RDFa

WARNING TO THE READER!

- ▶ This is an evolving slide set. This means:
 - it changes frequently
 - there may be bugs, inconsistencies
 - it may try to reflect the latest view of technology evolution but that is often a moving target
- "Frozen" versions are instantiated for a specific presentation, and those become stable



What is RDFa?

- ▶ For RDF people, it *sounds* very simple:
 - RDFa is a serialization of RDF embedded in XHTML, HTML, or XML in general



So why bother? Why should we care? Why is that of any importance?



RDFa may become the single biggest source of RDF triples on the Web after direct database access!

(well, this is a bit cheating, because the RDFa data may come from a database, too...)



Data for a Web of Data

- Apart from relational databases, most of the data on the Web are in... (X)HTML content
- New content is generated every day
- How would one get structured data from that information?



Authors of the "traditional Web"...

- ▶ Do not generate RDF/XML files separately
 - RDF/XML is complex
 - it requires a separate storage, generation, etc mechanism
 - that is also valid for, e.g., Turtle
 - even when authoring with, say, Emacs, creating an extra file is a load



Solution

- Add extra structured content to the (X)HTML pages
- Let processors extract those and turn into RDF



Existing approaches

Microformats

- reuses HTML attributes like @class, @title
- separate vocabularies (address, CV, ...)
- difficult to mix microformats (no concept of namespaces)
- possible to transform via, e.g., XSLT + GRDDL, but all transformations are vocabulary dependent



Existing approaches

Microdata

- adds new attributes to HTML5 to express metadata
- can use URI-s, it also fixes some vocabulary mappings (e.g., to Dublin Core elements)
- works for simpler usages, not well suited for complex vocabularies
- has no notion of datatypes, namespaces
- but generic processing becomes possible to generate RDF



Existing approaches

▶ RDFa

- adds new (X)HTML/XML attributes
- has namespaces and URIs at its core; i.e., mixing vocabulary is just as easy as in RDF
- complete flexibility for using Literals or URI Resources
- is a complete serialization of RDF
- generic processing becomes possible to generate RDF



RDFa is a complete bridge between the Web of Documents and the Web of Data



Therefore...

- ▶ It is very important for RDF experts to
 - know RDFa
 - parse it alongside Turtle, RDF/XML or other
 - when appropriate, generate RDFa pages



What does RDFa look like?



Main principles of RDFa

- ▶ RDFa means "RDF in attributes". I.e.:
 - all RDF contents are defined through XML attributes (no elements)
 - the XML/HTML tree structure is used
 - many of the attributes are defined by RDFa
 - some attributes (@href, @rel) are also reused
 - if possible, the text content is also reused (for literals) as well as @href values



What does this mean in practice?

- ▶ The same (X)HTML file:
 - is used, unchanged, by browsers
 - they ignore attributes they do not know
 - can be used by specialized processors (or APIs) to extract RDF triples

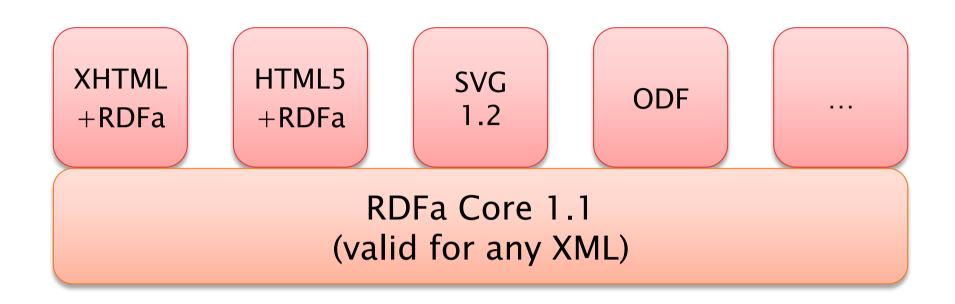


Before getting into details...

- ▶ The current Recommendation is RDFa 1.0
- ▶ There is an RDFa1.1 in the making, almost ready
- ▶ I will talk about RDFa1.1 and warn when the feature is not available in RDFa1.0



XML or (X)HTML?



