

Web Content Syndication

Juha Hautakangas

GFIS Developer

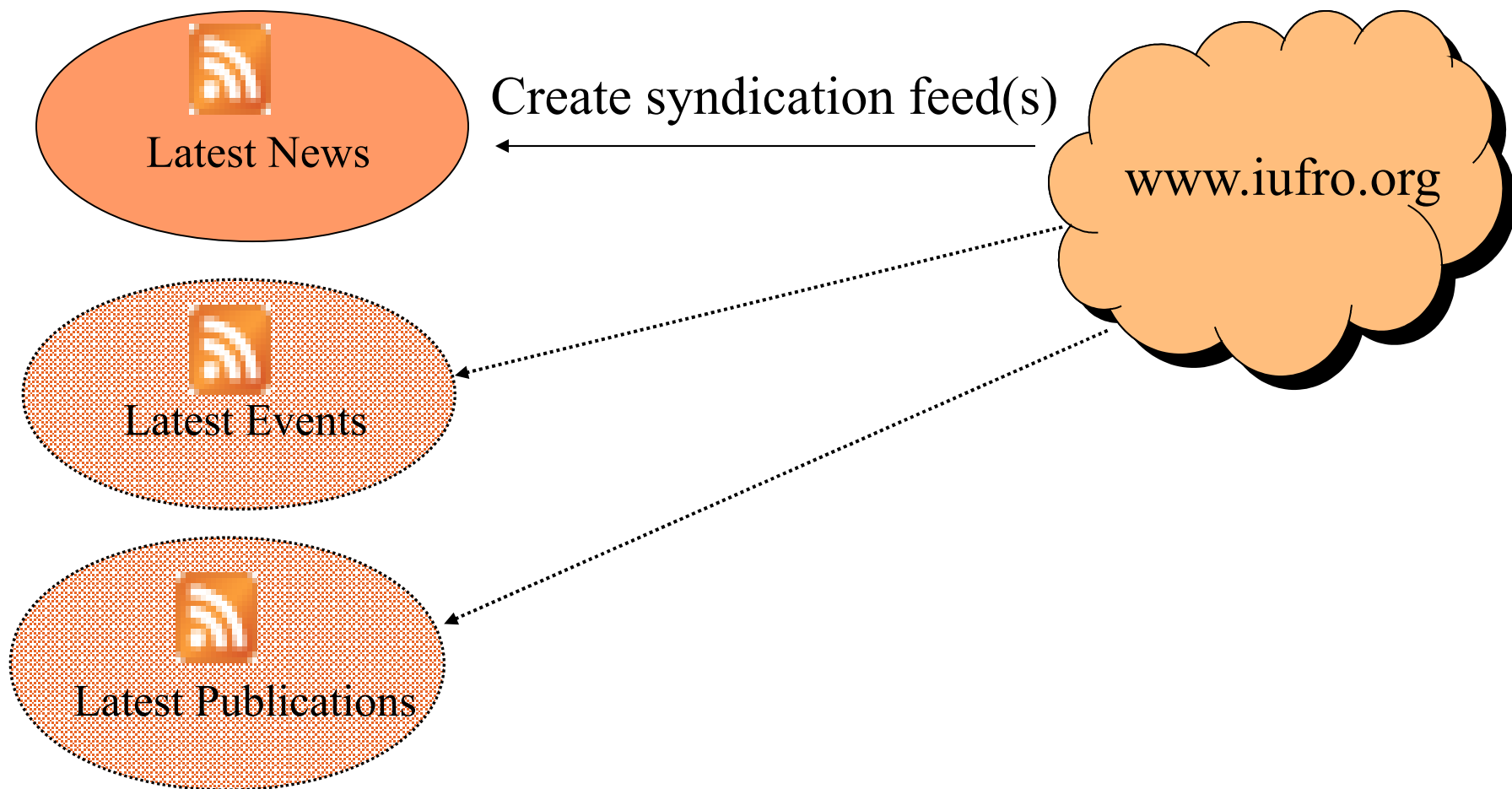
International Union of Forest Research Organizations

hautakangas@iufro.org

Syndication

- Making web feeds available from a website to provide users with a summary of the website's recently added / updated content
- Syndication feeds = Web feeds = RSS feeds = Atom feeds = XML feeds

