From Atom's to OWL's:
The new ecology of the WWW

Jim Hendler
Hendler@cs.umd.edu
http://www.cs.umd.edu/~hendler
From Atom*s to OWL§s: The new ecology of the WWW

Jim Hendler
Hendler@cs.umd.edu
http://www.cs.umd.edu/~hendler

* syntax
§ semantics
Acknowledgements

• In preparing this talk I have mostly ignored the advice of a great many people including Tim Berners-Lee, Dan Connolly, Wendy Hall, Eric Miller, Brand Nieman, Bijan Parsia, Guus Schreiber, Nigel Shadbolt, and Frank VanHarmelen

• I owe much to my research group
  – Details at http://www.mindswap.org/ (Our Semantic Web Portal)
The Web used to be easy (ca. 1990)

• Documents (HTML)
  – Emacs or vi via some cutting and pasting and it showed in your browser - woohoo!

• And Links (HTTP)
  – Install LibWWW, customize parameters, and you were up and running - woohoo!
But that didn't last long... (ca 2000)
Excelsior …

- **Syndication**
  - Atom, JSON, YAML…

- **Microformats**
  - hCard, hCalendar, hReview, XFN, …

- **XML specs**
  - XML 1.1, Xinclude, XML:id, …

- **Semantic Web specs**
  - RDF (XMP), RDFS, OWL, SPARQL, Rules …

- **Web Service languages**
  - WSDL 2.0, WSPL, WS Adressing, CDL, …
  - (and all of the above applied to services
    - UDDI v3, XOXO, XML Protocol, WSDL-RDF, …)
## What is this all about?

<table>
<thead>
<tr>
<th>Category</th>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text pages</td>
<td>A</td>
<td>Single pages</td>
</tr>
<tr>
<td>Music</td>
<td>A-</td>
<td>© and access issues</td>
</tr>
<tr>
<td>Images</td>
<td>B+</td>
<td>Lacks precision</td>
</tr>
<tr>
<td>Video</td>
<td>C+</td>
<td>Promising start</td>
</tr>
<tr>
<td>Personal Info</td>
<td>C</td>
<td>Splintered/ imprecise</td>
</tr>
<tr>
<td>Services</td>
<td>C-</td>
<td>Need to play together better</td>
</tr>
<tr>
<td>Data</td>
<td>F</td>
<td>Not working at a Web scale</td>
</tr>
</tbody>
</table>

### Web Search Report Card

XML 2005
Common Goal…

XML: Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet. (Wikipedia)

Sem Web: The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries (W3C SWA)

It's the data, stupid.
Increasingly Common Syntax

• The new ecology is all in XML …
Atom (XML)

<?xml version="1.0" encoding="utf-8"?>
<feed version="0.3" xmlns="http://purl.org/atom/ns#">
  <title>dive into mark</title>
  <link rel="alternate" type="text/html" href="http://diveintomark.org/"/>
  <modified>2003-12-13T18:30:02Z</modified>
  <author>
    <name>Mark Pilgrim</name>
  </author>
  <entry>
    <title>Atom 0.3 snapshot</title>
    <link rel="alternate" type="text/html" href="http://diveintomark.org/2003/12/13/atom03"/>
    <id>tag:diveintomark.org,2003:3.2397</id>
    <issued>2003-12-13T08:29:29-04:00</issued>
    <modified>2003-12-13T18:30:02Z</modified>
  </entry>
</feed>

Use XML in syndication
Adobe XMP (RDF-based)