- Formally:
 - RDFa WG defines Core and XHTML
 - HTML WG defines HTML5
- this tutorial uses XHTML examples



How does it work?







Unique identifier for RDFS Entailment.

"http://www.w3.org/ns/entailment/RDFS" is the URI. The <u>specification for the RDFS entailment</u> is part of the <u>RDF Semantics</u> W3C Recommendation.

For more information about RDF, please refer to the htm

Ivan Herman, ivan@w3.org, W3C, Semantic Web Activity Lead, 2009-05-03



```
        Unique identifier for <em>RDFS Entailment</em>.
```



```
about="http://www.w3.org/ns/entailment/RDFS"

property="http://purl.org/dc/terms/description">

Unique identifier for <em>RDFS Entailment</em>.
```

```
<http://www.w3.org/ns/entailment/RDFS>
... .
```



```
        Unique identifier for <em>RDFS Entailment</em>.
```



```
        Unique identifier for <em>RDFS Entailment</em>.
```

```
<http://www.w3.org/ns/entailment/RDFS>
     <http://purl.org/dc/terms/description>
          "Unique identifier for RDFS Entailment." .
```







Unique identifier for RDFS Entailment.

"http://www.w3.org/ns/entailment/RDFS" is the URI. The <u>specification for the RDFS entailment</u> is part of the <u>RDF Semantics</u> W3C Recommendation.

For more information about RDF, please refer to the <u>the RDF Concepts and Abstract Syntax</u> Recommendation .

Ivan Herman, ivan@w3.org, W3C, Semantic Web Activity Lead, 2009-05-03



```
<a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
        RDF Semantics.
</a>
```



```
a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
        RDF Semantics.
</a>
```

```
<http://www.w3.org/ns/entailment/RDFS>
....
```



```
<a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
     RDF Semantics.
</a>
```



```
<a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
        RDF Semantics.
</a>
```

```
<http://www.w3.org/ns/entailment/RDFS>
    <http://www.w3.org/2000/01/rdf-schema#seeAlso>
         <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> .
```



Is that it?



What we have is... Ntriples in HTML

- The combination of @about with @rel/@property and possibly @href covers most of we need...
- But this is too complex for authors



Just compare

with

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/terms/> .

<http://www.w3.org/ns/entailment/RDFS>
    rdfs:seeAlso <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> ;
    dc:description "Unique identifier for RDFS Entailment." .
```



The "Turtle" aspects of RDFa

- Use compact URI-s when possible
- Make use of XML structure for
 - shared subjects
 - shared predicates
 - create blank nodes
 - **=** ...



Compact URIs ("CURIE"s)

- Just like in Turtle:
 - define a prefix via @prefix
 - use prefix:reference to abbreviate a URI



CURIE definition and usage

can be replaced by:

Some details on @prefix

- Can be anywhere in the XML tree and is valid for the whole sub-tree
 - i.e., the html element is not the only place to have it
- The same @prefix attribute can hold several definitions:
 - prefix="dc: http://purl.org... rdfs: http://..."



Some details on @prefix

- An alternative (deprecated) syntax is
 - xmlns:dc="http://purl.org/dc/terms/"
- CURIEs and "real" URIs can be mixed
 - if an attribute value can be interpreted as a CURIE, fine
 - alternatively, it is considered as a URI
- CURIEs cannot be used on @href



RDFa 1.0 Warnings on CURIEs

- In RDFa 1.0
 - only the xmlns:XXX syntax is usable
 - CURIEs on @about can only be used with the syntax: about="[pref:ref]"
 - Only CURIEs can be used on, e.g., @property or @rel (no fallback on URIs)



Sharing subjects

- The basic principle: @about is inherited by children nodes
 - i.e., no reason to repeat it



Shared subject example



... yielding

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/terms/> .

<http://www.w3.org/ns/entailment/RDFS>
    rdfs:seeAlso
        <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> ;
    dc:description "Unique identifier for RDFS Entailment." .
```



