Putting Semantics on the Web with RDFa
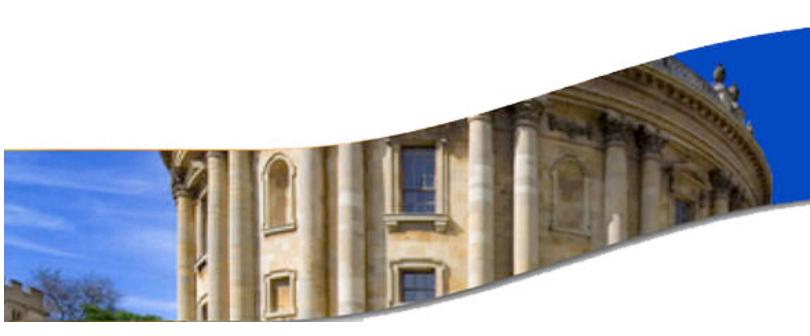
Jeni Tennison

jeni@jenitennison.com

Background

Microformats

RDFa in Metadata

RDFa in Content

Tools

Challenges

Conclusions

![XML summer school logo]

summer school

Background

www.xmlsummerschool.com

Web contains information

Humans can use it

What about computers?

   search engines

   mash-ups

How can computers make better use of the existing web of documents?

Parsing of irregular pages

    breaks when pages change

"Intelligent" guesses at meaning

    based on third-party interpretation

Hard to do
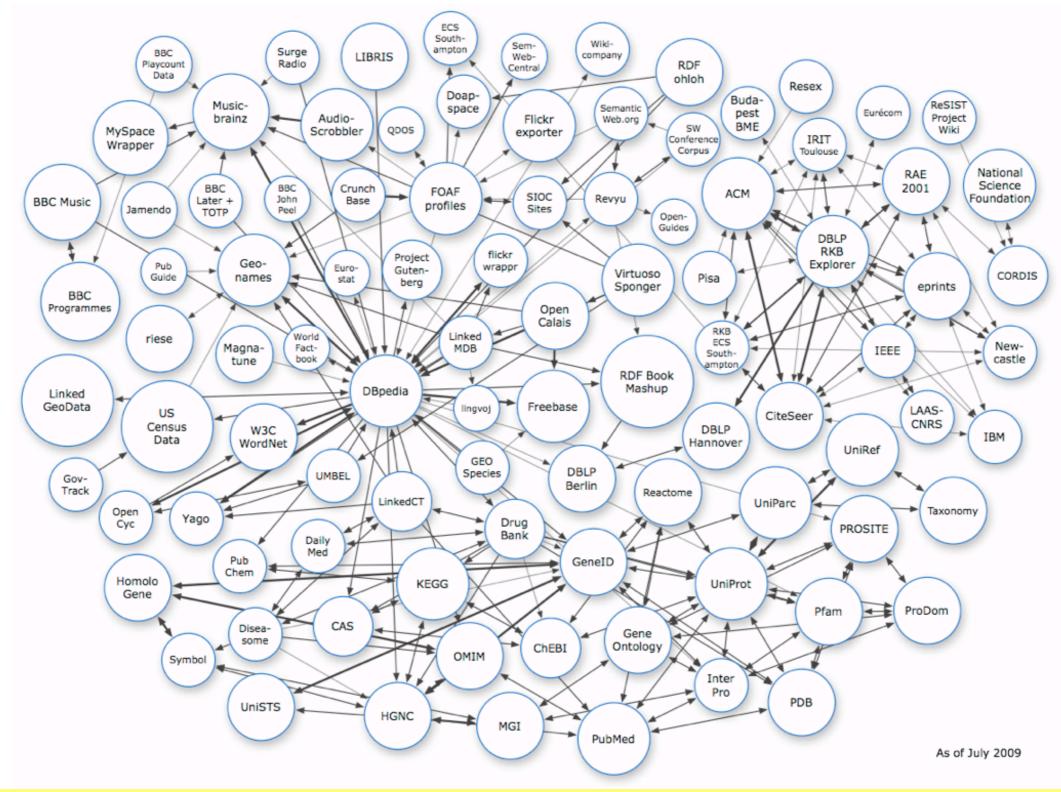
    HTML pages not designed for this

