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/

Plug!

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 analagous 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

The rest is detail

Feeds are great for outside-in value

Which we'll see today

=> 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.

RSS war

- RDF Site Syndication
  - 0.90
  - 1.0
- Really Simple Syndication
  - 0.91
  - 0.92
  - 0.94
  - 2.0

http://diveintomark.org/archives/2004/02/04/incompatible-rss

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

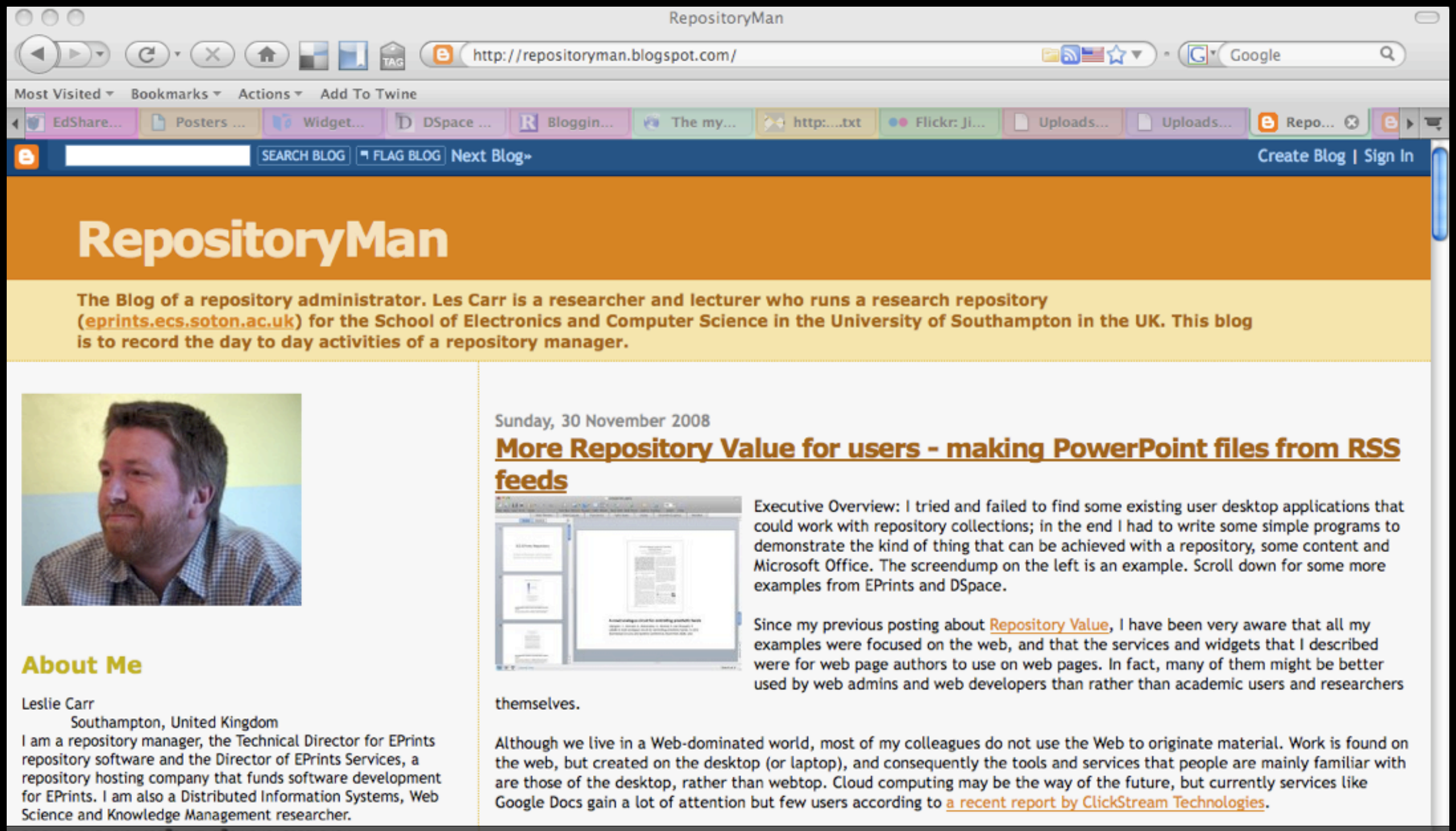
Convergence on Atom seems ineluctable.