Intricacies of literals



Date examples

This leads to:

```
@prefix dc: <http://...>
<...> dc:date "2010-07-05" .
```



Datatypes

This leads to:

```
@prefix dc: <http://...>
@prefix xsd: <http://...>
<...> dc:date "2010-07-05"^^xsd:date .
```



Controlling the literal

- The basic rule says: the (RDF) Literal is the enclosed text from the HTML content
- ▶ This is fine in 80% of the cases, but...
- ▶ It may not be natural in all cases! E.g.,
 - 2010-07-05 is the "official" ISO format (for xsd:date)
 - but "July 5, 2010" looks much more natural for a human...



Usage of @content

Also leads to:

```
@prefix dc: <http://...>
@prefix xsd: <http://...>
<...> dc:date "2010-07-05"^^xsd:date .
```



Subjects, and objects, and subjects again...



The rules until now

- What we said is:
 - @about sets the subject
 - @href sets the object
- But that is not always good enough
 - we do not always want active links (i.e., the "a" element)
 - what about other links in HTML?



@src also sets the object

yields:

```
<...> foaf:depiction <http://www.ex.org/img.png> .
```



We may not always want links...

- The RDFa @resource attribute is equivalent to @href
 - it sets the object, just like @href
 - but it is ignored by a browser, i.e., not a link!
 - e.g.,:



"Chaining"

Here is what we would like to have in RDFa

```
<http://www.w3.org/ns/entailment/RDFS>
    dc:creator <http://www.ivan-herman.net/foaf#me> .

<http://www.ivan-herman.net/foaf#me>
    foaf:mailbox <mailto:ivan@w3.org> ;
    foaf:workplaceHomepage <http://www.w3.org> .
```



"Chaining"

A straightforward way:

```
<body about="http://www.w3.org/ns/entailment/RDFS">
  <address>
    <span rel="dc:creator"</pre>
       resource="http://www.ivan-herman.net/foaf#me"/>
    <span about="http://www.ivan-herman.net/foaf#me">
      <a rel="foaf:mailbox"</pre>
          href="mailto:ivan@w3.org">ivan@w3.org</a>,
      <a rel="foaf:workplaceHomepage"</pre>
          href="http://www.w3.org">W3C</a>
    </span>
  </address>
```



"Chaining"

A straightforward way:

```
<body about="http://www.w3.org/ns/entailment/RDFS">
  <address>
    <span rel="dc:creator"</pre>
       resource="http://www.ivan-herman.net/foaf#me"/>
    <span about="http://www.ivan-herman.net/foaf#me">
      <a rel="foaf:mailbox"</pre>
          href="mailto:ivan@w3.org">ivan@w3.org</a>,
      <a rel="foaf:workplaceHomepage"</pre>
          href="http://www.w3.org">W3C</a>
    </span>
  </address>
```



"Chaining": when objects become subjects...

An alternative:

```
<body about="http://www.w3.org/ns/entailment/RDFS">
  <address>
    <span rel="dc:creator"</pre>
       resource="http://www.ivan-herman.net/foaf#me">
         <a rel="foaf:mailbox"</pre>
              href="mailto:ivan@w3.org">ivan@w3.org</a>,
         <a rel="foaf:workplaceHomepage"</pre>
              href="http://www.w3.org">W3C</a>
    </span>
  </address>
```



Chaining means

- Peresource (or @href) becomes a subject for the subtree
- ▶ This feature is a bit like in RDF/XML



Some extra features



Some extra features

- ▶ Blank nodes can be created using "_:XX"
- Shorthand for RDF types
- An API has been defined for Web Applications



Typing

- Typing can of course be done using @rel="rdf:type"
- ▶ But that is a widely used combination, so there is a separate @typeof attribute for that



Typing example

yields

```
<http://www.ivan-herman.net/foaf#me> a foaf:Person ;
  foaf:name "Ivan Herman" .
```



Profile files

- Prefix declarations can be collected in a separate file and referred to via a @profile attribute
 - the "profile file"
 - RDFa1.0 warning: this is an RDFa1.1 feature!



