

W3C RSS 1.0 News Feed Creation How-To

Eric Miller et al.

24 January 2001

last rev. \$Date: 2003/04/29 14:53:05 \$

Contents: [Really quick cheatsheet](#) · [RSS 1.0](#) · [Simple HTML Markup for Supporting RSS 1.0 generation](#) · [Markup description](#) · [Try it](#) · [Links](#)

[Really quick cheatsheet](#)

If you just want to add the capacity for an RSS feed to a page you own, with minimal additional work, here is a quick list of things to do.

1. Make sure the html/head/title element contains plausible data; it will be used as the title of the RSS news feed.
2. If the page contains a short high-level description of its subject, tag it as a span with class="description".
3. If the page does not contain a short high-level description of its subject, add it, and then tag it. This description will be used as the description of the RSS channel.
4. If you have a section of news items ("What's new?" or the like), use the markup described below to tag the items.
5. If you don't have a section of news items, add one and tag the items using the markup below. (Note that this is not, strictly speaking, essential, but it is far, far easier to use the system described here if you have a what's-new section on the page than if you are trying to use it elsewhere on the page.)

[RSS 1.0](#)

RSS 1.0 ("RDF Site Summary") is an RDF Vocabulary that provides a lightweight multipurpose extensible metadata description and syndication format. In short, it's a means for describing news and events so that they can be shared across the web. Examples of this in practice related to the W3C are the [W3C at meerkat channel](#), the [W3C news-is-free channel](#) and [W3C At A Glance](#). In short, RSS 1.0 is very cool way of describing, managing and making available to very broad audiences relevant and timely information.

This document provides a quick guide for a simple way of creating RSS 1.0 news feeds from your W3C related domain, activity or working group home page. A specific goal of which will be in effective integration of this information to provide team, member and public views of [what's going on at W3C](#). Rael Dornfest's [Writing RSS 1.0](#) provides a more detailed description of RSS 1.0 and links to additional tools and technologies for creating news feeds.

This document is designed to provide an explanation of *one* way which focuses on the use of a simple, descriptive set of HTML markup (discussed below) and a corresponding [XSLT transformation file](#). This work builds on Dan Connolly's [Site Summaries in XHTML](#) work with a focus of providing a more simple, descriptive set of HTML markup (discussed below) and a transformation capabilities that can be used for a larger set of web pages than the W3C's Home Page.

Please note that using this documentation for a working group's page means you can probably make the leap to using the [working group home page markup](#), which, besides allowing you to easily generate an RSS feed, will allow as much data as you markup to be used in other applications.

Simple HTML Markup for Supporting RSS 1.0 generation

With the addition of a small set of elements and descriptive classes, we can semantically annotate HTML documents to support the generation of RSS 1.0 feeds. The following example from the [Semantic Web Home Page](#) for instance provides the basis for describing the relevant feed:

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Semantic Web</title>
</head>

<body>
<blockquote>
<p><span class="summary">The Semantic Web is an extension of
the current web in which information is given well-defined
meaning, better enabling computers and people to work in
cooperation.</span> <cite>-- Tim Berners-Lee, James Hendler,
Ora Lassila, <a
href="http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html">
The Semantic Web</a></cite></p>
</blockquote>
```

News items associated with the feed are described using the following mark-up:

```
<h2 id="new" class="new">Events</h2>

<ul>

<li>

<div class="item" id="x20020125a">
<a class="link"
href="http://www.w3.org/2001/sw/news#x20020125a"><span
class="title">An RDF Schema for P3P</span></a> : <span
class="date">2002-01-25</span>, <span
class="description">The <a
href="http://www.w3.org/P3P/">P3P Working Group</a> has
published An <a
href="http://www.w3.org/TR/2002/NOTE-p3p-rdfschema-20020125">
RDF Schema for P3P</a> as a W3C Note. Based on The <a
href="http://www.w3.org/TR/2001/WD-P3P-20010928/">Platform
for Privacy Preferences 1.0 (P3P1.0) Specification</a>
Last Call Working Draft, the Note represents one possible
RDF schema for P3P. P3P simplifies and automates the
process of reading Web site privacy policies, promoting
trust and confidence in the Web.</span>
</div>
</li>
...
```

This HTML can then be transformed via a specific [XSLT file](#) to create the corresponding [RSS news feed](#). The above example RSS 1.0 looks like the following:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss:channel rdf:about="http://www.w3.org/2001/sw/Overview.rss">
<rss:title>W3C Semantic Web</rss:title>
<rss:link>http://www.w3.org/2001/sw/</rss:link>
<rss:description>The Semantic Web is an extension of the current web in
which information is given well-defined meaning, better enabling
computers and people to work in cooperation.</rss:description>

<rss:items>
<rdf:Seq>
<rdf:li rdf:resource="http://www.w3.org/2001/sw/news#x20020125a"/>
...
</rdf:Seq>
</rss:items>
</rss:channel>

<rss:item rdf:about="http://www.w3.org/2001/sw/news#x20020125a">
<rss:title>An RDF Schema for P3P</rss:title>
<rss:link>http://www.w3.org/2001/sw/news#x20020125a</rss:link>
<rss:date>2002-01-25</rss:date>
<rss:description>The <a href="http://www.w3.org/P3P/">P3P Working Group</a> has
published An <a href="http://www.w3.org/TR/2002/NOTE-p3p-rdfschema-20020125">
RDF Schema for P3P</a> as a W3C Note. Based on The <a href="http://www.w3.org/TR/2001/WD-P3P-20010928/">Platform
for Privacy Preferences 1.0 (P3P1.0) Specification</a>
Last Call Working Draft, the Note represents one possible
RDF schema for P3P. P3P simplifies and automates the
process of reading Web site privacy policies, promoting
trust and confidence in the Web.</rss:description>
```

```

<dc:date>2002-01-23</dc:date>
<rss:description rdf:parseType="Literal">
<span class="description">The <a href="http://www.w3.org/P3P/"
shape="rect">P3P Working Group</a> has published An <a
href="http://www.w3.org/TR/2002/NOTE-p3p-rdfschema-20020125"
shape="rect">RDF Schema for P3P</a> as a W3C Note. Based on The <a
href="http://www.w3.org/TR/2001/WD-P3P-20010928/"
shape="rect">Platform for Privacy Preferences 1.0 (P3P1.0)
Specification</a> Last Call Working Draft, the Note represents one
possible RDF schema for P3P. P3P simplifies and automates the
process of reading Web site privacy policies, promoting trust and
confidence in the Web.</span></rss:description>
</rss:item>
...