Silver Lining

Because of the RSS trainwreck, most libraries and tools support most versions of RSS / Atom for simple cases

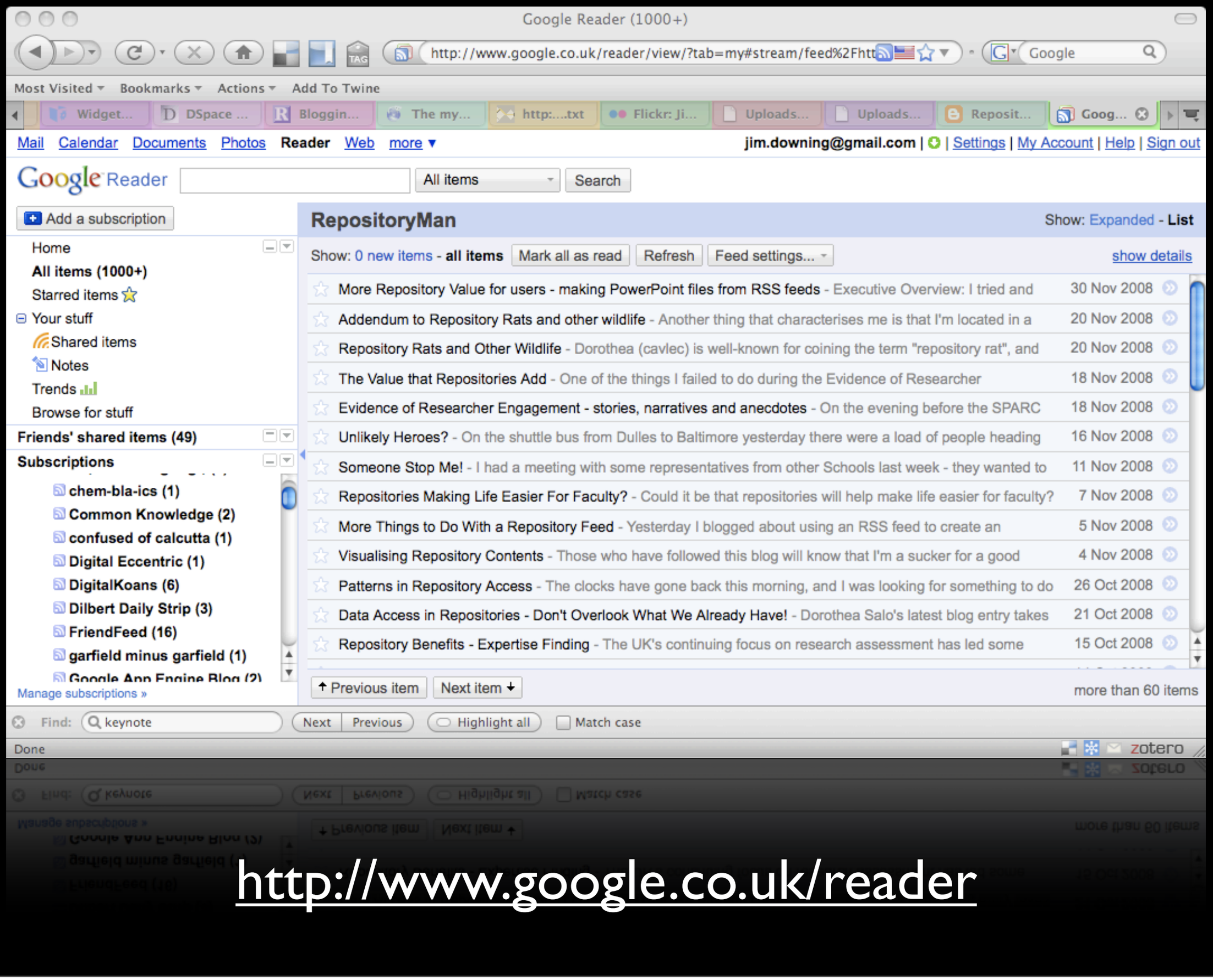
Web Feed Rally

+ a look under the bonnet



http://repositoryman.blogspot.com/

A common application for web feeds is to syndicate blog content.