"Term" declarations

- ▶ A profile file can also define a *term*:
 - assign a URI to a simple string
- The term can be used directly by authors, without prefixes



Term Example

- ▶ Say, file "http://ex.org/prof" defines a mapping:
 - "desc" → "http://purl.org/dc/terms/description"



Term Example

yields

```
@prefix dc: <http://purl.org/dc/terms/> .

<...> <http://purl.org/dc/terms/description>
    "Unique identifier for RDFS Entailment."
```



Terms are important...

- Usage of CURIEs and URIs is intuitive for RDF people...
- ▶ It is not for average HTML authors!
- Profile files can be published by major publishers:
 - Dublin Core, FOAF, ...
 - FaceBook, Google, ...
- ... and users can simply refer to the profiles

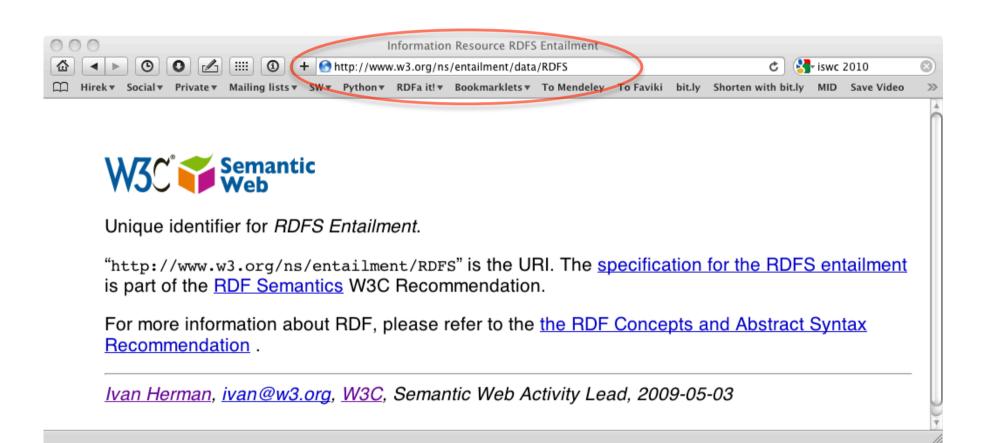


Publishing RDFa



A typical deployment pattern

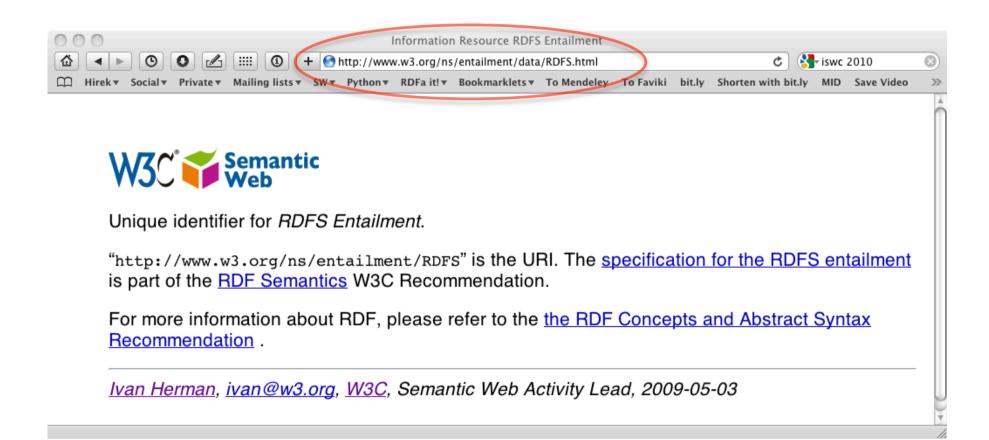
A browser usually asks for an HTML content:





A typical deployment pattern

Via content negotiations this goes to:





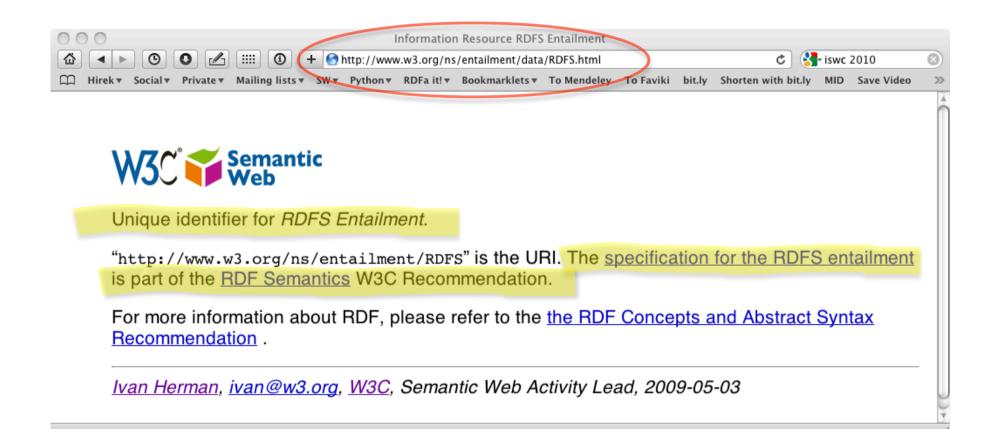
A typical deployment pattern

▶ But a client could ask for, say, Turtle:

```
000
                                       http://www.w3.org/ns/entailment/data/RDFS.ttl
                  (+ shttp://www.w3.org/ns/entailment/data/RDFS.ttl
                                                                                              iswc 2010
☐ Hirek ▼ Social ▼ Private ▼ Mailing lists ▼ SW ▼ Python ▼ RDFa it! ▼ Bookmarklets ▼ To Mendeley To Faviki bit.ly Shorten with bit.ly
EDICITY VIIV. VIICED.//WWW.WJ.OIG/1777/AHCHIL/VOCAD#/
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
ent:RDFS a ent:Entailment ;
      dc:creator <http://www.ivan-herman.net/foaf#me> ;
      dc:date "2009-05-03";
      dc:description "Unique identifier for RDFS Entailment";
      rdfs:comment "The specification for the RDFS entailment is part of the RDF Semantics
W3C Recommendation.";
      rdfs:isDefinedBy <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#rdfs entailment>;
      rdfs:seeAlso <a href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">http://www.w3.org/TR/2004/REC-rdf-mt-20040210/</a>.
<http://www.w3.org/ns/entailment/data/RDFS.html> dc:title "Information Resource RDFS
Entailment";
      xhv:stylesheet <a href="http://www.w3.org/StyleSheets/TR/base">http://www.w3.org/StyleSheets/TR/base</a>.
```



...corresponding to:





Where does the Turtle content come from?

- ▶ The triples are embedded in the HTML file
 - a client may know how to extract RDF triples directly from that file; or
 - an online "distiller" service is used; or
 - the server is set up to generate the Turtle file automatically and let content negotiation work (previous examples)



For example, use an online service...

```
http://www.w3.org/2007/08/pyRdfa/extract?format=turtle&uri=http%3A//www.w3.org/ns/entailment/data/RDFS.html
 \( \brace \) \( \brace \) \( \brace \) \( \brace \) \( \frac{1}{2} \) \( \brace \) \( \brace \) \( \frac{1}{2} \) \( \brace \) \( \bra
Hirek Social Private Mailing lists SW Python RDFa it! Bookmarklets To Mendeley To Faviki bit.ly Shorten with bit.ly MID Save Video Add Zemanta
@prefix dc: <http://purl.org/dc/terms/> .
@prefix ent: <http://www.w3.org/ns/entailment/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xhv: <http://www.w3.org/1999/xhtml/vocab#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
ent:RDFS a ent:Entailment ;
             dc:creator <http://www.ivan-herman.net/foaf#me> ;
             dc:date "2009-05-03";
             dc:description "Unique identifier for RDFS Entailment";
             rdfs:comment "The specification for the RDFS entailment is part of the RDF Semantics W3C
Recommendation.";
             rdfs:isDefinedBy <a href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#rdfs">http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#rdfs</a> entailment>;
             rdfs:seeAlso <a href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">http://www.w3.org/TR/2004/REC-rdf-mt-20040210/</a>.
<http://www.w3.org/ns/entailment/data/RDFS.html> dc:title "Information Resource RDFS Entailment" ;
             xhv:stylesheet <a href="http://www.w3.org/StyleSheets/TR/base">http://www.w3.org/StyleSheets/TR/base</a>.
```