```

Markup Description

The following is a more descriptive explanation of the associated mark-up.

It may help to know that an RSS feed is essentially a sequence of items, with a bit of additional collection- or feed-level information. Each item has a title, a link to the original source of the information (the news item is typically extracted from some larger context), and some descriptive text.

The stylesheet creates the feed this way:

- information about the channel. Each input document produces one RSS feed, and thus one channel.
 - the channel's URI (for 'rdf:about') is passed in as a parameter (this is handled by the service)
 - the channel title comes from the /html/head/title element in the input document
 - the channel link is passed in as a parameter
 - the channel description is the concatenation of every span element with class='summary'
- information about the items. Each item in the channel comes from a div with class='item' inside the HTML body
 - the item's rdf:about value comes from the href attribute value of the 'a' element with class='link' which appears inside that div (there should be only one such; if there are more, you won't like the result)
 - the item title, date, and description come from elements inside the div, where 'xxx' is 'title', 'date', or 'description'

Looked at the other way 'round:

- <title> (inside <head>): the title of the RSS 1.0 feed
- : the description of the RSS 1.0 feed
- <div class = "item">: an item associated with (i.e. contained in) the RSS 1.0 feed
- (inside an item): the absolute URI of the enclosing item
- (inside an item): the title of the enclosing item
- : The description of the enclosing item. (You are advised to use CSS to specify the font and properties such as text size and color. This will reduce the size of HTML files and make them easier maintain compared with using elements.
- (inside an item): the date of the enclosing item. **The date format for all W3C News Feeds should be uniform, please use YYYY-MM-DD.**

In case of doubt, it may be helpful to consult the [XSLT transformation sheet](#).

Note, 2003-04-24:

Try it

XSL file:

XML data:

Base:

Channel:

Page:

The *XSL file* is the stylesheet which does the scraping. Unless you are experimenting with a variant copy of it, don't change it from the default. The *XML data* is the XHTML page which contains the markup described above. The *Base* is the base URI to be used in absolutizing any relative URIs in the input; normally, making it the same as *XML data* will be the right thing. The *Channel* is the URI where the output RSS should go. The *Page* is the URI to which the RSS channel will link. (In case of doubt, make it identical to *XML data*.)

Once you're satisfied with the results, this service can be automated through CVS-commit triggered scripting. That is, once you make changes to your XHTML file and commit them to our CVS space, the XSLT can be run over it and the RSS file generated automatically and almost immediately. For now, please contact [Ryan Lee \(ryanlee@w3.org\)](mailto:ryanlee@w3.org) to get your feed set up.

Once this is in place and working to your satisfaction, please let [me](#) know, and I will [configure](#) the [W3C At A Glance](#) service to include this new content.

Links

To see examples, or for further information:

- [W3C at a glance](#). Top-level overview of the news feeds currently generated using this system.
- [W3C at a glance / W3C Architecture Domain / W3C Document Object Model](#). Sample news feed for a single Activity, derived from [the DOM Activity public home page](#), which can be consulted to see in context the HTML markup described above. The [RSS form](#) of the information for this Activity can also be seen.
- [Dominique Hazaël-Massieux], [W3C Organization HTML Profile](#), 21 January 2003. A description of markup used by a similar HTML-scraping tool for extracting information about W3C activities and groups.
- Rael Dornfest, [Writing RSS 1.0](#), O'Reilly Network, 25 August 2000.
- Ryan Lee, [Working Group Home Page Markup Guidelines](#).
- RSS-DEV Working Group, [RDF Site Summary \(RSS\) 1.0](#). General information and links.
- Dan Connolly, [Site Summaries in XHTML](#). August 2000, last revised May 2001.

[Eric Miller](#) <em@w3.org>, (W3C) Semantic Web Activity Lead
with help from Dominique Hazaël-Massieux, Ryan Lee, and Michael Sperberg-McQueen



Slid: howto.html, v 1.12 2003/04/29 14:53:05 cmsmcq Exp \$