http://www.google.co.uk/reader

ies Maps News Shopping Mail more ▾ jim.downing@gmail.com | Classic Home | My Account |

iGoogle

Google Search I'm Feeling Lucky

Advanced Search
Search Preferences
Language Tools

Get artist themes | Change theme from Encroaching Smoke V3.0 | Ad

Day
UK ...
e Day
uar...
s
ndar
der
nve...
vid...
ory...

RepositoryMan

- More Repository Value for users - making PowerPoint files from RSS feeds
- Addendum to Repository Rats and other wildlife
- Repository Rats and Other Wildlife

Google Calendar

« December 2008 »

S	M	T	W	T	F	S
30	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3
4	5	6	7	8	9	10

SWORD - rwidgets.co.uk

<sword />

The **Simple Web-service Offering Repository Deposit** (SWORD) project is a lightweight protocol for deposit in the form of a profile of the Atom Publishing Protocol.

Click here to visit the official website

Click here to start using the widget

share powered by netvibes

Google Reader (1000+)

All items (1000 refresh mark all as read

Enter Location

Select a different location:
Country/Region: United States
City:

Save

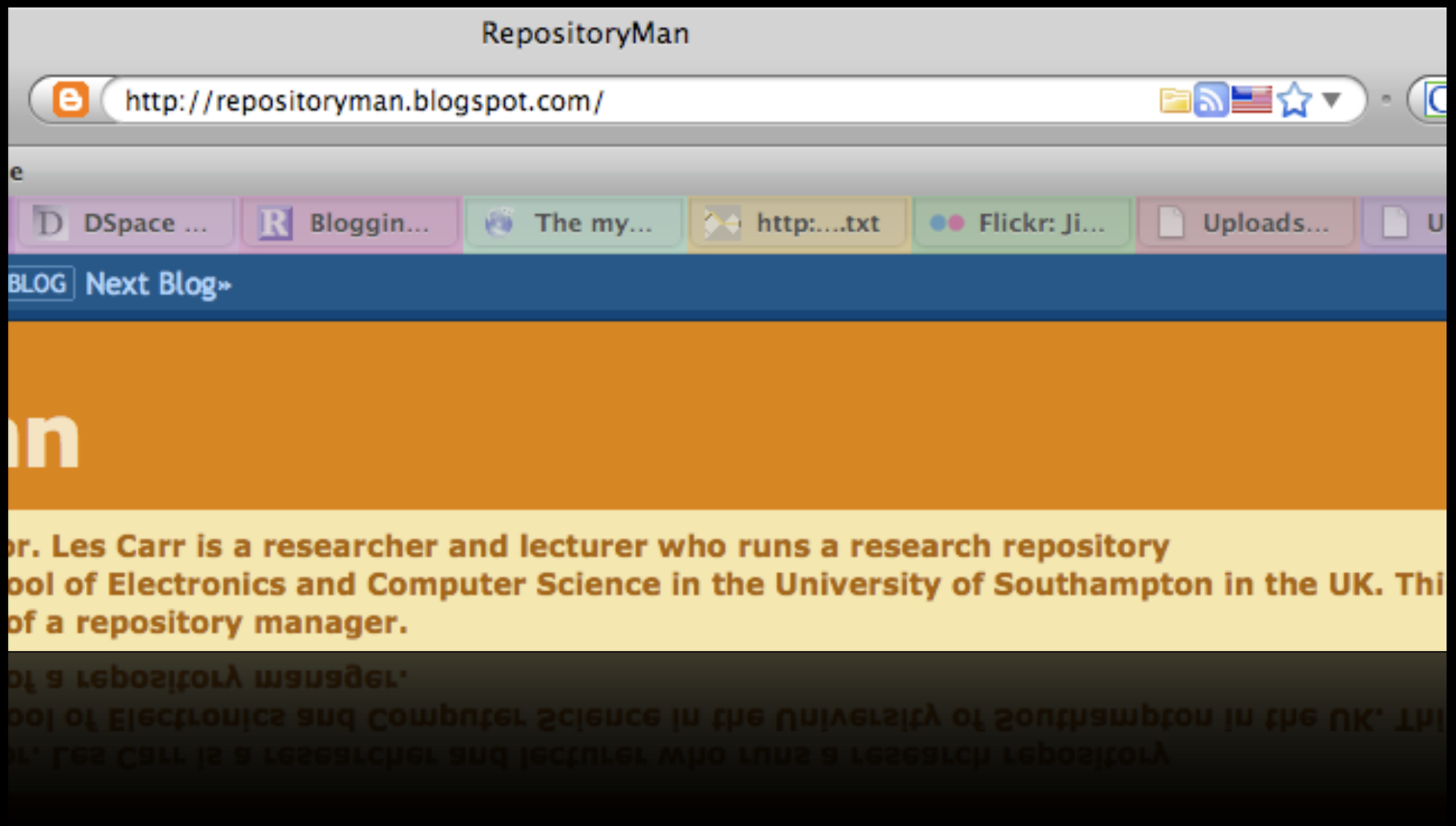
Currency Conversion

Amount
From Australian dollar
To British pound

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.

