ALL FEEDS LEAD TO ROME

RSS 1.0
RSS 0.90
RSS 0.92
RSS 2.0
ATOM

ROMA LOCUTA EST. CAUSA FINITA EST
ROME HAS SPOKEN. THE CAUSE IS FINISHED

Rss and atOM utilitiEs ROME

ESCAPE syndication hell, a java developer perspective

Patrick Chanezon, Developer, Sun
Alejandro Abdelnur , Developer, Sun

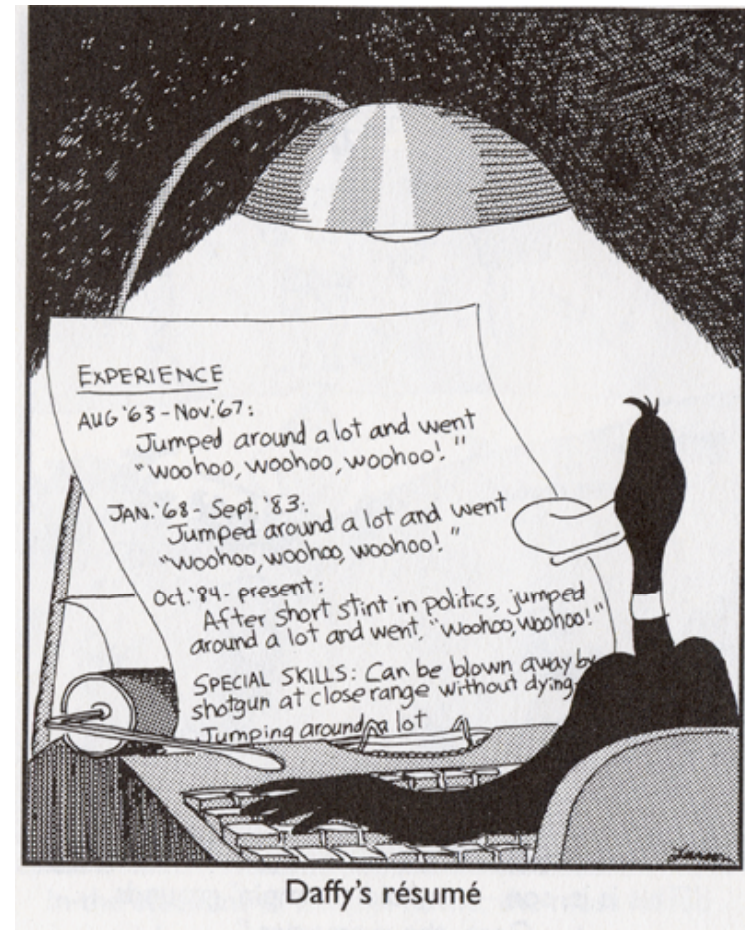ROME-XML

ROME-XML
ROME-XML

ROME-XML

Agenda

- Introduction: syndication revolution
- What is ROME
- The ROME library
- Fetcher
- Modules
- Plans and usage

P@ in a slide

- France - work remotely
- MyNetscape - Sun Portal
- Blogs.sun.com
- RSS 1999 - Weblogs 2002
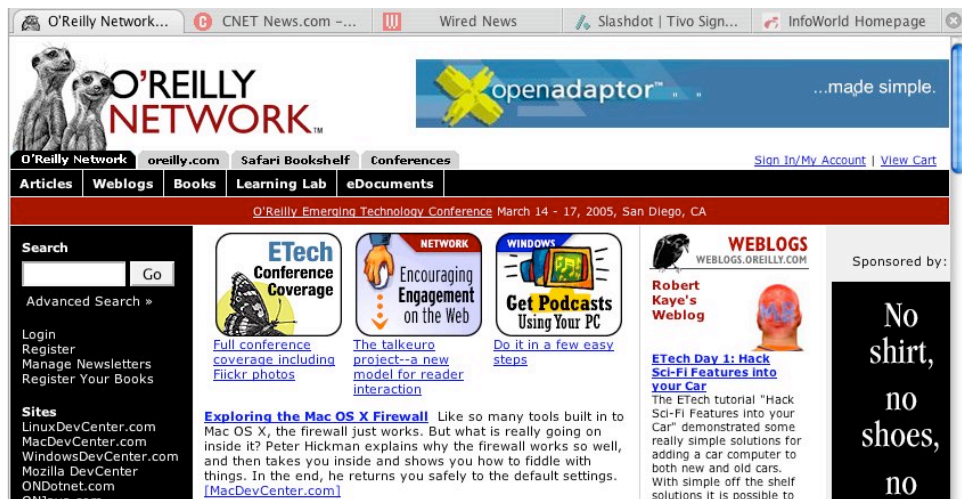- Fascinated by syndication revolution

