terms and data: **RIF**

SKOS

- ▶ Represent and share classifications, glossaries, thesauri, etc
 - for example:
 - Dewey Decimal Classification, Art and Architecture Thesaurus, ACM classification of keywords and terms...
 - classification/formalization of Web 2.0 type tags
- ▶ Define classes and properties to add those structures to an RDF universe
 - allow for a quick port of this traditional data, combine it with other data

Example: the term “Fiction”, as defined by the Library of Congress

The screenshot shows a web browser window displaying the Library of Congress Authorities & Vocabularies page for the term "Fiction". The browser's address bar shows the URL <http://id.loc.gov/authorities/sh85048050>. The page header includes the Library of Congress logo and navigation buttons for "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS". The breadcrumb trail reads "The Library of Congress > Authorities & Vocabularies > Fiction".

The main content area is titled "Authorities & Vocabularies" and includes a "Return" link. Below this is a search box with the placeholder text "Enter search terms..." and a "GO" button. Two tabs, "Details" and "Visualize", are visible, with "Visualize" currently selected.

The "Fiction" entry is displayed with the following information:

- URI:** `<http://id.loc.gov/authorities/sh85048050#concept>`
- Type:** Topical Term
- Alternate Labels:** Fiction--Philosophy; Metafiction; Novellas (Short novels); Novels; Stories
- Broader Terms:**
 - [Literature](#)
 - [Prose literature](#)
- Narrower Terms:**
 - [Adventure stories](#)
 - [Allegories](#)
 - [Alternative histories \(Fiction\)](#)
 - [Bildungsromans](#)
 - [Biographical fiction](#)

Example: the term “Fiction”, as defined by the Library of Congress

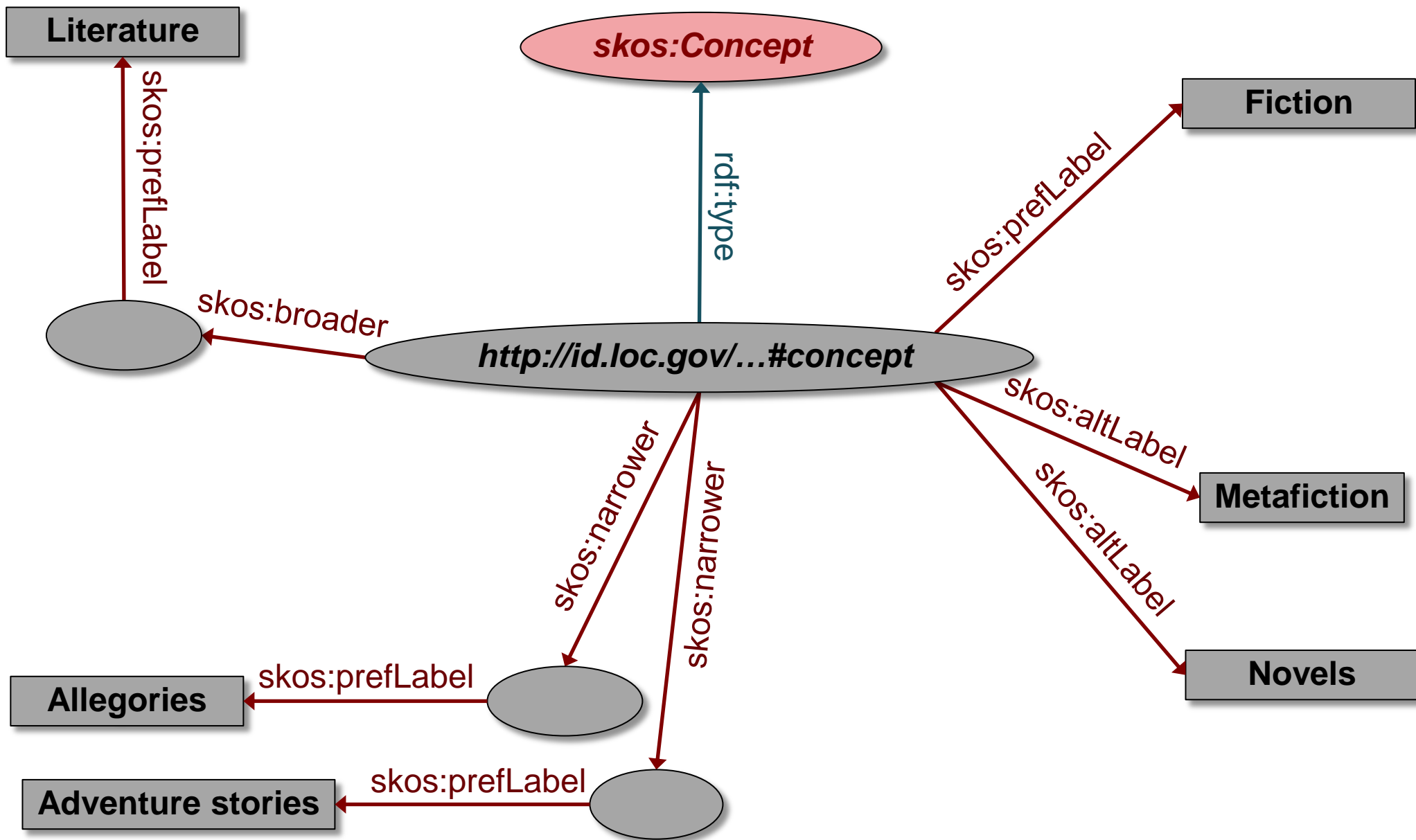
The screenshot shows a web browser window displaying the Library of Congress Authorities & Vocabularies page for the term "Fiction". The browser's address bar shows the URL <http://id.loc.gov/authorities/sh85048050>. The page features a navigation bar with the Library of Congress logo and buttons for "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS". Below the navigation bar, the breadcrumb trail reads "The Library of Congress > Authorities & Vocabularies > Fiction". The main heading is "Authorities & Vocabularies". A "Return" button is visible. A search box contains the text "Enter search terms..." and a "GO" button. Below the search box, there are two tabs: "Details" (selected) and "Visualize". The "Details" tab displays the following information:

- Fiction**
- URI:** <http://id.loc.gov/authorities/sh85048050#concept>
- Type:** Topical Term
- Alternate Labels:** Fiction--Philosophy; Metafiction; Novellas (Short novels); Novels; Stories
- Broader Terms:**
 - Literature
 - Prose literature
- Narrower Terms:**
 - Adventure stories
 - Allegories
 - Alternative histories (Fiction)
 - Bildungsromans
 - Biographical fiction

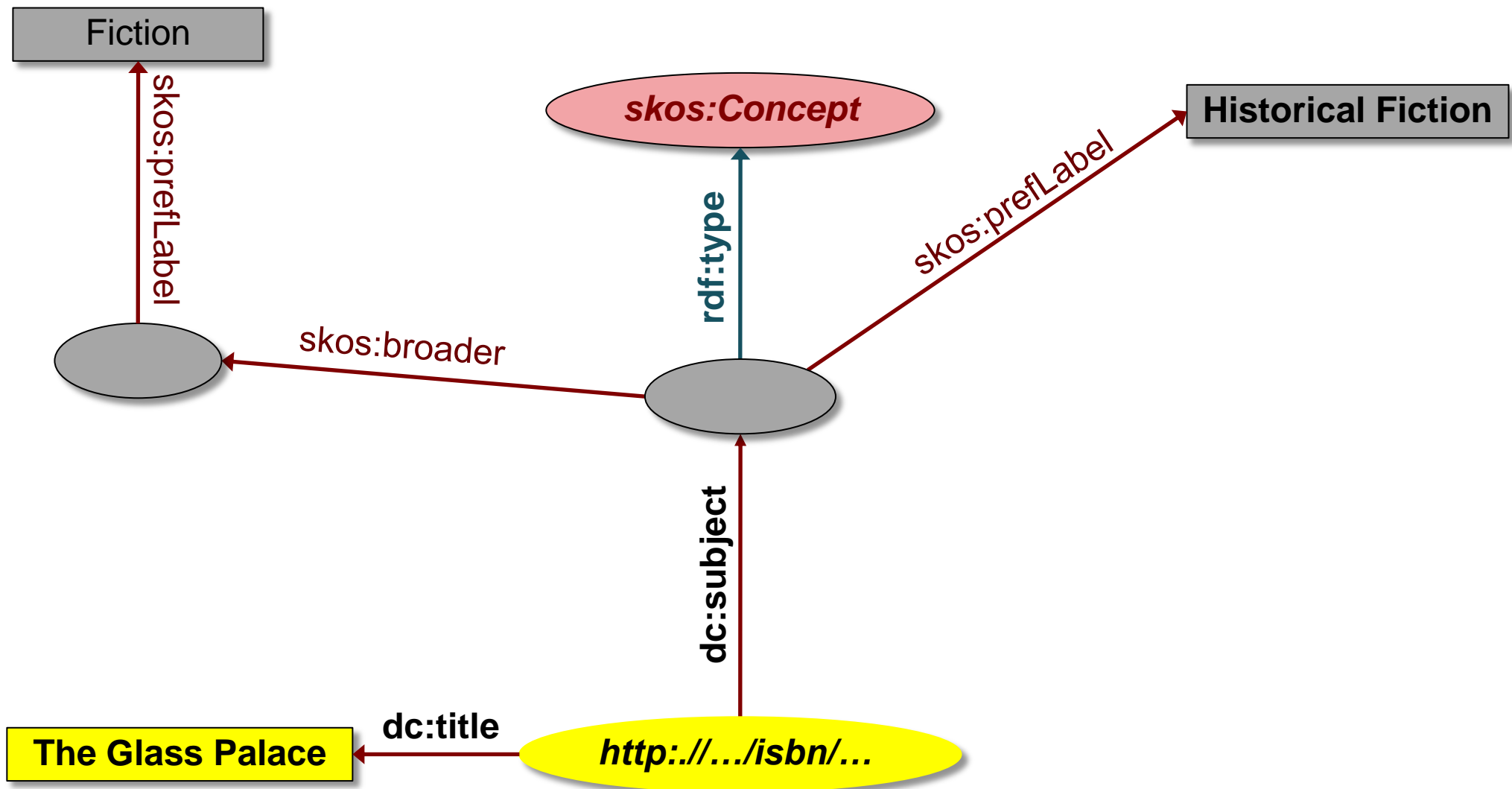
Thesauri have identical structures...

- ▶ The structure of the LOC page is fairly typical
 - label, alternate label, narrower, broader, ...
 - there is even an ISO standard for these
- ▶ SKOS provides a basic structure to create an RDF representation of these

LOC's "Fiction" in SKOS/RDF



Usage of the LOC graph



Same serialized

```
<http://.../isbn/000651409X>
  dc:title "The Glass Palace"@en;
  dc:subject <http://id.loc.gov/authorities/sh85061165#concept>;
  ...

<http://id.loc.gov/authorities/sh85061165#concept>
  a      skos:Concept;
  skos:prefLabel "Historical Fiction"@en;
  skos:broader <http://id.loc.gov/authorities/sh85048050#concept>;
  ...

<http://id.loc.gov/authorities/sh85048050#concept>
  a      skos:Concept;
  skos:prefLabel "Fiction"@en;
  skos:narrower <http://id.loc.gov/authorities/sh85061165#concept>;
  ...
```

SKOS terms overview

- ▶ Classes and Properties:
 - Basic description (Concept, ConceptScheme,...)
 - Labeling (prefLabel, altLabel,...)
 - Documentation (definition, historyNote,...)
 - Semantic relations (broader, narrower, related,...)
 - Collections (Collection, OrderedCollection,...)
 - Concept mappings (broadMatch, narrowMatch,...)

Importance of SKOS

- ▶ SKOS provides a simple bridge between the “print world” and the (Semantic) Web
- ▶ Thesauri, glossaries, etc, from the library community can be made available
 - LOC is a good example
- ▶ SKOS can also be used to organize, eg, tags, annotate other vocabularies, ...

Importance of SKOS

- ▶ Anybody in the World can refer to common concepts
 - they mean the same for everybody
- ▶ Applications may exploit the relationships among concepts
 - eg, SPARQL queries may be issued on the library data+LOC