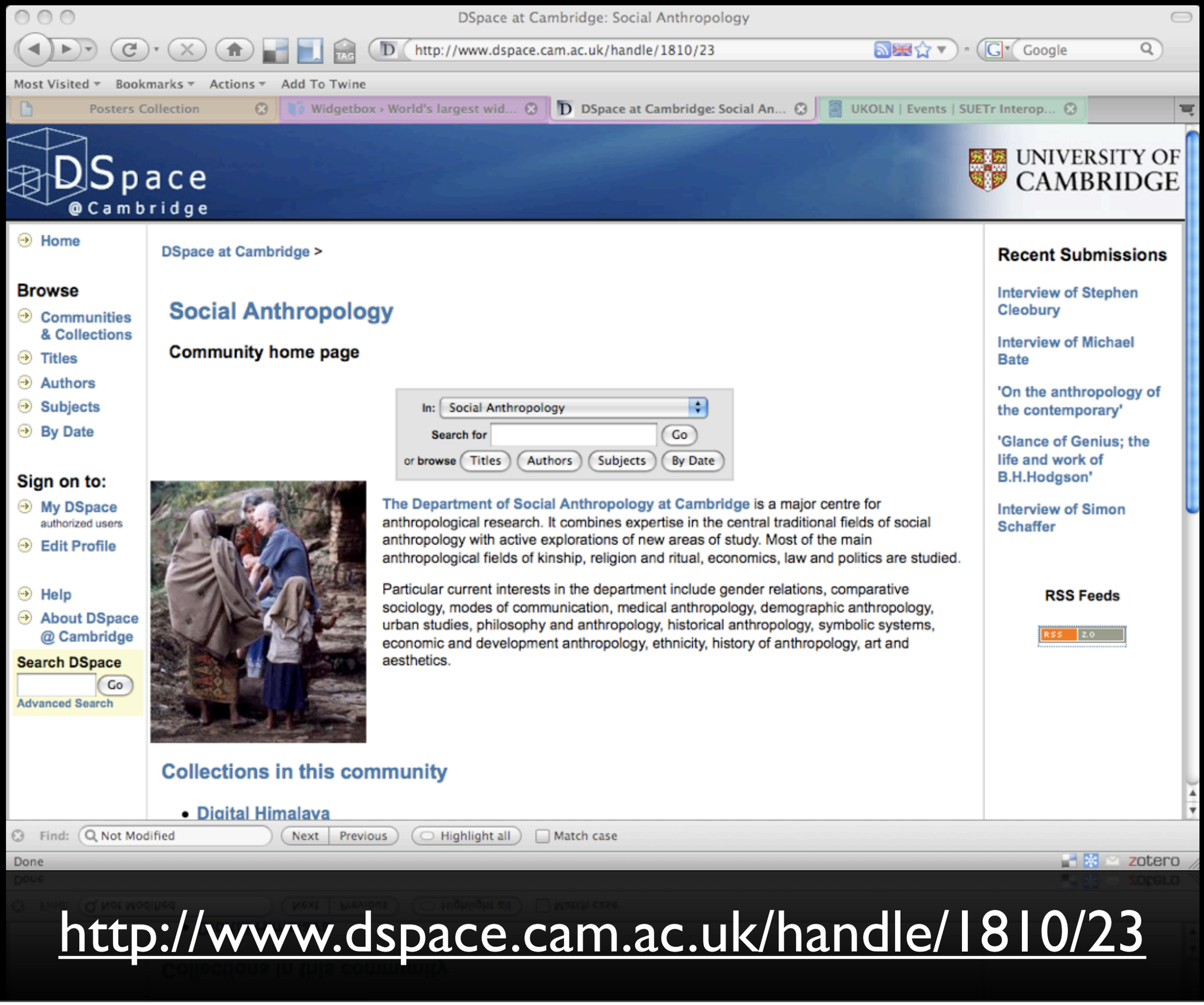
BTW - Auto-discovery of interoperability services from HTML pages is applicable to many different types of service, and is just as important there.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xh
<html dir='ltr' xmlns='http://www.w3.org/1999/xhtml' xmlns:b='http://www.google.com
<head>.
<meta content='text/html; charset=UTF-8' http-equiv='Content-Type'/>.
<meta content='true' name='MSSmartTagsPreventParsing'/>.
<meta content='blogger' name='generator'/>.
<link href='http://www.blogger.com/favicon.ico' rel='icon' type='image/vnd.microsoft.ico
.
<link rel="alternate" .
      type="application/atom+xml" .
      title="RepositoryMan - Atom" .
      href="http://repositoryman.blogspot.com/feeds/posts/default" />.
<link rel="alternate" .
      type="application/rss+xml" .
      title="RepositoryMan - RSS" .
      href="http://repositoryman.blogspot.com/feeds/posts/default?alt=rss" />.
.
<link rel="service.post" type="application/atom+xml" title="RepositoryMan - Atom" href=
<link rel="EditURI" type="application/rsd+xml" title="RSD" href="http://www.blogger.com
<link rel="me" href="http://www.blogger.com/profile/16951479417243623642" />.
<link rel="openid.server" href="http://www.blogger.com/openid-server.g" />.
<title>RepositoryMan</title>.
<style id='page-skin-1' type='text/css'><!--.
```

Consuming Atom

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

There are loads and loads of



Not just for blog content – a collection in my IR.

Google Reader (1000+)

http://www.google.com/reader/view/feed/http%3A%2F%2Fwww.dspace.c

Most Visited | Bookmarks | Actions | Add To Twine

Posters Collection | Widgetbox › World's largest wid... | Google Reader (1000+) | UKOLN | Events | SUETr Interop...

Mail | Calendar | Documents | Photos | Reader | Wet | **An update for the Google Reader offline feature is available. Refresh.** | gmail.com | Settings | My Account | Help | Sign out

Google Reader | All items | Search

Add a subscription

- Home
- **All items (1000+)**
- Starred items
