Web Feeds

Jim Downing, University of Cambridge

Introduction
Quick poll on who uses them?
Who has published them?
Who knows the difference between RSS 0.90 and 0.91
Developer Happiness Days
http://dev8d.jiscinvolve.org/
In the next half hour

- General wittering about interop and standards
- Some details about RSS family, and about Atom
- What and how to feed and consume
- Advanced stuff, efficient feeds, archive feeds
Another View of Interoperability
Ways of adding value to the resources in your repository

- Do It Yourself
- Let other people do it in other systems
Interoperability
= 
Increasing the probability of reducing the costs of working with other systems

... and you don’t necessarily know what the other systems are yet. It isn’t just about everything working automagically together, which tends to be either very simple integration, or else more like harmonization in approach.
Serendipity vs Predictable ROI

Because the value in interoperability is about probability, it’s like winning on the horses; magic when it happens, expensive when it doesn’t.

It’s rather difficult to predict ROI on a single interoperability event, so you spread your risks.

Study form, look at mechanisms and techniques that are delivering value in analogous systems and transfer them.

Focus on the resources, rather than the services.
Standards

- Fit for purpose
- Adoption
- Motivation

Standards can be useful if they are adopted, because they increase the probability that two systems will understand each other. They fail when they don’t decrease the cost (by not providing the desired solution), and when they don’t get adopted (usually because they don’t provide the desired solution).
Fit to solution + adoption + motivations of standardizers

BUT “Open Standard” is not a simple tick box. De facto standards are often more powerful, and a non-standard protocol that works will beat an open standard that doesn’t.
Web Feeds
Definition, Origins of Species
Web feeds

A web feed (or news feed) is a data format used for providing users with frequently updated content. Content distributors syndicate a web feed, thereby allowing users to subscribe to it.

Wikipedia
http://en.wikipedia.org/wiki/Web_feed

“Web Feed” is the generic term for RSS and Atom feeds.
Repositories exist on the web

Feeds are great for outside-in value

=> They’re great for repositories

A conceit for the sake of this presentation: the web is the only delivery channel that really matters. So web feeds are the second most important thing about your repository after the web pages.
Although a single, ordained standard may have saved effort, there were significant political and technical issues that separated the two camps – it’s unlikely that an open standard based on any of these protocols would have won the whole community over.
This is the only de facto standard the community were able to agree on.
Atom

• All the features of the others
• Extensibility of RDF RSS
• Most of the simplicity of Really Simple Syndication.
• Less tech-religion
• Atom is closest to having universal adoption

Convergence on Atom seems ineluctable.
Silver Lining

Because of the RSS trainwreck, most libraries and tools support most versions of RSS / Atom for simple cases.
Web Feed Rally
+ a look under the bonnet
A common application for web feeds is to syndicate blog content.
http://www.google.co.uk/igoogle/
Show me the code!

[Off to JEdit to look at some Atom XML.]
Auto-Discovery

So how does a feed reader like GReader know how to get hold of the feed from Les’s page? How does Firefox know how to find the feed?

Auto-discovery is an essential part of usability for web-feeds. You probably lose 50% of the potential users without it.

e.g. Until recently you couldn’t put a feed URL straight in to GReader.

BTW – Auto-discovery of interoperability services from HTML pages is applicable to many different types of service, and is just as important there.
Consuming Atom

- Google Reader
- Widgets
- Aggregators
- pipes.yahoo.com
- www.pageflakes.com
- et cetera, et cetera

There are loads and loads of
http://www.dspace.cam.ac.uk/handle/1810/23

Not just for blog content – a collection in my IR.
DSpace Community: Social Anthropology

Show: 10 new items - all items  Mark all as read  Refresh  Feed settings...-

☆ Interview of Stephen Cleobury
Title: Interview of Stephen Cleobury
Authors: Cleobury, Stephen
Abstract: An interview on the life and work of the musician Stephen Cleobury, Director of Music at King's College, Cambridge
Description: Interviewed and filmed by Alan Macfarlane on 4th July 2008 and edited by Sarah Harrison

☆ Interview of Michael Bate
Title: Interview of Michael Bate
Authors: Bate, Michael
Abstract: An interview with the biologist Michael Bate of the University of Cambridge on his life and times
Description: Filmed on 2nd July 2008 by Alan Macfarlane and edited by Sarah Harrison

☆ 'On the anthropology of the contemporary'
Title: 'On the anthropology of the contemporary'

8:06 AM (seconds ago)
What can we feed?

- Any set / stream of content
  - Collection
  - Topic
  - Author
  - Personalised interest

- Different types of content
  - Textual
  - Images
  - Audio / video
  - Data
Extending Atom

- Any namespace is allowed
- Dublin Core metadata
- RDF
This view shows aggregation – a number of different calendar feeds
Aggregability

- Data model is close enough the same between feed formats
- Always sorted by date modified

Universal sort order is a vital piece of this all working together. If you have a specific need to sort by something else, that’s OK (except by the Atom spec), but you’ll lose some interop.
Geo-tagged feeds

Uploads from Jim Downing, with geodata

DSC00721
Tue, Apr 22, 2008 11:05
Jim Downing posted a photo:

Paul goes ape.