Syndication

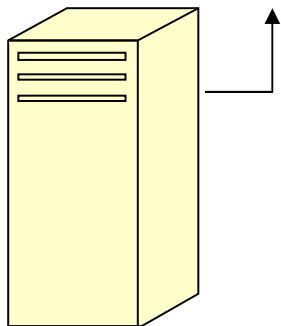


Subscription

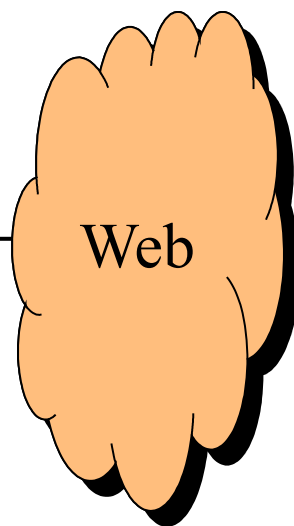
- To use a feed you need to *subscribe* to it

RSS/XML feed

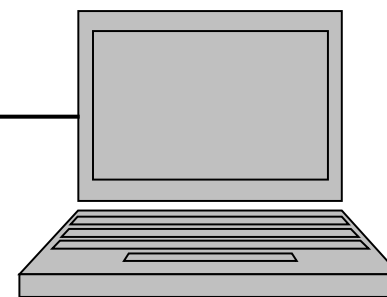
```
.....  
<item>  
  <title>Science Initiative webpages launched</title>  
  <link>http://www.iufro.org/news/article/181/</link>  
  <description>Read about this IUFRO led CPF initiative on  
    the adaptation of forests to climate change.</description>  
  <pubDate>Wed, 06 Feb 2008 14:08:00 +0100</pubDate>  
</item>  
.....
```