Inaccurate

www.xmlsummerschool.com

Data labelled as data

   known/consistent semantics

Data referencing data

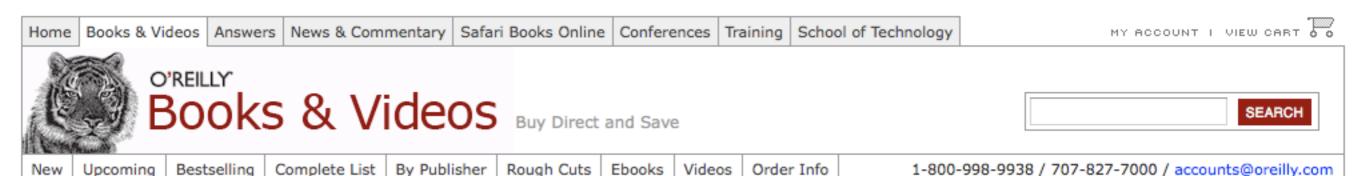   known/consistent identifiers

Easier to process

   even automatically

www.xmlsummerschool.com

As of July 2009

www.xmlsummerschool.com

7

RDFa Websites: O'Reilly

RDFa Websites: BBC

www.xmlsummerschool.com

RDFa Websites: Gazette

www.xmlsummerschool.com

Operator

microformats, some RDFa

Tails Export

microformats

Semantic Radar

RDF / RDFa

Extraction

W3C pyRdfa: http://www.w3.org/2007/08/pyRdfa/extract?uri=

Viewing

OpenLink Data Explorer: http://demo.openlinksw.com/DAV/JS/rdfbrowser/index.html?uri=

Microformats

www.xmlsummerschool.com

summer school

Microformat basics

rel-license

hCard

hCalendar

Good and bad aspects

www.xmlsummerschool.com

Simple way of embedding metadata in HTML

Use existing attributes

`class`, `rel` attributes

<abbr `title=""`*>* pattern

microformats.org

Link to license

Use rel attribute

Implies license for whole page

```
<a href="http://creativecommons.org/licenses/by/
2.0/" rel="license">cc by 2.0</a>
```

www.xmlsummerschool.com

People and organisations
Based on vCard

```
<p class="vcard">
  <span class="fn">Jeni Tennison</span><br />
  <a href="mailto:jeni@jenitennison.com"
     class="email">jeni@jenitennison.com</a><br />
  <abbr class="tel" title="+44 20 82240520">
    +44 (0)20 8224 0520
  </abbr>
</p>
```

www.xmlsummerschool.com

hCalendar

Events

Based on iCalendar

```
<div class="vevent">
  <h3 class="summary">RDFa Course</h3>
  <p><abbr class="dtstart"
      title="2009-02-06T09:30:00">9:30</abbr> -
    <abbr class="dtend"
      title="2009-02-06T16:30:00">16:30 2nd Feb
    </abbr></p>
</div>
```

www.xmlsummerschool.com

summer school

Common semantics

   everyone needs to talk about people, places, events

Easy to use

Easy to do things with

   Operator, Tails Export plugins

www.xmlsummerschool.com

Accessibility issues

`title` attribute isn't meant for machine-readable data

Mixing can be hard

Limited in extent

Not extensible

RDFa in Metadata

www.xmlsummerschool.com

RDFa background

Identifying subjects

Expressing relationships

Expressing properties

www.xmlsummerschool.com

summer school

RDF in (X)HTML

encodes data in attributes

like extensible microformats

RDF "API" within web pages

Recommendation 2007

Quick Example

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
   "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dctype="http://purl.org/dc/dcmitype/"
   xmlns:dc="http://purl.org/dc/terms/">
<head>
   <title>RDFa</title>
   <base href="http://www.jenitennison.com/id/rdfa-course" />
   <link rel="rdf:type" href="[dctype:Text]" />
   <meta property="dc:title" content="RDFa" />
   <link rel="dc:creator" href="/id/jeni" />
</head>
...
</html>
```
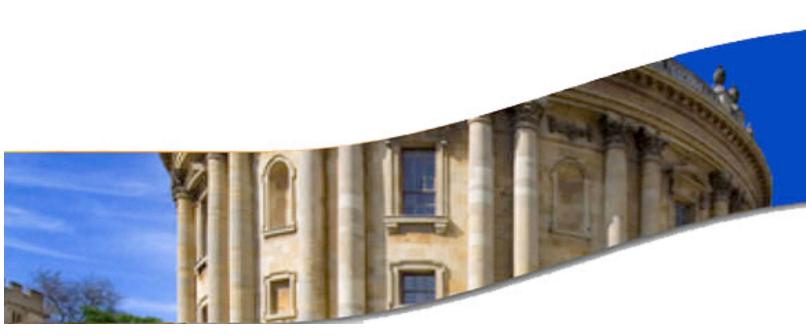
www.xmlsummerschool.com

summer school

Shorthand for URIs

declare namespaces as normal

colon separator from rest of URI

Safe CURIEs in []s

in attributes that hold URIs