Media files

DSC00721
Tue, Apr 22, 2008 11:05
Jim Downing posted a photo:

Media files
Enclosures

• Just ‘blessed’ links
• Anything can be enclosed

<link href="http://example.com/podcasts/show001-usenglish.mpg" rel="enclosure" length="21472922" type="audio/mpeg" />
<entry>
  <title>Summary page for crystal structure from DataBlock 2007sot1536 in CIF b816002csup1 from article b816002c in issue 2008/44 of Royal Society of Chemistry, Chemical Communications.</title>
  <link href="http://wwmm.ch.cam.ac.uk/crys...02csup1_2007sot1536.cif.summary.html"/>
  <id>urn:uuid:8de35128-15c8-4c5a-91c4-e413a478f2c9</id>
  <summary>Summary page for crystal structure from DataBlock 2007sot1536 in CIF b816002csup1 from article b816002c in issue 2008/44 of Royal Society of Chemistry, Chemical Communications.</summary>
  <updated>2008-11-15T06:31:59Z</updated>
  <link rel="enclosure" href="http://wwmm.ch.cam.ac.uk/crystaleye/sum...536.complete.cml.xml" type="chemical/x-cml" length="103999"/>
  <link rel="enclosure" href="http://wwmm.ch.cam.ac.uk/crystaleye/su...l_2007sot1536_l.small.png" type="image/png" length="5348"/>
  <content type="xhtml">
    <div xmlns="http://www.w3.org/1999/xhtml">
      <img src="http://wwmm.ch.cam.ac.uk/crystaleye/summary/rsc/cc/2008/44/data/b816002c/b816002csup1_2007sot1536/b816002csup1_2007sot1536_l.small.png"/>
    </div>
  </content>
</entry>

http://wwmm.ch.cam.ac.uk/crystaleye/feed/atom/feed.xml
Progressive enrichment; works in simple reader for simple purposes, works in more sophisticated consumers for more sophisticated stuff.
**EdShare: Collaboration in a University for learning and teaching**

Magpi Evans, Les Carter, Selvendran Francois, Janet Metcalf, Tim McLeod, Debra Morris, Sue Whitaker

**Approach:**
- Extending the EPrints platform to develop an institutional repository
- Supporting collaboration and sharing in learning and teaching
- Creating a web 2.0 content-centred environment
- Enabling tagging and allocation of metadata to capture and present multi-media resources
- Supporting collaborative partnerships and new communities to enrich the student experience

**Motivation:**
- To enable people to:
  - Share knowledge, from across the institution,
  - Work on:
    - Identifying, describing, archiving, and preserving educational materials, resources, and processes
    - To be able to:
      - View, access, find, and associate with the repository,
      - Contribute to development of the content

**Sharing through EdShare:**
- Me - initial sparks, emerging pedagogy, mature and complete "stuff"
- The University - sharing between disciplines, collaboration across faculties
- The World - showcasing the products of the whole institution

www.edshare.soton.ac.uk
Advanced Topics
‘Secure’ feeds

• Many feed readers cannot cope with properly auth secured feeds
• Do you want to encourage your users to give their repo username and password to Google?
• Usual approach is to use obfuscated URLs
# E-Tags

<table>
<thead>
<tr>
<th></th>
<th>Client</th>
<th>Server</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First GET</strong></td>
<td>GET /feed.atom</td>
<td>200 OK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E-Tag:foobar</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;feed xmlns:....</td>
</tr>
<tr>
<td><strong>No updates</strong></td>
<td>GET /feed.atom</td>
<td>304 Not Modified</td>
</tr>
<tr>
<td></td>
<td>If-None-Match:foobar</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Updated</strong></td>
<td>GET /feed.atom</td>
<td>200 OK</td>
</tr>
<tr>
<td></td>
<td>If-None-Match:foobar</td>
<td>E-Tag:barblat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;feed xmlns:....</td>
</tr>
</tbody>
</table>
## Modification Date

<table>
<thead>
<tr>
<th></th>
<th>Client</th>
<th>Server</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First GET</strong></td>
<td>GET /feed.atom</td>
<td>200 OK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;feed xmlns:...</td>
</tr>
<tr>
<td><strong>No updates</strong></td>
<td>GET /feed.atom</td>
<td>304 Not Modified</td>
</tr>
<tr>
<td><strong>Updated</strong></td>
<td>GET /feed.atom</td>
<td>200 OK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;feed xmlns:...</td>
</tr>
</tbody>
</table>
Archived Feeds

- RFC 5005
- http://www.ietf.org/rfc/rfc5005.txt
- A RESTful, simple approach to harvesting
- If combined with metadata embedding and enclosures, is a persuasive alternative to OAI-PMH
Subscription Feed

<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <title>Example Feed</title>
  <link href="http://example.org"/>
  <link rel="self" href="http://example.org/index.atom"/>
  <updated>2003-12-13T18:30:02Z</updated>
  <author>
    <name>John Doe</name>
  </author>
  <id>urn:uuid:60a76c80-d399-11d9-b93C-0003939e0af6</id>
  <entry>
    <title>Atom-Powered Robots Run Amok</title>
    <link href="http://example.org/2003/12/13/atom03"/>
    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
    <updated>2003-12-13T18:30:02Z</updated>
    <summary>Some text.</summary>
  </entry>
</feed>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:fh="http://purl.org/syndication/history/1.0">
  <title>Example Feed</title>
  <link rel="current" href="http://example.org/index.atom"/>
  <fh:archive/>
  <link rel="prev-archive" href="http://example.org/2003/10/index.atom"/>
  <updated>2003-11-24T12:00:00Z</updated>
  <author>
    <name>John Doe</name>
  </author>
  <id>urn:uuid:60a76c80-d399-11d9-b93c-0003939e0af6</id>
  <entry>
    <title>Atom-Powered Robots Scheduled To Run Amok</title>
    <link href="http://example.org/2003/11/24/robots_coming"/>
    <id>urn:uuid:cdef5c6d5-gff8-4ebb-assa-80dwe44efkjo</id>
    <updated>2003-11-24T12:00:00Z</updated>
    <summary>Some text from an old, different entry.</summary>
  </entry>
</feed>
Conclusions

- Web feeds are an important part of the web
- Loads of tools for dealing with them
- Most data can be represented in a feed
- Progressive enrichment
- Important way of adding value to IR users