Building Interoperable Metadata

Hal Abelson, Ben Adida

MIT
Computer Science & Artificial Intelligence Laboratory
8 March 2006

http://ben.adida.net/presentations/
“the important thing about the semantic web is the web part.”

paraphrased from Chris Welty, Dan Connolly, Hal Abelson.
"If you think of semantic web as an interoperability technology, then it becomes capable of serendipitous behavior, of opportunism."

Ora Lassila, yesterday.
Why You Should Care

• The web really works because it is interoperable.

• What’s true for human-rendered data is true for machine-readable metadata.

• So how can web publishers make metadata interoperable?
Talk Outline

• What is Web Interoperability?
• Principles of Interoperable Metadata
• Technology for Web Metadata
• Real-World Examples
Interoperability
Technorati

Yahoo

Passatempos ... Personal ... photo ... Photo
Poetry ... Politics ... random ... Rants ... R
Science ... software ... Sport ... Sports ... s
thoughts ... Travel ... TV ... Video ... Web

Box Office Bugs Canadian Career Cartoon Ca
Condoleezza Rice Dick Cheney Dictator Doctors D
Federal Reserve Fresh Game Homestand Games G
Groin Injury H5n1 Haiti Hope Hunger Strike H
Extremism Japan Lawyers Lead Martin Michelle
Gada.be

Technorati

Tags: sunset
Related tags: Photos, Sky, Photography, Photo, Beach, Travel, Love...

Show 053 Ali'i Beach Park
Date: Feb. 11 2006 Locations: Haleiwa's A one-stop surf check today at Ali'i Beach Park. Special Guest "Eye Witness Surf Report" from the legendary...

After The Sunset
My parents were in town for a week at the end of January and we watched After The Sunset with Pierce Brosnan, Woody Harrelson, and Salma Hayek. I...

Sunset
Sailing Down the Stream of Con...

RSS Metasearch
sunrise

MSN
Magazine
Recipes | Food | Wine | Garden | Travel | Home | Life in the West
Sunrise, Calendars and Local Time

Google Blogs
Winter Sunset
Superbowl Sunset
Peaceful sunset

Flickr
Single-App Interoperability
Multi-App Interoperability
a Metadata Nirvana
Meeting the Players
Web Publishers

Blogr™

shutr™

calendr
Tool Builders

DeskBlog
All Your Blogs

Mycal
Desktop Calendar
Users

Ulrich

Ursula
They all want to share data as easily as possible.
Ulrich Gourmet

Great new Sushi place

Posted by Ulrich User 13 minutes ago

it was amazing, I tell you. The freshest fish I've seen in years. And the toro...

Fish Stinks said 3 minutes later:

I don't know what you're talking about, that sushi place made me sick.
Great New Sushi Place

it was amazing, I tell you. The freshest fish I've seen in years. And the toro...

licensed under a <a href="http://creativecommons.org/licenses/by-nc-nd/2.5/">CC License</a>.
Attribution: http://flickr.com/photos/adactio/89364513/
### Calendar

**FEBRUARY**
- Sun 26
- Mon 27
- Tue 28

**MARCH**
- Wed 1
- Thu 2
- Fri 3
- Sat 4

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Semantic Tech</td>
</tr>
<tr>
<td>8</td>
<td>sleep</td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

**Tags:** work conference talk

Suggestions: work conference talk
Great new Sushi place

Published: Tuesday, February 14, 12:05 PM

it was amazing, I tell you. The freshest fish I've seen in years. And the toro...
Great new Sushi place

it was amazing, I tell you. The freshest fish I've seen in years. And the toro...

Posted by Ulrich User 12 days ago

Posted in sushi, moderate price

License: SOME RIGHTS RESERVED

Author: Ulrich
Principles of Interoperable Metadata
Principle: Publisher Independence

- Trackbacks
- URL commenting on

- Camera and Lens Info
- Resolution, Size, etc..
Principle: Data Reuse

<h2>Great New Sushi Place</h2>
<p>it was amazing, I tell you. The freshest fish I've seen in years. And the toro... </p>

licensed under a <a href="http://creativecommons.org/licenses/by-nc-nd/2.5/">CC License</a>. 
Principle:
Self-Containment

License:  

Author: Ulrich
Principle: Schema Modularity

Posted in sushi, moderate price

Tags
- food
- sushi
- salmon
- prawn
- ginger
- e-kagen
- japanese
- brighton

DeskBlog
All Your Blogs
Principle: Evolvability
• Publisher Independence
• Data Reuse
• Self-Containment
• Schema Modularity
• Schema Evolvability
Technology for Web Metadata

- XML
- RDF
- Microformats
- GRDDL
- RDF/A
XML

http://shutr.net/schema.xsd
Beautiful Sushi.

Taken on January 15th, 2006.

licensed under a

<a href="http://creativecommons.org/licenses/by/2.5/">CC License</a>.
<?xml version="1.0"?>
<Photos xmlns="http://shutr.net/schema.xsd">
  ...
  <Photo id="123">
    <title>Beautiful Sushi</title>
    <creator>Ulrich</creator>
    <date>2006-01-15</date>
    <license url="http://cc.org/licenses/by/2.5/">
      ...
    </license>
  </Photo>
</Photos>
<Photos xmlns="http://shutr.net/schema.xsd">
  ...
  <Photo id="123">
    <title>Sushi</title>
    <creator>Ulrich</creator>
    <date>2006-01-15</date>
    <license url="http://cc.org/licenses/by/2.5/>
  ...
  </Photo>
</Photos>