<http://www.iufro.org/>

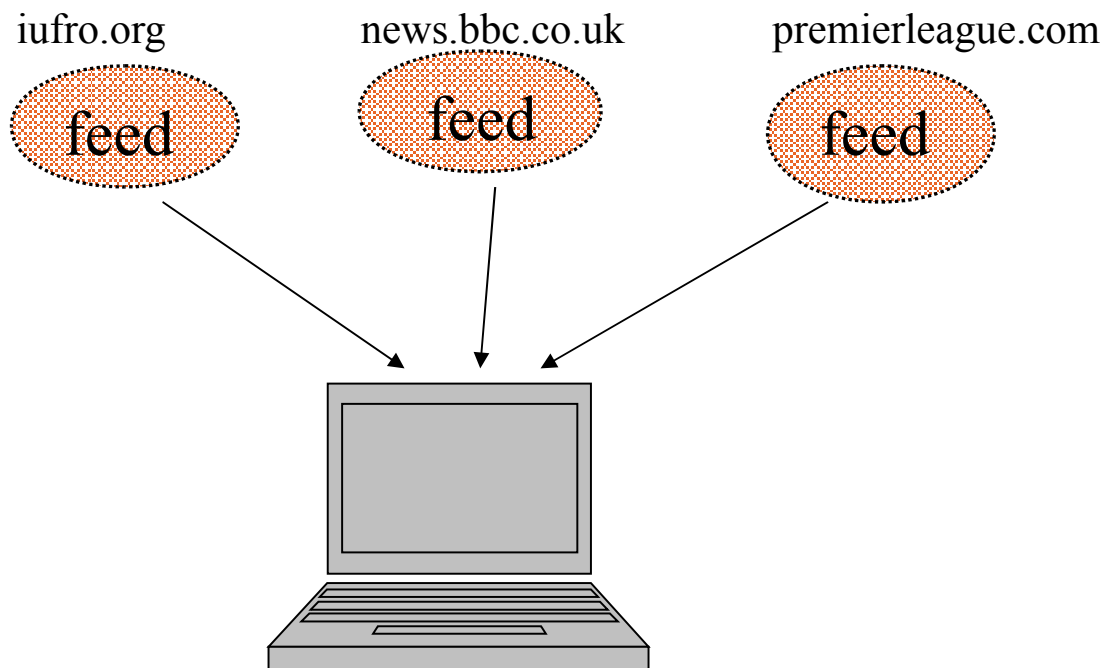


Subscribe to a feed



Aggregation

- Combining multiple feeds into single location for easy viewing



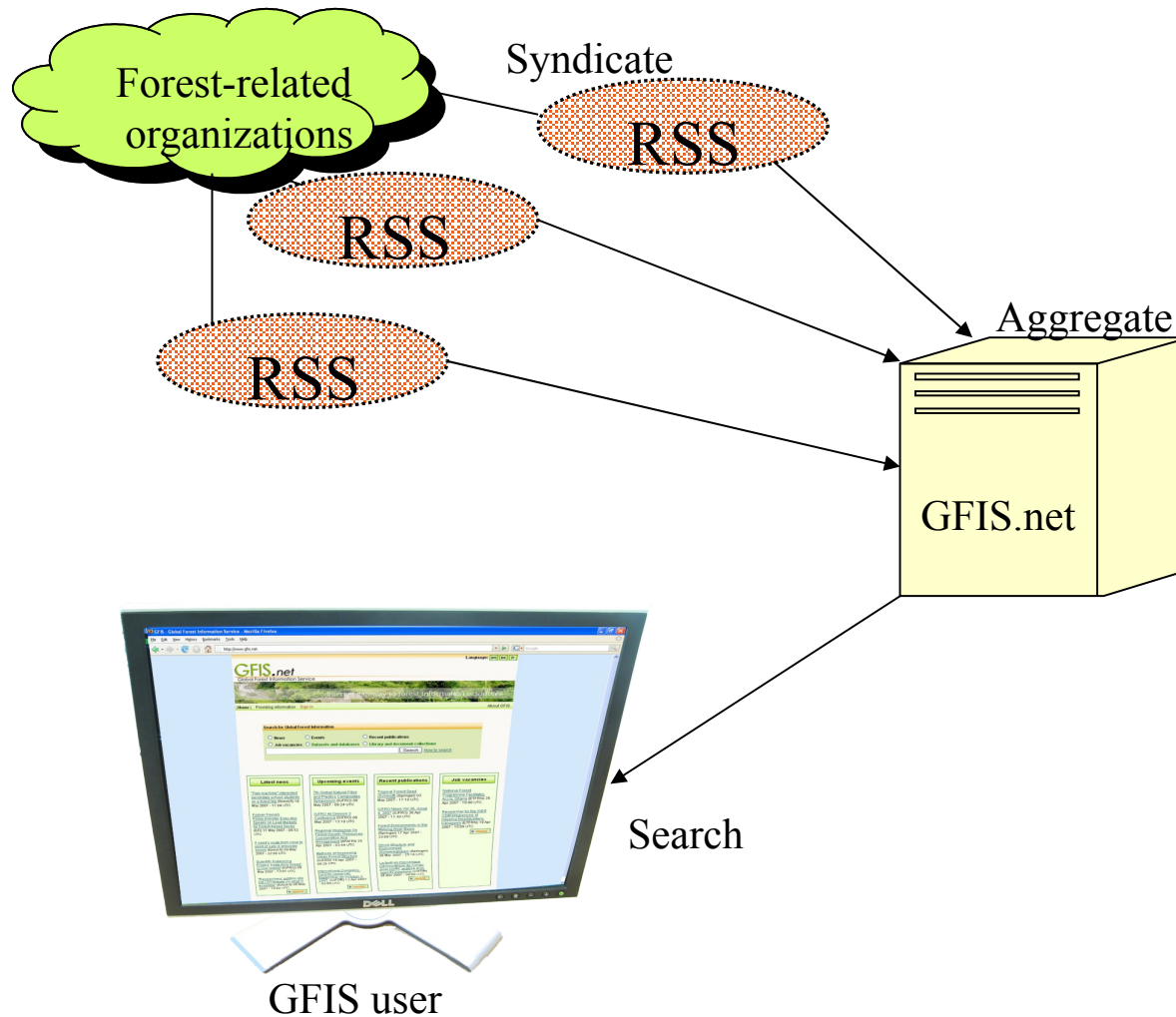
Aggregator applications

- Modern web browsers
 - Firefox / Safari / Opera / Internet Explorer 7
- Specialized feed readers
 - Web / Desktop-aggregators
 - A lot of free and commercial software exist