```
<link rel="rdf:type" href="[dctype:Text]" />
<meta property="dc:title" content="RDFa" />
<link rel="dc:creator" href="/id/jeni" />
```

www.xmlsummerschool.com

Usually the base URI
   location of the web page
   use <base> element

Use *about* attribute

```
<link rel="dc:creator" href="/id/jeni" />
<meta about="/id/jeni" property="foaf:name"
   content="Jeni Tennison" />
```

www.xmlsummerschool.com

Relationships

In <head>, use <link>

Use rel/rev for relation

  can have multiple relations

Use href for object

```
<link rel="rdf:type" href="[dctype:Text]" />
<link rel="dc:creator" href="/id/jeni" />
```

Properties

In <head>, use <meta>

Use property for property

  can have multiple properties

Use content for value

```
<meta property="dc:title" content="RDFa" />
<meta property="dc:identifier"
  content="http://www.jenitennison.com/id/rdfa-course" />
```

Datatypes

Values that aren't natural language
  strings

  dates, booleans, URIs

Use *datatype* attribute

```
<meta property="dc:modified"
  content="2009-02-06"
  datatype="xsd:date" />
```

RDFa in Content

Use natural content of page

RDFa attributes

   allowed on any element

Content (can) provide property value

Nesting

   provides context for statements

   supports providing multiple values

summer school

```
<div about="/id/rdfa-course">
  <h3 property="dc:title">RDFa Course</h3>
  <p property="dc:description" lang="en">
    This course covers the basics of RDFa...
  </p>
  <p>
    Last modified:
    <span property="dc:modified" datatype="xsd:date"
      content="2009-02-06">
      6th Feb 2009
    </span>
  </p>
</div>
```

Use href if you want a link

Use resource if you don't

```
<div about="/id/rdfa-course">
   <h3 property="dc:title">RDFa Course</h3>
   <p>
     by <span rel="dc:creator" resource="/id/jeni">
     Jeni Tennison</span>
     (<a rel="cc:license"
         href="http://creativecommons.org/...">
       license</a>)
   </p>
</div>
```

www.xmlsummerschool.com
33

Nesting Context

Usually inherit *about*

   also `src` attribute

New resource provided by

  `resource` attribute

  `typeof` attribute

  `"hanging" rel/revs`

www.xmlsummerschool.com

```
<div about="/id/rdfa-course">
  <h3 property="dc:title">RDFa Course</h3>
  <p rel="dc:creator" resource="/id/jeni">
    by
    <span property="foaf:name">Jeni Tennison</span>
  </p>
</div>
```

```
</id/rdfa-course>
  dc:title "RDFa Course" ;
  dc:creator </id/jeni> .

</id/jeni> foaf:name "Jeni Tennison" .
```

```
<div about="/id/rdfa-course">
  <h3 property="dc:title">RDFa Course</h3>
  <p rel="dc:creator">
    by
    <span property="foaf:name">Jeni Tennison</span>
  </p>
</div>
```

```
</id/rdfa-course>
  dc:title "RDFa Course" ;
  dc:creator [ foaf:name "Jeni Tennison" ] .
```

www.xmlsummerschool.com

```
<div about="/id/rdfa-course">
  <h3 property="dc:title">RDFa Course</h3>
  <p rel="dc:creator">
    <span about="/id/jeni">
      by
      <span property="foaf:name">Jeni Tennison</span>
    </span>
  </p>
</div>
```

```
</id/rdfa-course> dc:creator </id/jeni> .
</id/jeni> foaf:name "Jeni Tennison" .
```

www.xmlsummerschool.com

Example Typeof

```
<div about="/id/rdfa-course">
  <h3 property="dc:title">RDFa Course</h3>
  <p rel="dc:creator">
    <span typeof="foaf:Person">
      by
      <span property="foaf:name">Jeni Tennison</span>
    </span>
  </p>
</div>
```

```
</id/rdfa-course> dc:creator [
  a foaf:Person ;
  foaf:name "Jeni Tennison" ] .
```

www.xmlsummerschool.com

General Tips

Interactions between attributes are complicated

test everything with a good parser
easiest if you avoid blank nodes
try to pair typeof with about

www.xmlsummerschool.com

Tools

www.xmlsummerschool.com

summer school

CSS & Javascript

GRDDL

Parsers & browsers

Search engines

Scrapers

www.xmlsummerschool.com 41

Style differently

Provide alternative views

Ubiquity

  http://code.google.com/p/ubiquity

rdfQuery

  http://code.google.com/p/rdfquery

GRDDL

Gleaning RDF from HTML

Provide an XSLT stylesheet