- Your stuff
  - Shared items
  - Notes
  - Trends
  - Browse for stuff
- **Friends' shared items (48)**
- **Subscriptions**
  - **DC's Improbab... (1)**
  - **DSpace Community: Dig... (1)**
  - **Lifehacker (57)**
  - **O'Reilly Radar (3)**
  - **Tapestry and HiveMind (2)**
- **d (898)**
  - **CrystalEye: All Struc... (17)**
  - **Delicious DSpace (114)**
  - **dspace « Wo... (1)**
  - **James Paice's Recent ... (53)**
  - **JL Java Announcements (103)**
  - **Liberal Education Today (50)**
  - **MacOSXHints.com (59)**
  - **Science Blog - (119)**
  - **TechCrunch (382)**
- **DSpace Community: Soc... (10)**

Manage subscriptions »

DSpace Community: Social Anthropology

Show: **Expanded** - List

Show: 10 new items - **all items** | Mark all as read | Refresh | Feed settings... | show details

Interview of Stephen Cleobury

8:06 AM (seconds ago)

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

Add star | Share | Share with note | Email | Keep unread | Add tags

Interview of Michael Bate

8:06 AM (seconds ago)

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

Add star | Share | Share with note | Email | Keep unread | Add tags

'On the anthropology of the contemporary'

8:06 AM (seconds ago)

Title: 'On the anthropology of the contemporary'

↑ Previous item | Next item ↓

26 items

Done

zotero

iGoogle

http://www.google.co.uk/ig

Most Visited | Bookmarks | Actions | Add To Twine

Posters Collection | Widgetbox › World's largest wid... | iGoogle | UKOLN | Events | SUETr Interop...

Web Images Maps News Shopping Mail more ▾ jim.downing@gmail.com | Classic Home | My Account | Sign out

iGoogle™

Google Search I'm Feeling Lucky

Advanced Search
Search Preferences
Language Tools

Search: the web pages from the UK

Home Add a tab

New! Get artist themes | Change theme from Encroaching Smoke V3.0 | Add stuff »

DSpace Community: Social Anthropology

- Interview of Stephen Cleobury
- Interview of Michael Bate
- 'On the anthropology of the contemporary'

RepositoryMan

- More Repository Value for users - making PowerPoint files from RSS feeds