GFIS is an Aggregator

- GFIS reads RSS feeds maintained by forest-related organizations on their forest-related web resources
- Provides users with tools to search/browse information in the feeds
- <http://www.gfis.net/>

Syndication and GFIS



RSS / XML

- RSS is a dialect of XML
- An RSS feed (a channel) is a container for items
- An item represents a summary of a news, an event or some other kind of resource

RSS 2.0

- A simple, easy to understand format which can be used to publish frequently updated content
- RSS feed is an XML file that can be read by special feed reading software or a web service
- Creating RSS feeds is not difficult but there are certain rules you have to follow

The XML Declaration

- `<?xml version="1.0" encoding="UTF-8" ?>`
- Processing instruction that identifies the document as being XML
- Essential information about the syntax
 - most importantly: the character set encoding

RSS: Channel

- RSS document contains
 - channel (feed)
 - title – title of the channel (e.g. "IUFRO news")
 - link – URL to the website corresponding to the channel (e.g. "http://www.iufro.org/")
 - description – description of the channel (e.g. "Latest news from IUFRO")

RSS Example: Channel

- **A channel is a container for a number of items**

```
<channel>
```

```
  <title>IUFRO News</title>
```

```
  <link>http://www.iufro.org/ </link>
```

```
  <description>Latest news from IUFRO</description>
```

```
  <item>.....</item>
```

```
  <item>.....</item>
```

```
  <item>.....</item>
```

```
  ...
```

```
</channel>
```

RSS: Item

- Single resource (e.g. a news)
- Contains descriptive fields such as *title*, *link* and *description*
- Acceptable fields described in the RSS 2.0 Specification
 - *Extendable using namespaces*

RSS Example: Item

<item>

<title>Opportunities and risks of wood energy production**</title>**

<link><http://www.fao.org/newsroom/en/news/2007/1000709/index.html>**</link>**

<description>

The use of wood energy can help reduce greenhouse gas emissions and contribute to poverty reduction. But the use of wood for fuel can also result in deforestation or forest degradation if forests are not managed sustainably.

</description>

<pubDate>Mon, 19 Nov 2007 23:00:00 GMT**</pubDate>**

</item>

Extending RSS

- RSS 2.0 is extendable
 - "A feed may contain elements and attributes not described in the RSS 2.0 Specification, only if those elements and attributes are defined in a namespace."
- You can use existing namespaces or define one yourself

GFIS Information Types

- GFIS information types:


- News
- Events
- Recent publications
- Job vacancies



RSS

- Each information type has a bit different content inside the <item>

GFIS Information Types

- GFIS information types:
 - News
 - Events
 - Recent publications
 - Job vacancies
-  RSS
- Library and document collections (FAO AGRIS)
 - Datasets and databases (USGS/NBII Clearinghouse)
 - Educational resources and Experts coming as new information types
- Each information type has a bit different content inside the <item>

News

- Any information about recent events or happenings in the field of international forestry
- For example information about forest resources, their users and institutions that are involved in forestry

Example of a GFIS News Feed

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel>
    <title>The title of the channel</title>
    <link>The URL of the organization</link>
    <description>The description of the channel</description>
    <item>
      <title>The title of the news</title>
      <link>The URL of the news</link>
      <description>The description of the news</description>
      <pubDate>Tue, 13 Jun 2006 16:00:01 +0200</pubDate>
    </item>
  </channel>
</rss>
```

Events

- For example exhibitions, online meetings, conferences and workshops
- Not counted as an event: Press-conferences, inaugurations or launching activities, courses and internal meetings
- Metadata for an event provides descriptive information that serves as the basis for the discovery of the purpose, location, duration, responsible agents, links to related meetings and resources of the meeting

Example of a GFIS Events Feed (Item)