Daffy's résumé

- As disruptive as the web: potential to change the ways we work and play
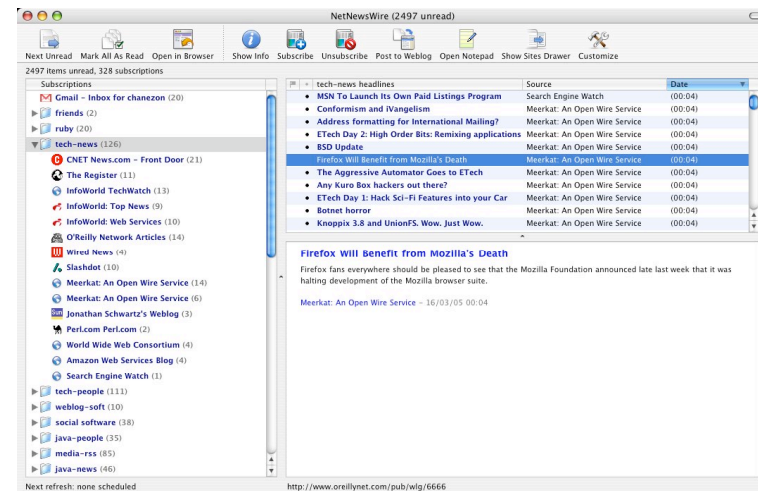- The products are actually useful to me

What is syndication?

- Have a client scan what's new on a server
- Simple XML format to make it easier on the client: avoid screenscraping.