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about=""
        xmlns:pdf="http://ns.adobe.com/pdf/1.3/">
        <pdf:Producer>Acrobat Distiller 7.0.5 for Macintosh</pdf:Producer>
    </rdf:Description>
    <rdf:Description rdf:about=""
        xmlns:dc="http://purl.org/dc/elements/1.1/">
        <dc:creator>
            <rdf:Seq>
                <rdf:li>James Hendler</rdf:li>
            </rdf:Seq>
        </dc:creator>
        <dc:title>
            <rdf:Alt>
                <rdf:li xml:lang="x-default">XMLideas.ppt</rdf:li>
            </rdf:Alt>
        </dc:title>
    </rdf:Description>
    <rdf:Description rdf:about=""
        xmlns:xapMM="http://ns.adobe.com/xap/1.0/mm/">
        <xapMM:DocumentID>uuid:93277c40-5534-11da-a3f2-000a95d6b344</xapMM:DocumentID>
        <xapMM:InstanceID>uuid:9327882b-5534-11da-a3f2-000a95d6b344</xapMM:InstanceID>
    </rdf:Description>
</rdf:RDF>
```

Embed meta-data in PDF, etc.
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/>"
    <foaf:Person>
        <foaf:name>Jim Hendler</foaf:name>
        <foaf:title>Dr</foaf:title>
        <foaf:firstName>Jim</foaf:firstName>
        <foaf:surname>Hendler</foaf:surname>
        <foaf:mbox_sha1sum>be972c7a602683f7cf3c7a1fd0949c565debe4d3
            </foaf:mbox_sha1sum>
        <foaf:workplaceHomepage rdf:resource="http://owl.mindswap.org"/>
    </foaf:Person>
</rdf:RDF>
FEA-RMO (OWL)

<owl:Ontology rdf:about="file:/C:/Models/eGov/FEA/all.owl">
  <owl:versionInfo>Version 1.0</owl:versionInfo>
  <dc:contributor>Dean Allemang, TopQuadrant</dc:contributor>
  <dc:description>FEA Reference Model Ontology</dc:description>
    <owl:imports rdf:resource="http://purl.org/dc/elements/1.1"/>
  <dc:source>FEAPMO</dc:source>
  <dc:source>The Performance Reference Model Version 1.0, FEAPMO</dc:source>
  <dc:date>January 2005</dc:date>
  <dc:contributor>Irene Polikoff, TopQuadrant</dc:contributor>
  <dc:contributor>Ralph Hodgson, TopQuadrant</dc:contributor>
</owl:Ontology>

  <owl:disjointWith>
  </owl:disjointWith>
</owl:Class>

<owl:Class rdf:about="http://www.osera.gov/owl/2004/11/fea/brm#SF_IntelligenceOperations">
  <owl:disjointWith>
  </owl:disjointWith>
</owl:Class>

Application specific vocabulary (ontology) definition
So what is the difference?

Documents: XML is a system for defining, validating, and sharing **document formats**. XML uses tags to distinguish document structures, and attributes to encode extra document information. XML will look very familiar to those who know about SGML and HTML. (XML FAQ)

Links: RDF is based on the idea of **identifying things using Web identifiers** (called Uniform Resource Identifiers, or URIs), and describing resources in terms of simple properties and property values. This enables RDF to represent simple statements about resources as a graph of nodes and arcs representing the resources, and their properties and values. (RDF Primer)
Extending the Web

• Bringing new document types, media types, services, and (especially) data to the Web requires BOTH
  – More complex document processing
    • Formats improvement, correctness checking, subdocument components
  – More complex link types
    • Must capture, but not be limited by, structure and models
    • Must have a link "labeling" and ability to "reason" about those links
Compare Xlink

Which provides some info around a link - but no external referent to an entity (i.e. no URI is assigned to this link information)
To RDF

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>

Which assigns specific URIs to the individual fields (i.e. the URI in the rdf:about is a persistent and externally addressable URI) and each "property" (fullname, mailbox, etc.) is defined by a URI)
Or in graph (link!) form

- URIs for the relations is important
  - common naming
  - dereferencing!
Adding the Semantics

- This is an RDF "Triple" (resources can also be URIs)
- Make the URI dereferencible for a machine-readable description
  - That is, put the ontology at that URI!
  - RDFS and OWL are basically XML dialects for this
Adding the **Semantics**

- **Ex:** Asserts that email address is a many to one relation (i.e. different people with same email address will be considered equivalent)
  - Can merge multiple FOAF files describing same user from different applications
Adding the Semantics

- RDFS and OWL provide other property restrictions as well
  - Which are properties needed for data modeling, domain vocabularies, etc.
- This is the heart of the Semantic Web
  - Why we talk links, not documents
Real Progress Occurring

• Increasing Semantic Web Uptake
  – Start ups in the space
    • Cerebra, Siderean, SandPiper, Data Grid…
  – Bigger players buying in
    • Adobe, Cisco, HP, IBM, Nokia, Oracle, Sun, Vodaphone… announcements/use in '05
  – Government Projects in and across agencies