```
<item>
  <title>The title of the event</title>
  <link>The URL of the event</link>
  <description>The description of the event</description>
  <pubDate>Tue, 13 Jun 2006 16:00:01 +0200</pubDate>
  <ags:dateStart xsi:type="dcterms:W3CDTF">
    YYYY-MM-DD
  </ags:dateStart>
  <ags:dateEnd xsi:type="dcterms:W3CDTF">Y
    YY-MM DD
  </ags:dateEnd>
  <ags:location>
    <ags:locationCity>Location city of the event</ags:locationCity>
    <ags:locationCountry xsi:type="dcterms:ISO3166">
      Country of the event (A three-letter country code)
    </ags:locationCountry>
  </ags:location>
</item>
```

Recent Publications

- Recent publications from organizations
- For example forestry journals and policy documents

Example of a GFIS Publications Feed (Item)

<item>

<title>**The title of the publication**</title>

<link>**The URL of the publication**</link>

<description>**The description of the publication**</description>

<pubDate>**Tue, 13 Jun 2006 16:00:01 +0200**</pubDate>

<dc:creator>

The party responsible for the publication

</dc:creator>

<dc:subject>

Describe the content of the resource using keywords

</dc:subject>

</item>

Job Vacancies

- A brief description of a post with specific terms of reference within the information provider institution
- A job vacancy announcement includes details of the closing date for applications

Example of a GFIS Job Vacancies Feed (Item)

```
<item>  
  <title>The title of the job vacancy</title>  
  <link>The URL of the job vacancy</link>  
  <description>  
    The description of the job vacancy  
  </description>  
  <pubDate>Tue, 13 Jun 2006 16:00:01 +0200</pubDate>  
  <ags:dateEnd xsi:type="dcterms:W3CDTF">  
    YYYY-MM-DD  
  </ags:dateEnd>  
</item>
```

Datasets and Databases

- Datasets
 - Information encoded in a defined structure (for example, lists, tables, databases), intended to be useful for direct machine processing
- Databases
 - Collections of related information, organized and presented to serve a specific purpose, and which allow for rapid query and retrieval
- Stored in the USGS / NBII Metadata Clearinghouse (information system)

Library and Document Collections

- Books and continuing resources ('serials', for example periodicals, newspapers, yearbooks, etc.)
- Can be print publications or electronic publications
- Stored in the FAO AGRIS information system

Check Your Creation: Validate

- RSS / XML feeds must be well-formed and valid
 - Well-formedness: feed conforms to all syntax rules of XML
 - Validity: feed also conforms to some semantic rules
 - RSS Specification
 - GFIS Application Profile
- You can validate your RSS feeds at:
 - <http://validator.w3.org/feed/>
 - <http://feedvalidator.org/>

Why Syndicate Your Content?

- Syndication is a free and easy way to advertise your online information
 - Syndication vs. publishing on a single website only
- Get more visitors
 - Notifying users when your website has new content pulls them on Your website to check out the new content
- Reduce the amount of disappointed users
 - Users get disappointed when they visit a website and find out it hasn't been updated since their last visit

Why Syndicate Your Content?

- Very little resources required
 - Some technical effort in the beginning (depending on the infrastructure of your existing website)
 - Database-driven content (CMS / Dynamic scripts) is the best/easiest option (maintained automatically)
- Feeds are used by other websites and people directly – they are not for GFIS only

XML Namespaces?

- Used for providing uniquely named elements and attributes in an XML file
- Defined by W3C recommendation: Namespaces in XML
- An XML file may contain elements or attributes from multiple namespaces

Existing Namespaces

- Many existing namespaces can be used to extend RSS, for example:
- Dublin Core Metadata Element Set (DCMES)
 - A generic vocabulary for describing a wide range of resources
 - `xmlns:dc="http://purl.org/dc/elements/1.1/"`
- Agricultural Metadata Element Set (AgMES)
 - For describing agricultural information resources, developed by FAO
 - `xmlns:ags="http://purl.org/agmes/1.1/"`

- An element set for describing a wide range of networked resources
- The elements provide means to describe documents, images, sound files, and other networked information objects
- Consists of 15 metadata elements:
 - title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights

- An element set for describing a wide range of networked resources
- The elements provide means to describe documents, images, sound files, and other networked information objects
- Consists of 15 metadata elements:
 - title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights

The Agricultural Metadata Element Set

- Developed by Food and Agriculture Organization of the United Nations for describing agricultural information resources

DCMES and AgMES Elements in the GFIS RSS Feeds

- GFIS extends RSS using DCMES and AgMES to define its application profile
- DC elements are used in feeds for publications
- AgMES elements are used in feeds for events and job vacancies
- When elements or attributes from a namespace are used, the RSS document must contain the correct XML namespace definition:

```
<rss version="2.0" xmlns:dc="http://purl.org/dc/elements/1.1/">
```