From 10 sites a day

To 328 feeds

Syndication revolution

- Blogs: 1 year from 2 to 7 millions

Blogs Tracked by Technorati

**Weblogs Tracked
March 2003 - February 2005**

Doubling in size approx.
every 5 months

Technorati
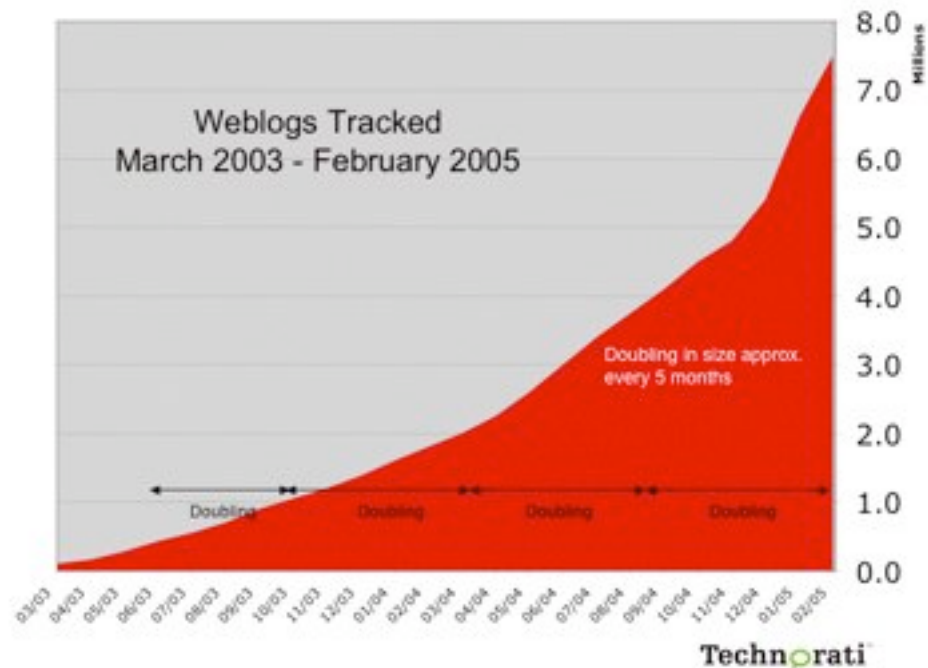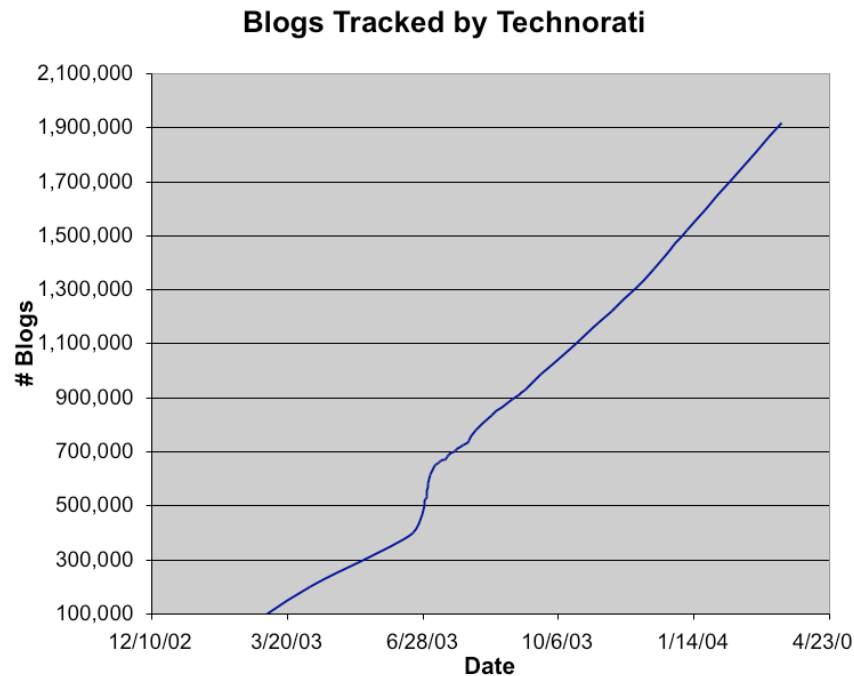
- Blogs are just the tip of the syndication iceberg:
  - podcasting, video, banks, news, business processes

5

Syndication application typology

Producer

Blog Clients
Marsedit, Bloged,
w:bloggar

Blog Servers
MT, WordPress, Roller
Online Services
Flickr, Delicious, Netflix,
Amazon, xWiki, Gmail

Client

Desktop
Syndication
Suites
Radio, NNW+Marsedit

Server
Syndication
Engines
Merkaat, Feedster,
Feedburner, AllConsuming

Server

Client Side Readers
NetNewsWire, NewsGator,
Straw, Jnn, Radio

Server Side Readers
Bloglines, Rojo, Feedtagger
Portals
MyNetscape, MyYahoo,...

Browsers, mail clients
FireFox, Safari RSS,
Thunderbird

Consumer

What it looks like

```
<?xml version="1.0" encoding="UTF-8"?>
<feed version="0.3" xmlns="http://purl.org/atom/ns#">
   <title>Gmail - Inbox for chanezon</title>
   <link rel="alternate" href="http://gmail.google.com/gmail"
   type="text/html"/>
   <modified>2005-0315T18:25:42Z</modified>
<entry>
   <title>RubyGems, RPA and best tool for the job</title>
   <summary>On Tue, 15 Mar 2005 22:44:44 +0900, Mauricio
   Fernández &amp;lt; …</summary>
   <link rel="alternate" href="http://gmail.google.com/gmail"
   type="text/html"/>
   <modified>2005-03-15T18:24:32Z</modified>
   <issued>2005-0315T18:24:32Z</issued>
   <id>tag:gmail.google.com,2004:1164874688543396343</id>
   …
</entry>
```