    • US, EU, Japan
  – Life Sciences/Pharma an increasingly important market
    • Health Care and Life Sciences Interest Group at W3C
  – Many open source tools available
    • Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Onto(xxx)
So…

- XML derives from a focus on the documents
  - Structuring the internals of the nodes of the Web graph (tree structure)
  - Providing structure documents for data description
- RDF/Sem Web derives from a focus on the links between them
  - Labeling the links
  - Adding "restrictions"
  - "Merging" the data from disparate descriptions
Applied to the Web

Jim Hendler, Professor, University of Maryland & Director of Semantic Web and Agent Technology, Maryland Information and Network Dynamics Laboratory

From Atoms to OWLs the new ecology of the Semantic Web
Over the past couple of years, Semantic Web deployment has really started rolling. Successes have included adoption of RDF by major corporations and the development of new ontology-based technologies of use for many enterprise and web applications. Despite this, controversy still seems to abound with respect to both the relationship of the Semantic Web to XML, and the use of these technologies. This talk will explain what the Semantic Web is all about and, perhaps more importantly, attempt to dispel two pervasive myths -- that XML and the Semantic Web are incompatible, and that XML is able to do all that the Semantic Web promises without reinventing the semantic extensions inherent in RDF and OWL.

Bio: Jim Hendler is a Professor at the University of Maryland and the Director of Semantic Web and Agent Technology at the Maryland Information and Network Dynamics Laboratory. He has joint appointments in the Department of Computer Science, the Institute for Advanced Computer Studies and is an affiliate of the Institute for Systems Research. He has authored about 200 technical papers in the areas of artificial intelligence, Semantic Web, agent-based computing, and high-performance processing. Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former Chief Scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a US Air Force Exceptional Civilian Service Medal in 2002, and is a member of the World Wide Web Consortium's Semantic Web Coordination Group. He is the Editor in Chief of IEEE Intelligent Systems and is on the Board of Reviewing Editors for Science.
XML Markup

>Adds the syntax:
which arguments go where
how to parse the elements
That define the DOCUMENT
But for the *semantics* we need information about Talks, Subjects, People, Events, etc. and the roles this item plays in them.
Thus, the **Semantic Web**
Not just annotation...

Media formats, Data, Web service descriptions, etc. can be merged via their "semantic relations"
Content-based metadata

http://www.mindswap.org/2003/PhotoStuff/
"Folksonomy" vs. Ontology

Flickr - protein
Sem Web - protein

Different levels of precision needed for different users
Merging data

Piggy Bank, MIT Simile project, is a Firefox addition that allows the merging of RDF data on the fly
Merging services

Information management capabilities
Discovery, Filtering, Composition

(Evren Sirin, Bijan Parsia)
Get a B&N price (In Euros)
Of a particular book
In its German edition?
Reasoning about access control

1) If $X$ is AC rep of $Y$, $X$ can delegate W3C member access rights in $Y$.
2) Kari is AC rep of Elisa.

1) If $X$ is employee of Elisa, $X$ has W3C member access rights.
2) Tiina is employee of Elisa.

Tiina: I have W3C member access rights
Proof: Alan 1, Alan 2, Kari 1, Kari 2
Which gets us back to protocols…

(A) User requests a resource.

(b) 401 error provides access rules.

(C) Proof is generated and pointer is sent in new HTTP-Get request.

(D) Proof is checked, and confirmed, and the transaction succeeds.
But can't this just be done in XML?

• Yes, it can … and it has
  – RDF/XML is legal XML
  – OWL has an XML Schema and XSLT to move that to the (normative) RDF/XML interchange format
  – SPARQL returns an XML (data) document
  – GRDDL and RDF/A provide XHTML to RDF mapping

• No need to reinvent the wheel, but…
Playing nice together

- Good news: Slash vs. Hash is over
  - The TAG has spoken
    - The armies are burying their dead and returning home as we speak
- And coordination improving
  - Cf. WSDL 2.0 OWL; SPARQL and XQuery, ...
- But, Semantic Web needs some URIs from the XML world
  - Example: OWL cannot use complex XML schema data types
  - Similarly Xlink and Xpath issues
Punch line

• We work towards a common goal
  – Raising the grade for data (and meta-data) on the Web
• XML and RDF/RDFS/OWL can and do coexist very well
  – XML for documents and structure
  – RDF for linking and merging
  – OWL for vocabularies and organization
    • RDF and XML forms
• The Semantic Web is here
  – don't reinvent it