... or set up the server...

```
RewriteEngine On
RewriteBase /ns/entailment/data/

RewriteRule RDFS.ttl
/2007/08/pyRdfa/extract?format=turtle&
uri=http://www.w3.org/ns/entailment/data/RDFS.html [L]
```



Publishing RDFa

- RDFa gives an easy way of publishing RDF data on the Web
- Often, the same RDF data is available in different formats, including RDFa
 - it is up to the client to choose which one to use
 - Web Applications would rely on RDFa
 - non-RDFa aware clients would rely on RDF/XML
 - etc.



Publishing via CMS systems

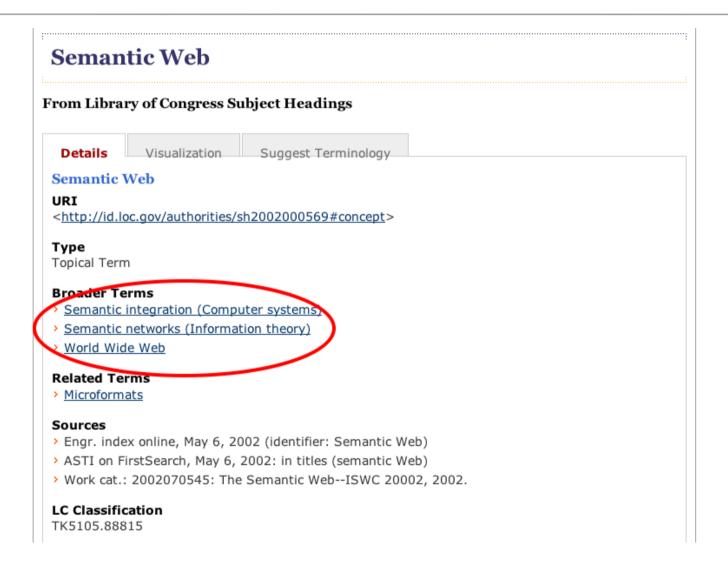
- Drupal 7 has RDFa at its core, i.e.,
 - Drupal 7 clients and users get RDFa whether they care or not!



Some deployment examples



LOC example





LOC example

```
000
                               Source of http://id.loc.gov/authorities/sh2002000569
                    Topical
Term
                 <
                    <h3>Broader Terms</h3>
                    <
                           <a href="http://id.loc.gov/authorities/sh2004000479#concept" rel="skos:broader</pre>
                              <span property="skos:prefLabel" xml:lang="en">Semantic integration (Comput
systems)</span>
                           </a>
                       <
                           <a href="http://id.loc.gov/authorities/sh92004914#concept" rel="skos:broader</p>
                              <span property="skos:prefLabel" xml:lang="en">Semantic networks (Information)
theory)</span>
                           </a>
                       <
                           <a href="http://id.loc.gov/authorities/sh95000541#concept" rel="skos:broader</pre>
                              <span property="skos:prefLabel" xml:lang="en">World Wide Web</span>
                           </a>
```



Consuming RDFa

- Various search engines begin to consume RDFa
 - Google, Yahoo, ...
 - they may specify which vocabularies they "understand"
 - this is still an evolving area
- ▶ Facebook's "social graph" is based on RDFa



Google's rich sniplet

- Embedded metadata (microformat or RDFa) is used to improve search result page
 - at the moment only a few vocabularies are recognized, but that will evolve over the years