Syndication formats

- 9 formats, depending how you count
- 1998 CDF, iSyndicate, scriptingNews
- 1999, Netscape, 0.90 (RDF), 0.91(XML)
- Userland, 0.91, 0.92, 0.93, 0.94, 2.0
- Rss-dev: RSS 1.0(RDF), recently 1.1
- Personality issues, syndication war: 1.0/2.0 fork
- Atom, IETF
  - Atom 0.3, now 0.5-"do-not-use-in-production"

Parser Libraries

- Python, Universal Feed Parser, by Mark Pilgrim
- Java
  - ROME **ROME**
  - Apache Jakarta Commons Feedparser (from Rojo folks)
  - We're discussing merging the 2 projects
  - Common JavaOne talk in June
- Perl, haven't looked but I'm sure TMTOWTDI
- Ruby
  - Looked incomplete, still young
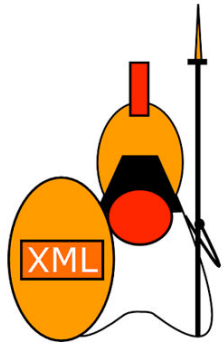
Why ROME?

"Show me your code and conceal your data structures, and I shall continue to be mystified. Show me your data structures, and I won't usually need your code; it'll be obvious."
Fred Brooks, "The Mythical Man Month"

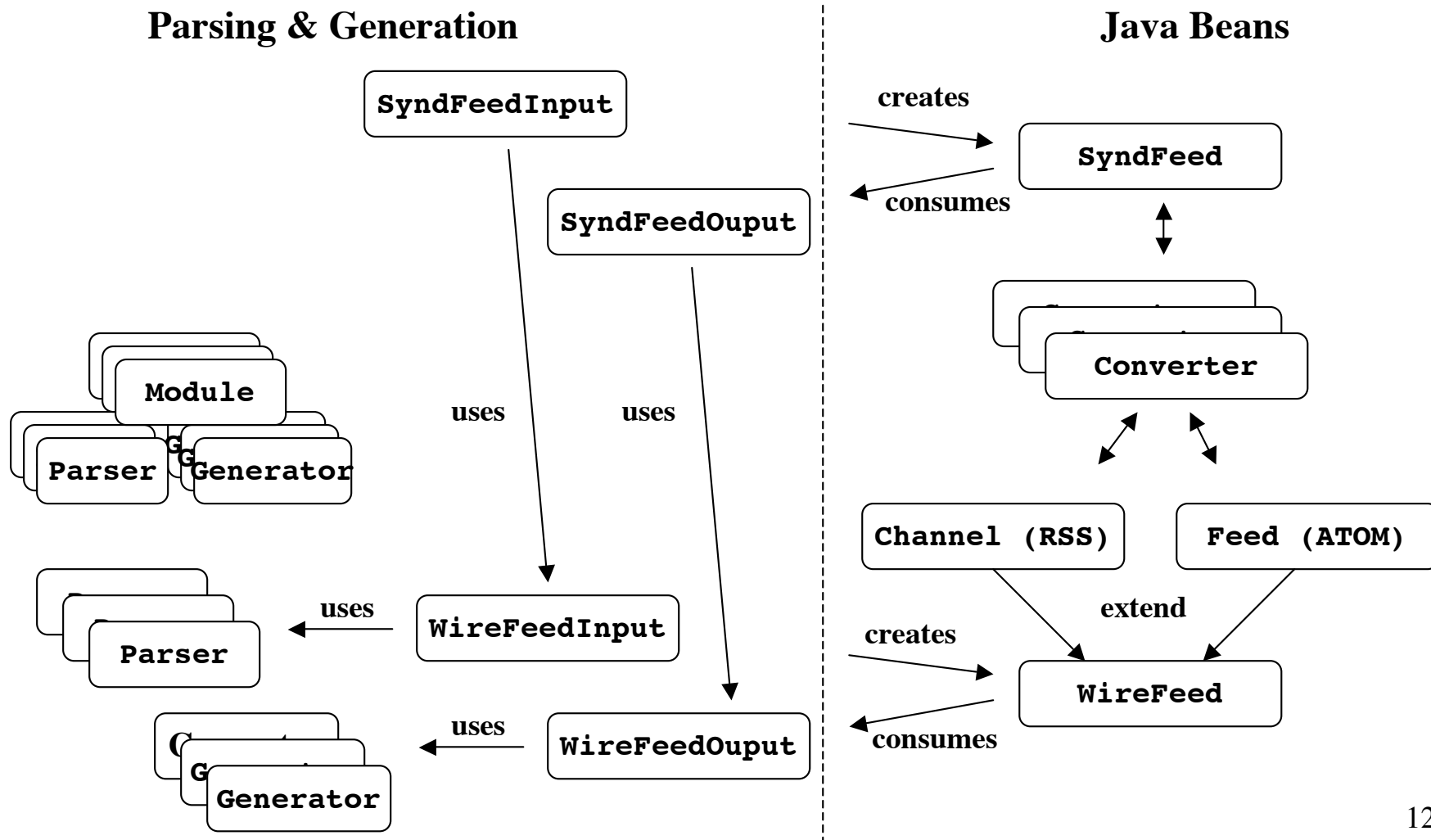- Obvious? Feeds versions are highly inconsistent
- Need for an Easy, Simple, Complete, Abstract, Powerful and Extensible way of handling feeds in java
- Existing solutions were incomplete, stale, had an unfriendly Java API

