RSS
Syndication, pipes and web content republishing

Thanks to Erik Wilde, UC Berkeley
dret@berkeley.edu
RSS is...

• A family of XML formats used to publish frequently updated works
  – blog entries, news headlines, audio, and video

• A mechanism for content syndication and user aggregation
  – emerging business models for video & music
  – potential scholarly publishing models
Updated Content

• Early Web content was **static content**
  – or updated very infrequently
  – not a requirement to reuse content in different contexts

• Frequently **updated content** quickly became a very common scenario
  – as commercial interests took over the Web, users should have a reason to re-visit a site
  – presenting a stream of new content creates the image of a live Web site
Beyond HTML

• Two major use cases where HTML is not sufficient
  1. users want an efficient way to get the updated content from a site
  2. sites want to aggregate updated content from other sites and re-publish it

• Syndication formats are designed to support these two use cases
  – container formats for updated items
  – a small amount of metadata about these items for automated processing & styling
Aggregation Intermediaries

• User-configured portals mediate between content providers and readers
  – users select a number of feeds as their preferred information sources
  – the feeds can be presented on a single personalized Web page
  – *by sharing and re-publishing items, users are becoming aggregation intermediaries*
RSS is a way for information displayed in a multimedia web page to be communicated across the web as a list of time-stamped items.

While the stories are in list form, they can be filtered or processed in lots of ways.

The list of items (aka RSS feed) can be displayed as a table of contents, as a summary, or in bespoke browser interface or ‘widget’.
RSS History

• The Myth of RSS Compatibility provides a good overview
  – http://diveintomark.org/archives/2004/02/04/incompatible-rss
• RSS is a schoolbook example for why standards are a good thing
  – RSS 0.9 was created for the My Netscape portal in March 1999
  – RSS 0.91 (a simplification) was introduced in July 1999 (as an interim solution)
  – RSS was without an owner, and different parties claimed/denied ownership
  – RSS 1.0 was created by an informal developer group
  – RSS 0.92 (and 0.93 and 0.94) were published without acknowledging RSS 1.0
    – finally, RSS 2.0 was released as a follow-up to the RSS 0.9x versions
• Using RSS has become an exercise in managing a menagerie of versions
Basic RSS Structure

```xml
<?xml version="1.0"?>

<channel>
    ....
</channel>

<item>
    ....
</item>

<item>
    ....
</item>

<item>
    ....
</item>

....More Items
```
<rss version="0.91">
  <channel>
    <title>XML.com</title>
    <link>http://www.xml.com/</link>
    <description>XML.com features a rich mix of information and services for the XML community.</description>
    <language>en-us</language>
    <item>
      <title>Normalizing XML, Part 2</title>
      <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
      <description>In this second and final look at applying relational normalization techniques to W3C XML Schema data modeling, Will Provost discusses when not to normalize, the scope of uniqueness and the fourth and fifth normal forms.</description>
    </item>
  </channel>
</rss>
RSS 1.0