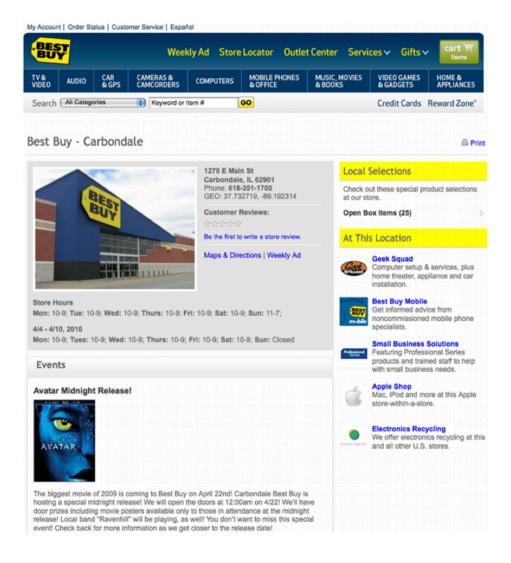


Effects of, e.g., Google

- A number of popular sites publish RDFa as part of their normal pages:
 - Tesco, BestBuy, Slideshare, The London Gazette,
 Newsweek, MSNBC, O'Reilly Catalog, the White House...
 - Creative Commons snippets are in RDFa

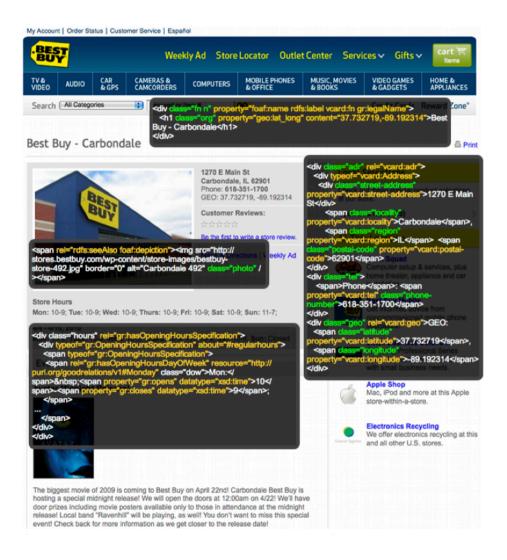


BestBuy Example for RDFa Usage





BestBuy Example for RDFa Usage



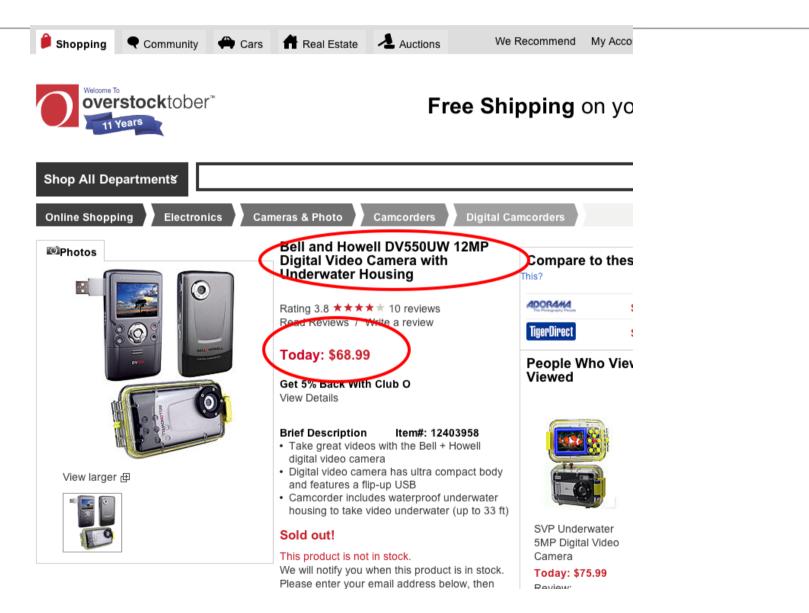


Effects on BestBuy

- Reported in a BestBuy blog:
 - GoodRelations+RDFa improved Google rank tremendously
 - 30% increase in traffic on BestBuy store pages
 - Yahoo observers a 15% increase in click-through rate
- Not bad... ⊕



Overstock.com example





Overstock.com example