What is ROME?
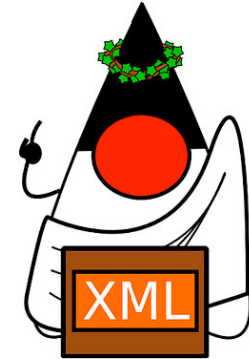
- A Java library
- RSSs/ATOM feeds parsers and generators
- Homogeneous Java representation of feeds
- Built on top of Java Collections and JDom
- Within Java feeds are Java Beans

How does ROME work?

Parsing & Generation

Java Beans

SyndFeedInput

SyndFeedOuput

Module

Parser | Generator

uses   uses

Parser

uses

WireFeedInput

Generator

uses

WireFeedOuput

creates

SyndFeed

consumes

Converter

Channel (RSS)   Feed (ATOM)

extend

creates

WireFeed

consumes

12

Sample Code
Converting any feed to RSS 1.0

```
Reader reader = ...
Writer writer = ...

SyndFeedInput input = new SyndFeedInput();
SyndFeed feed = input.build(reader);


feed.setFeedType("rss_1.0");


SyndFeedOutput output = new SyndFeedOutput();
output.output(feed,writer);
```
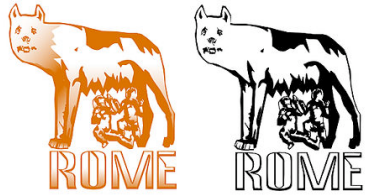
Pluggability

- Parsers and generators per feed type
- Module parsers and generators per feed type
- Converters
- ROME looks for `/rome.properties` in all classpath entries and aggregates them
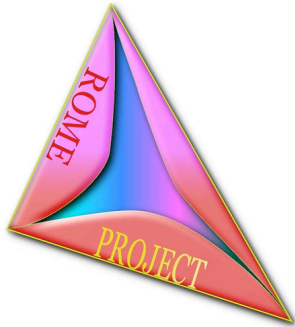- Easy to replace or reconfigure parsers, generators and converters