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://purl.org/rss/1.0/" ...>
  <channel rdf:about="http://www.xml.com/cs/xml/query/q/19">
    <title>XML.com</title>
    <link>http://www.xml.com/</link>
    <description>XML.com features a rich mix of information and services for the XML community.</description>
    <items>
      <rdf:Seq>
      </rdf:Seq>
    </items>
  </channel>
  <item rdf:about="http://www.xml.com/pub/a/2002/12/04/normalizing.html">
    <title>Normalizing XML, Part 2</title>
    <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
    <description>In this second and final look ....</description>
    <dc:creator>Will Provost</dc:creator>
    <dc:date>2002-12-04</dc:date>
  </item>
</rdf:RDF>
```

RDF declares a set of resources to be part of the channel.

Each item uses RDF statement syntax to give metadata about the resource. Each item is outside the channel element.
<rss version="2.0" xmlns:dc="http://purl.org/dc/elements/1.1/">
  <channel>
    <title>XML.com</title>
    <link>http://www.xml.com/</link>
    <description>XML.com features a rich mix of information and services for the XML community.</description>
    <language>en-us</language>
    <item>
      <title>Normalizing XML, Part 2</title>
      <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
      <description>In this second and final look at applying relational normalization techniques to W3C XML Schema data modeling, Will Provost discusses when not to normalize, the scope of uniqueness and the fourth and fifth normal forms.</description>
      <dc:creator>Will Provost</dc:creator>
      <dc:date>2002-12-04</dc:date>
    </item>
  </channel>
</rss>

Combination of RDF-style metadata declarations, but no RDF. Items back inside the channel.
Case for Content Management

- RSS is very rarely produced by hand
  - by definition, RSS contains redundant information for a specific purpose

- If a Content Management System (CMS) is used, RSS can be generated
  - basic metadata can be generated by the CMS (title, author, date)
  - better tagging of content results in better tagging of feeds
  - well-tagged feeds are better foundations for large-scale reuse of feed items

- Blogging is simply a specialized case of a CMS
  - Web-based interface for controlling everything
  - strictly time-ordered sequenced of published items
  - navigation features primarily based on the time-specific facets of the blog (maybe tags)
  - all blogging tools include feed support
Widgets

Widgets (or gadgets) are visual components for a web page. Some are useful, some are clever, some are just pretty.

Collection of repository items (articles, lectures, artwork etc) that you see as an HTML page are exported across the web as an RSS feed.

RSS feed:
• item 1,
• item 2,
• item 3...

A widgetmaker web site lets you design and create a widget using an RSS feed of your choice.

This widget makes a 3D gallery display from the previews of each item in my “latest publications” feed
Using Widgets on Your Site

The widget site stores your design and gives you some HTML to copy and paste into your web page to invoke the widget.

RSS feed:
- item 1,
- item 2,
- item 3...
Spruce up your own home page with a widget from widgetbox.com

My recent papers from a research repository RSS feed

Note the use of document thumbnails.

This widget was created for me by widgetbox.com using a repository feed that delivers my recent publications. I just had to paste the HTML it made into my Web page.
Recent analyses from a data repository RSS feed

Note the use of chemical structure thumbnails.

...or one from KickApps.com...

Simon Coles, Manager of the National Crystallography Centre

My role as manager is primarily to actively lead the team, by example, in the activities of collection, refinement and publication of crystal structure data. On a daily basis I am responsible for overseeing the administrative aspects of the service, ensuring the smooth running of the laboratory and interacting with our users.

I personally devise and conduct more advanced experiments than those generally provided, such as variable temperature and charge density investigations. This is part of my remit to provide more in-depth, or research-based, experiments and seek new direction for service activities, in order to maintain our position as a world-leading facility.

Latest Crystal Structures

At a higher level it is my duty to define procedure and protocol for the operation of the Service and assist in the generation of reports to our funding bodies and management panel. I also contribute to the pursuit of funding to operate the service and further its capabilities. Another strand of my work for the service is promotion, outreach and training. In this respect I regularly present aspects of service work at international conferences and author journal articles describing our novel developments that further the field of chemical crystallography. As part of our training program it is important that we use visual aids to promote better communication.

Just because a subject is scientific or technical doesn’t mean that a visual representation of its data can’t promote better communication.
Create a pagecast at PageFlakes.com

A brief bio and flattering photo

the only bit that I had to write!

My school’s latest news items RSS feed

Recent blog entries from my students’ RSS feed

My recent lectures from the teaching repository RSS feed

My recent papers from the research repository RSS feed
Create a timeline at dipity.com

My recent papers from the same research repository feed as before. Same data, new visualisation. Here, the datestamp from each article is used to place it on the timeline.
or a calendar at 30boxes.com

Recent deposits from the KULTUR project art repository RSS feed. Once again, the datestamp from each article is used to place it on the calendar. Thumbnails optional.
Display your collection on an image wall at cooliris.com

These images come from the KULTUR project art repository RSS feed. You can see them in a widget, or a special browser.
...it’s useful for viewing large collections of visual material

The image wall also works well with thumbnail images of posters from a research repository.
Variety of Ways of Using RSS

- RSS consumed directly by browser
  - Flash widget
  - AJAX
- RSS consumed by server-side cgi script to produce HTML page
  - use bespoke PHP library (e.g. splinteredfeed)
  - use XSLT
- RSS consumed by command-line XSLT to produce
  - dynamic HTML page
  - static HTML page
  - static data for incorporation into dynamic HTML

- *BEWARE security sandboxes!!*
RSS XSLT example (pt 1)