- Addendum to Repository Rats and other wildlife
- Repository Rats and Other Wildlife

Google Calendar

« December 2008 »

S	M	T	W	T	F	S
30	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3
4	5	6	7	8	9	10

SWORD - rwidgets.co.uk

<sword />

The **Simple Web-service Offering Repository Deposit** (SWORD) project is a lightweight protocol for deposit in the form of a profile of the Atom Publishing Protocol.

Click here to visit the official website

Click here to start using the widget

share powered by netvibes

Google Reader (1000+)

All items (1000 refresh mark all as read

Interview of Stephen Cleobury
from DSpace Community: Social Anthropology

Interview of Michael Bate
from DSpace Community: Social Anthropology

'On the anthropology of the contemporary'
from DSpace Community: Social Anthropology

Enter Location

Select a different location:
Country/Region: United States
City:

Save Cancel

Currency Conversion

Amount
From Australian dollar
To British pound Convert

Google Docs

Search recent docs

JimDowningBioSemTech2009 4 Dec
LensfieldSemTech2009Abstract 4 Dec

Transferring data from www.google.com...

zotero

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

www.dipity.com

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
2432961077_3ca5f77435_o.jpg (JPEG Picture)

DSC00711
Tue, Apr 22, 2008 11:05

Jim Downing posted a photo:

Play Events

Dec 7, 2008 3:15 PM PST
EOC/NA @ DPI Cluster Meeting (Jena, Germany)

Sep 17, 2008 1:57 AM PDT
ojd20 JISC meeting London

Sep 15, 2008 8:01 AM PDT
LHR – SEA Flight BA1053 departs Sun, 28 Sep 2008 13:55 (London time).

Sep 15, 2008 8:01 AM PDT
SEA – LHR Flight BA0048 departs Sun, 5 Oct 2008 18:45 (Seattle time).

Sep 15, 2008 8:01 AM PDT
Online Check-in opens for

Timeline | Flipbook | List | Map | Full Screen | Embed Widget | ShareThis

SHOW SOURCES

Enclosures

- Just ‘blessed’ links
- Anything can be enclosed

<link href=”http://example.com/podcasts/show001-usenglish.mpg”
rel=”enclosure”
length=”21472922”
type=”audio/mpg” />