<Photos xmlns="http://flickr.net/schema.xsd">
  ...
  <Photo id="456">
    <title>Sushi 2</title>
    <creator>Ursula</creator>
    <date>2006-01-16</date>
    <license url="http://cc.org/licenses/by-nc/2.5/>
  ...
  </Photo>
</Photos>
# XML Report Card

<table>
<thead>
<tr>
<th>Publisher Independence</th>
<th>everyone decides their own schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Reuse</td>
<td>data is duplicated</td>
</tr>
<tr>
<td>Self-Containment</td>
<td>HTML and XML are separate files</td>
</tr>
<tr>
<td>Schema Modularity</td>
<td>attributes are not reusable</td>
</tr>
<tr>
<td>Evolvability</td>
<td>XSLT could be applied</td>
</tr>
</tbody>
</table>
RDF

This work is licensed under a Creative Commons License.
**RDF - A Refresher**

- Everything is a triple: subject, property, object.
- Every item is either a URI or a literal (e.g. 3, “Ben”)

```xml
<cc:license <http://cc.org/licenses/by/2.5/>.
<dc:creator "Ben Adida".
</images/sunset.jpg> dc:title "Sunset on the Beach".
```
Beautiful Sushi.
Taken on January 15th, 2006.
licensed under a
<a href="http://cc.org/licenses/by/2.5/">CC License</a>.
## RDF Report Card

<table>
<thead>
<tr>
<th>Publisher Independence</th>
<th>everyone decides their own schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Reuse</td>
<td>data is duplicated</td>
</tr>
<tr>
<td>Self-Containment</td>
<td>HTML and RDF are entirely separate</td>
</tr>
<tr>
<td>Schema Modularity</td>
<td>attributes are reusable</td>
</tr>
<tr>
<td>Evolvability</td>
<td>additional fields are fine</td>
</tr>
</tbody>
</table>
Beautiful Sushi.

Taken on January 15th, 2006.

licensed under a
<a rel="license"
href="http://creativecommons.org/licenses/by/2.5/">CC License</a>.
“If some usage of the web is too nascent to have converged, we can’t easily codify it, so we choose to pass on the problem.”

Ryan King, Microformats.org

http://microformats.org/blog/2006/01/17/microformats-are-semantic-markup/
<table>
<thead>
<tr>
<th>Publisher Independence</th>
<th>schemas are tightly controlled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Reuse</td>
<td>HTML rendered data is reused</td>
</tr>
<tr>
<td>Self-Containment</td>
<td>Same file, but WHOLE file.</td>
</tr>
<tr>
<td>Schema Modularity</td>
<td>attributes are not reusable, though some MFs are composable.</td>
</tr>
<tr>
<td>Evolvability</td>
<td>may happen with GRDDL.</td>
</tr>
</tbody>
</table>
GRDDL

XML Schema
XML
XML
RDF
HTML
XSLT
XSLT
XSLT
Beautiful Sushi.

Taken on January 15th, 2006.

licensed under a CC License
# GRDDL Report Card

<table>
<thead>
<tr>
<th>Publisher Independence</th>
<th>everyone decides their schema, but it all depends on the transform.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Reuse</td>
<td>transform reuses data</td>
</tr>
<tr>
<td>Self-Containment</td>
<td>no correspondence</td>
</tr>
<tr>
<td>Schema Modularity</td>
<td>attributes are reusable if exposed as RDF.</td>
</tr>
<tr>
<td>Evolvability</td>
<td>additional fields are fine, but controlled by the transform</td>
</tr>
</tbody>
</table>
RDF/A

- Product of W3C’s Task Force on RDF-in-HTML.
- Joint effort of HTML and SemWeb working groups.
- Working Draft of the RDF/A Primer available in the next 2 weeks.
Beautiful Sushi.

taken by

Ulrich

on 2006-01-02

licensed under a

CC License
# RDF/A Report Card

<table>
<thead>
<tr>
<th>Publisher Independence</th>
<th>everyone decides their own schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Reuse</td>
<td>similar to MF</td>
</tr>
<tr>
<td>Self-Containment</td>
<td>chunk of HTML is self-contained</td>
</tr>
<tr>
<td>Schema Modularity</td>
<td>attributes are reusable</td>
</tr>
<tr>
<td>Evolvability</td>
<td>same as RDF</td>
</tr>
</tbody>
</table>
RDF/A

Elegant Degradation

• Fully marked up XHTML

• Client software interprets as much as it can understand, structuring it for the user to see.

• Anything the client doesn’t understand is just XHTML-rendered.
Calendr Events

Lessig - Free Culture
Larry will be giving his talk on Free Culture at CSAIL on 2006-03-08 14:00-05 for about 1 hour
Calendr Events

Calendar.

Calendr Events.

available under a [CC license](http://creativecommons.org/licenses/by/2.5/) CC license.

Lessig - Free Culture

Larry will be giving his talk on Free Culture at CSAIL on 2006-03-08 14:00-05 for about 1 hour.
Real-World Examples
Creative Commons

- Including CC in a data model.

- Sometimes, users control only a small part of the HTML of their page. That HTML should be entirely self-describing!

- Other times, publishers want to publish XML with CC data. How can a generic CC client detect that?

- RDF/A for CC HTML; GRDDL for CC XML.
Libraries / MIT DSpace

- OAI effort to standardize archive formats.
- the XML schema has been forked 150 times among 800 institutions.
- Would RDF/A prevent forking? No, of course not, but what is compatible would remain compatible.
- GRDDL
• Publisher Independence
• Data Reuse
• Self-Containment
• Schema Modularity
• Schema Evolvability
So...

Existing Datasets: GRDDL

New HTML applications: RDF/A

Questions?

CSAIL