```xml
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet version="1.0"
 xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
 xmlns:xhtml="http://.../" xmlns:rdf="http://.../
 xmlns:rss="http://.../" xmlns:dc="http://.../">
 <xsl:output method="html"/>

 <xsl:template match="rdf:RDF">
  <ul id="deliciouslist">
   <xsl:apply-templates select="rss:item"/>
  </ul>
 </xsl:template>
</xsl:stylesheet>
```

Turn the RDF element into an unordered list...

Download complete example from http://www.splintered.co.uk/experiments/69/
RSS XSLT example (pt 2)

<xsl:template match="rss:item">
  <li>
    <a>
      <xsl:attribute name="href">
        <xsl:value-of select="rss:link" />
      </xsl:attribute>
      <xsl:value-of select="rss:title" />
    </a>
  </li>
</xsl:template>

…and turn the each item into an <li> element containing a link to the item’s URL.
Delicious Bookmark List

This is the page that we’re going to get a feed from...
As an RSS Feed

<table>
<thead>
<tr>
<th>Delicious/splintered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biscuit Tin - Random is good</td>
</tr>
<tr>
<td>The next President of the United States - The Big Picture - Boston.com</td>
</tr>
<tr>
<td>Callie Shell - Obama - Digital Journalist</td>
</tr>
<tr>
<td>YouTube - YouTube Captions and Subtitles</td>
</tr>
<tr>
<td>mindforks: Using .attr for ARIA in jQuery UI</td>
</tr>
<tr>
<td>WordPress » Youtuber « WordPress Plugins</td>
</tr>
<tr>
<td>Audio Player Wordpress plugin - 1 Pixel Out</td>
</tr>
<tr>
<td>Twitter API Wiki / FrontPage</td>
</tr>
<tr>
<td>The Google Maps Image Cutter - UCL Centre for Advanced Spatial Analysis</td>
</tr>
<tr>
<td>QuirksBlog: Slides from accessible Ajax workshop at Fundamentos Web</td>
</tr>
<tr>
<td>northtemple - JavaScript and screen readers</td>
</tr>
<tr>
<td>Vintage and Retro Typography Showcase I Inspiration I Smashing Magazine</td>
</tr>
<tr>
<td>Rubitone (Rubik + Pantone)</td>
</tr>
<tr>
<td>China Channel Firefox Add-on - Experience the censored Chinese internet at home!</td>
</tr>
<tr>
<td>Hack that Phone - Make and use themes with WinterBoard - Section 2 ...</td>
</tr>
</tbody>
</table>

...this is what the feed looks like in an RSS reader...
Processed with SplinteredFeed

• As an HTML document
• Not much improvement – but it’s just a start 😊

...and this is the result of the transformation.
Yahoo Pipes: sed for RSS
Yahoo Pipes: sed for RSS

1. Take THREE independent blog feeds that you want to mash together

2. Prepend each item’s title with the name of its blog

3. Mash them all up together (union + sort by time)

4. So you end up with a combined blog feed, whose items can still be distinguished. E.g. tech, science & political news.
Result: Pipes used in widgets

Widgets showing latest news from...

Repositories Worldwide

...different open source products...

...twelve independent bloggers...

...various techno-political issues
Not So Simple Syndication

- Syndication is a powerful concept
- RSS versions make the RSS life harder than necessary
- Current RSS versions are stable, but many extensions exist
- Eventually RSS will be replaced by Atom
Suggested Exercises

• Make a Web page from a BBC News feed using XSLT
• Make an HTML fragment as above, but embed it into another news page using <iframe> in the HTML, AJAX on the client browser or PHP include on the server.
• Create a KickApps carousel widget from a Flickr RSS feed and embed it in a web page
• Mash several news feeds together using Yahoo Pipes, filtering on “Global Warming” or “Climate Change”
• Generate your own RSS feed from e.g. your lecture timetable.