XML Goodies

- XmlReader
  - Deals with charset encoding inconsistencies
- XmlHealer
  - Trims XML streams
  - Resolves HTML entities

ROME Pros

- Simple to use
- Single Java representation of feeds
- Lean (implementation and dependencies)
- Fully pluggable
- Well documented
- Widely used, got momentum
- It's Java

ROME Cons

- Loss of information at SyndFeed level: intersection, not union

- Loss of information at conversion

- Potential DOM overhead (could be minimized with SAX/StAX based parsers)

Roman Numbers (03/2005)

- Age: **IX** months
- Members: **XII** developers, **XVI** users
- Mailing list: ~ **C** per month
- Project hits: ~ **M$\overline{\text{V}}$** per month
- Public API: **XV** interfaces, **XXXII** classes
- JAR size: **CLX** kilobytes

Technical issues when dealing with feeds

- Liberal vs. strict parsers (strict against what?-)
- Inefficient client/server protocol
  - Clients GET a url periodically
  - Server sometimes dynamically generate feed
  - How to save processing and bandwidth on both sides?
- Application level
  - RSS 2.0 attribute, /. Banning item
- Protocol level
  - Conditional GET and RFC3229 Delta Encoding feed oriented IM methods

ROME Fetcher

- Done by Nick Lothian
- HTTP fetching
- GZip
- Etags & lastModified
- Delta encoding support
- Hook for cache

Conditional GET

- Changed between HTTP 1.0 and 1.1
  - If-Modified-Since and Last-Modified
  - If-None-Match and Etag
- Details at
  http://fishbowl.pastiche.org/2002/10/21/http_conditional_get_for_rss_hackers

RFC3229 Delta Encoding

- Bob Wyman (PubSub) proposal after Microsoft MSDN server collapsed last summer
- Define 2 new instance-manipulation methods for feeds: feed and f-range

| GET /atom.xml HTTP/1.1 Host: bar.example.net If−None−Match: "321" A−IM: feed, gzip	HTTP/1.1 226 IM Used ETag: "4321" IM: feed, gzip Date: Tue, 13 Sep 2004… Cache−Control: no−store, im	GET /atom.xml HTTP/1.1 Host: bar.example.net If−None−Match: "321" A−IM: feed, f−range, gzip F−range: entries=1−20	HTTP/1.1 226 IM Used ETag: "4321" IM: feed, f−range, gzip Date: Tue, 13 Sep 2004 18:30:05 GMT Cache−Control: no−store
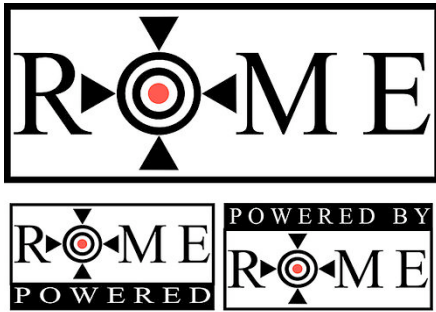
Syndication extensibility

- Namespaced modules for the various formats: dublin core, prism, etc

- People start to play with Data blogging, or structured blogging (Bob Wyman, Joe Reger, Russell Beattie)

- Will we see a similar figth than for SOAP doc/literal versus rpc/encoded?