```
 <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...1_2007sot1536_1.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_1.small.png"/>
    </div>
   </content>
 </entry>
```


http://wwmm.ch.cam.ac.uk/crystaleye/feed/atom/feed.xml

Google Reader

All items | Search

Home
All items (64)
Starred items
Trends
Your shared items
Friends' shared items
Jean-Claude Bradley
Manage friends »

Add subscription | Discover »

Show: updated - all | Refresh

- a (2)
  - Jimmy Nilsson's weblog (2)
- c (22)
  - BioMed Central (1)
  - DSpace Wiki - Recent ... (2)
  - Jon's Radio (1)
  - Lifehacker (16)
  - O'Reilly Radar (2)
- d (40)
  - CrystalEye: All Struc...
  - Delicious DSpace (2)
  - James Paice's Recent ... (1)
  - JL Java Announcements (8)
  - Liberal Education Today (2)
  - MacOSXHints.com (5)
  - Science Blog - (5)
  - TechCrunch (17)

Manage subscriptions »

CrystalEye: All Structures

Feed settings... | Expanded view | List view

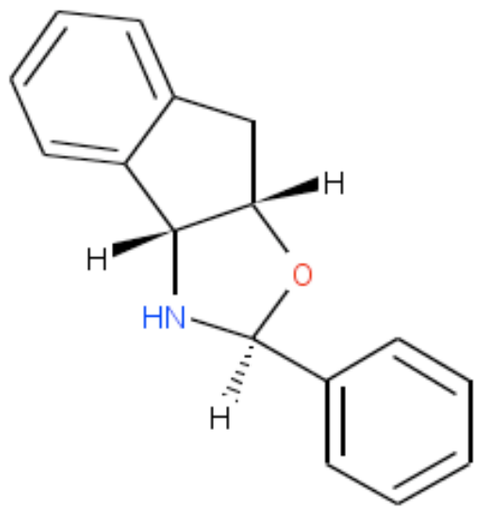
Show: 0 new items - all items | Mark all as read | Refresh | show details

Original enclosure

Add star | Share | Email | Mark as read | Edit tags: d

Summary page for crystal structure from DataBlock shelxl in CIF b719624esup1 from article b719624e in issue 2008/7 of Royal Society of Chemistry, Organic and Biomolecular Chemistry.

Mar 26, 2008 (yesterday)



Original enclosure

Add star | Share | Email | Mark as read | Edit tags: d

Summary page for crystal structure from DataBlock aba2 in CIF

Mar 26, 2008 (yesterday)

Previous item | Next item | more than 60 items

Progressive enrichment; works in simple reader for simple purposes, works in more sophisticated consumers for more sophisticated stuff.

cooliris

Determination in DC Polymeric Power
he Presence of Space Charge and
ure Gradient under dc Conditions

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

	Client	Server
First GET	GET /feed.atom	200 OK E-Tag:foobar <feed xmlns:....
No updates	GET /feed.atom If-None-Match:foobar	304 Not Modified
Updated	GET /feed.atom If-None-Match:foobar	200 OK E-Tag:barblat <feed xmlns....

Modification Date

	Client	Server
First GET	GET /feed.atom	200 OK Last-Modified:Sat, 29 Oct 1994 19:43:31 GMT <feed xmlns:...
No updates	GET /feed.atom If-Modified-Since: Sat, 29 Oct 1994 19:43:31 GMT	304 Not Modified
Updated	GET /feed.atom If-Modified-Since: Sat, 29 Oct 1994 19:43:31 GMT	200 OK Last-Modified:Sat, 30 Nov 1994 19:43:31 GMT <feed xmlns:...

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"/>
 <link rel="prev-archive"
  href="http://example.org/2003/11/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>
```

Archive Feed Page

```
<?xml version="1.0" encoding="utf-8"?>
<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"/>
 <link rel="self" href="http://example.org/2003/11/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