```
<head
  profile="http://www.w3.org/1999/xhtml/vocab
           http://www.w3.org/2003/g/data-view">
  ...
  <link rel="transformation"
    href="/styles/RDFa2RDFXML.xsl" />
  ...
</head>
```

www.xmlsummerschool.com

Human-readable HTML

GRDDL metadata

```
<head profile="/profile.html">...</head>
```

```
<head profile="http://www.w3.org/2003/g/data-view">
  ...
  <link rel="profileTransformation"
    href="/styles/toRDF.xsl" />
  ...
</head>
```

Online service

Understands GRDDL

Converts to

   RDF/XML

   Turtle

   JSON etc etc

www.xmlsummerschool.com

pyRdfa

    http://www.w3.org/2007/08/pyRdfa/
      extract?uri=

librdfa

    http://rdfa.digitalbazaar.com/librdfa/
      rdfa2rdf.py?uri=

summer school

OpenLink RDF Browser

http://demo.openlinksw.com/DAV/JS/ rdfbrowser/index.html?uri=

SIOC Browser

http://sparql.captsolo.net/browser/ browser.py?url=

www.xmlsummerschool.com

Yahoo! Search

  specialist SearchMonkey apps

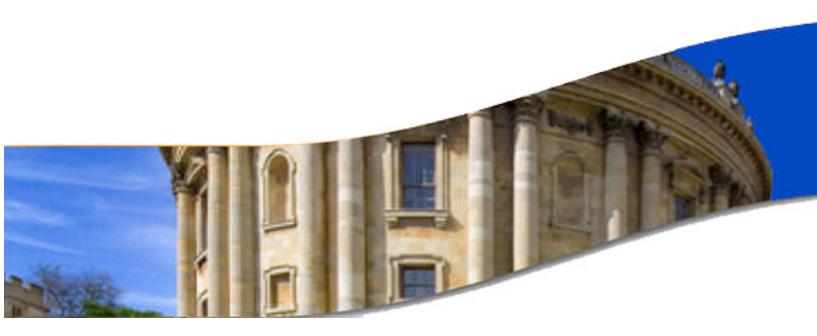Semantic search engines

  Swoogle

  SWSE: http://swse.deri.org/

summer school

Your website is your API

Parsable

Regular semantics

Expose resources

import into RDF database

run SPARQL

Challenges

www.xmlsummerschool.com

summer school

Current status

- requires namespace declarations
- if you want to be valid, use XHTML
  - requires your own DTD

Upcoming changes make it easier

- embedding in other formats (inc HTML)
- profiles and terms
- RDFa API

Semantic Content

Some web pages built on data(bases)

    easy to expose semantics

    may want to embed more data

Some content isn't yet marked up

    adding semantics is hard

    natural language processing

www.xmlsummerschool.com

Linked data identifiers

    providing good URIs for web pages

    giving required status codes

Ontologies

    choosing which to use

    creating your own

Microdata / RDFa & the future

Conclusions

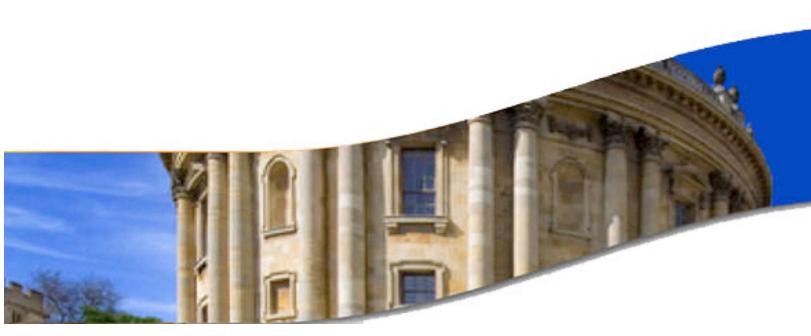www.xmlsummerschool.com

Recommend

Get your websites in order

   good URIs for things

Try to use microformats

   if you have relevant data

Start using RDFa

   snippets such as licences

   metadata in the head

   embedded in content

RDFa Primer

http://www.w3.org/TR/xhtml-rdfa-primer/

Linked Data on the Web

http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

www.xmlsummerschool.com

summer school

Any Questions?

www.xmlsummerschool.com