From content to applications

MyNetscape in 1999
Invent RSS to aggregate more content in its Portal

A9 march 2005
Leverages RSS to aggregate more search results in its… Portal:-)
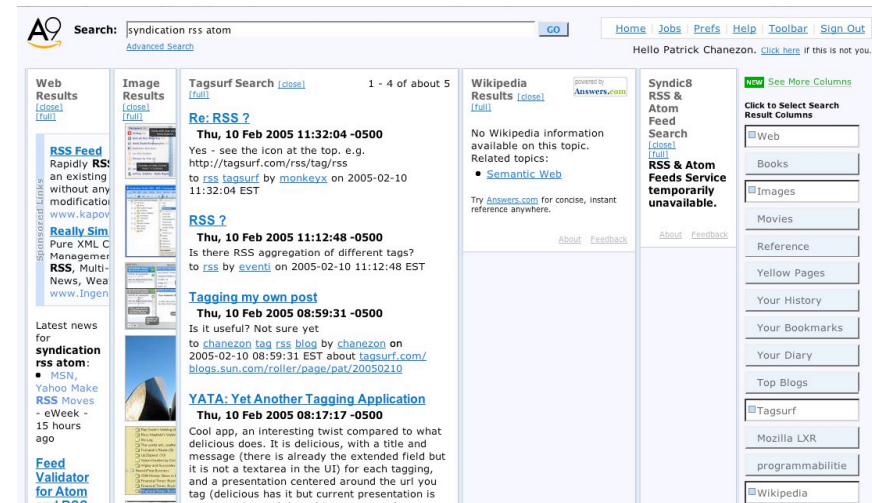
Tomorrow? Syndication formats becomes REST web services main envelope format: UBL may be embedded in Atom more than in SOAP!

Syndication as the new envelope?

- REST vs SOAP(WS Stack) war
- Amazon APIs, proposed SOAP and REST
  - 70% developers use REST
  - What is important is the schema
- Atom could become the new envelope format for REST based WS, using industry standard schemas
  - When will we first see UBL embedded in an Atom feed?
- Example: Joe Reger data blogging proposal. Embedding arbitrary meta data in his site's feeds, for ex: triathlon related. More like rpc encoding.
- Bob Wyman's structured blogging proposal. Based on schemas.

ROME Modules

- ROME was designed to handle namespaced extensions

- We let you plug in your own namespace parser to handle them, return some bean of your own

- By default we handle: dc, syndication

- Plans: prism, yahoo media, a9 opensearch, UBL (just for the fun of it)

Other Sub-Projects

R✹ME

Mark Preston
5257 Silverthorne rd.
Westerville, Ohio 43081
music2myears@insight.rr.com
614-598-0583

Greetings

My thinking with this design came from your summery: Rome, Atom, and Java. Not only did I
want to incorrporate the utilities, but show the strength in Rome. When I see this logo I think
POWER. Thank you for your consideration. Mark

Powered By R✹ME   **Powered By R✹ME**   **R✹ME**

Powered By R✹ME   **Powered By R✹ME**   **R✹ME**

Powered By R✹ME   **Powered By R✹ME**   **R✹ME**

- # Aqueduct (in progress)
  - ## Persistent API for SyndFeeds
  - ## Pluggable implementation (Hibernate & Castor)
- # Modules (in progress)
 - ## Mod_Prism and more

Plans

- Dave Johnson, new apache umbrella project, blogtools
  - Roller
  - Roller Planet (ROME based aggregator)
  - Atom client and server API impl based on ROME
- Kevin Burton, FeedParser
- Common JavaOne Talk in june
- Discuss if we merge all projects
  - Feedparser events based ROME parsers
  - Common http client library

Who is using ROME?

- Sun
- xWiki, SnipSnap
- Roller
- ScheduleWorld
- FeedPod (text2voice, adds mp3 enclosures)
- parss, from the antville guys
- Public Interactive
- Blog-City
- Reger.com
- abclinuxu.cz
- CNET

ALL FEEDS LEAD TO ROME
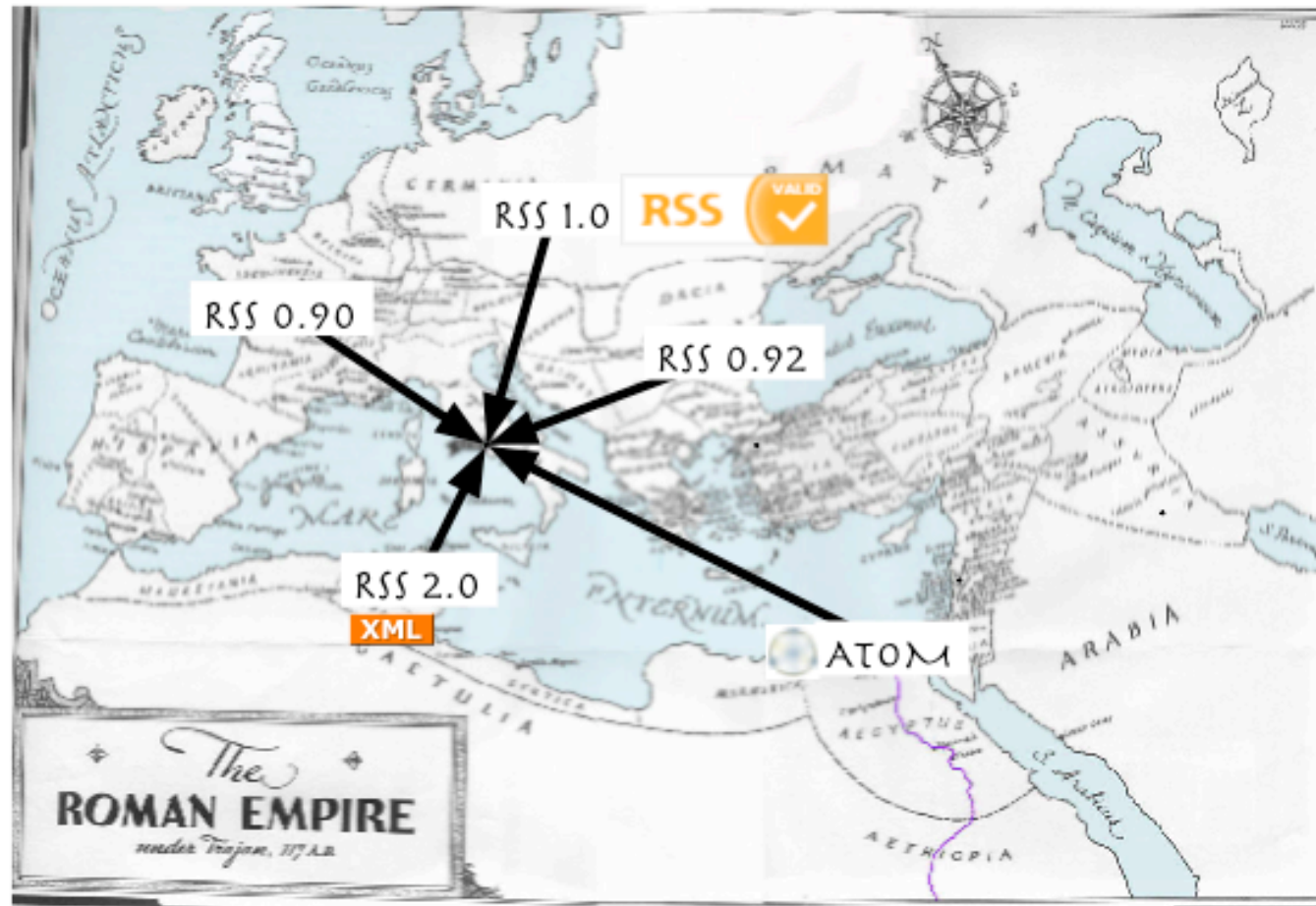
ROMA LOCUTA EST. CAUSA FINITA EST
ROME HAS SPOKEN. THE CAUSE IS FINISHED

Questions?

- We have a logo contest: which logo did you prefer?

- Vote on http://wiki.java.net/bin/view/Javawsxml/RomeLogoContest

Backup Slides

Syndication and search

- Issues
  - Eliminate duplicates: feedster not very good at that
    - Continuum from journalist who blogs to linkblogs
    - Ripped, mixed and burnt feeds will create more and more duplicates
- Opportunities
  - Search results provided as feeds
    - like a search in the future
  - Search aggregation through RSS
    - A9's new OpenSearch spec
  - Search within feeds
    - Already structured metadata
  - Subs list + social net + tagging: use algorithms (Bayesian, associations)