



# Engenharia de Dados e Conhecimento

## Feeds Syndication



RSS Advisory Board on March 30, 2009

<http://www.rssboard.org/rss-specification>

# **RSS 2.0 SPECIFICATION (VERSION 2.0.11)**



- Its name is an acronym for *Really Simple Syndication*.
- It is one of the first XML application whose use became widespread on a global scale, thanks to the web.
- RSS itself is little more than a way of summarizing the contents of a web page.
  - Allows administrators to news sites, weblogs and sites updated very often a standardized way of telling the world "what's new" on my site / page.



- The information is delivered as an XML file called "RSS feed".
- RSS data is small and fast-loading. It can easily be used by services like cell phones, for example.
- Special type of applications named aggregators can gather and sort out RSS feeds, coming from several sources, and create new RSS documents to be consumed by simpler programs.

# RSS Aggregator – Google News

universidade de aveiro



Google News

Search

Headlines Local For You U.S. ▾

SECTIONS

- Top Stories
- World
- U.S.
- Business
- Technology
- Entertainment
- Sports
- Science
- Health
- Manage sections

World

**Catalonia crisis: Spain moves to suspend autonomy**  
BBC News · 1h ago

RELATED COVERAGE

**Spain to suspend Catalonia's autonomy - BBC News**  
BBC News

Spain threatens to take over Catalonia's government as constitutional crisis looms  
Washington Post · 1h ago

Spain Sets Stage to Take Control of Catalonia in Independence Fight  
New York Times · 1h ago

View full coverage →

MORE ABOUT

- Catalonia
- Spain

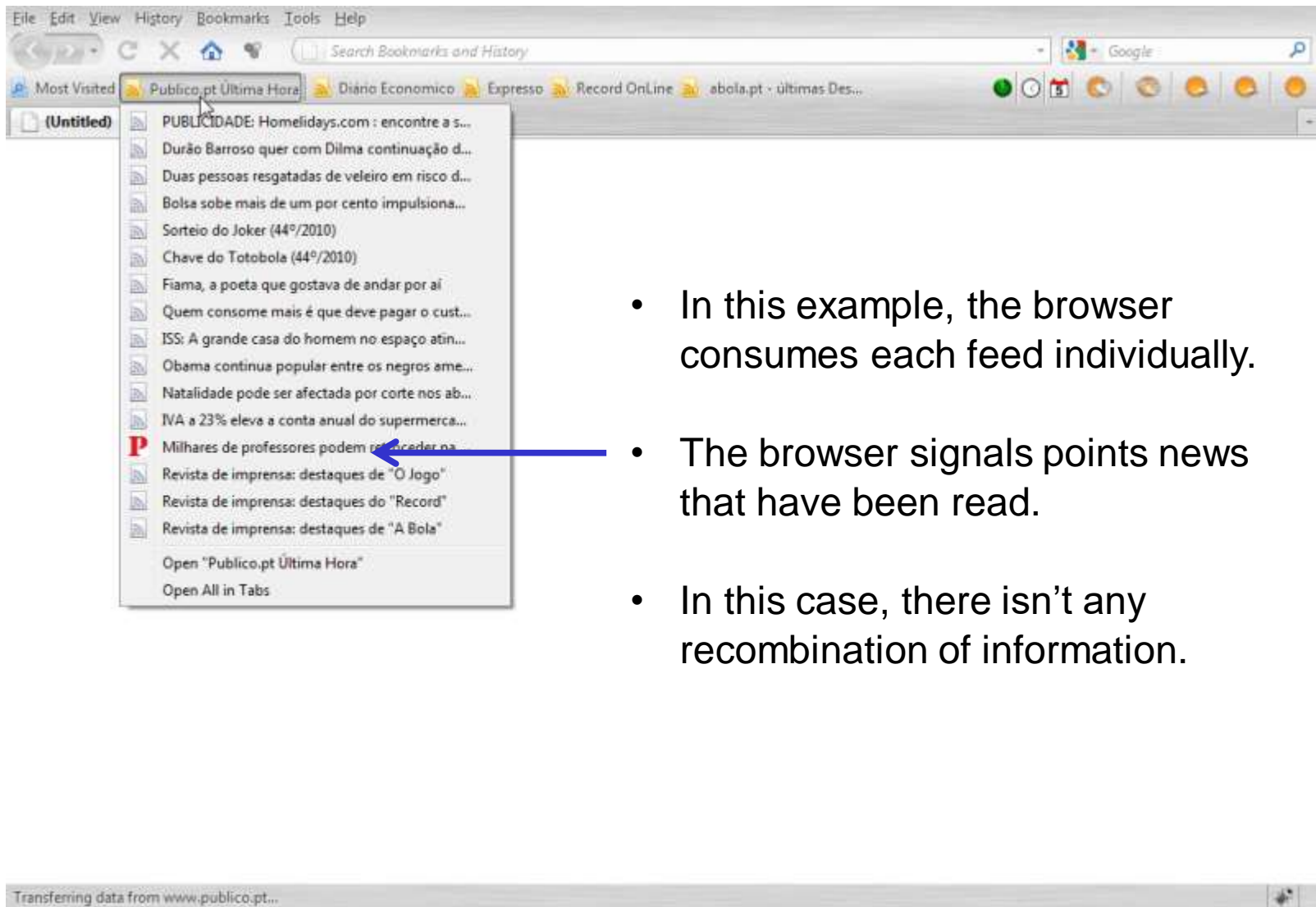
**Quebec Bars People in Face Coverings From Receiving Public Services**  
New York Times · 13h ago

RELATED COVERAGE

Related

- Xi Jinping
- China
- North Korea
- Donald Trump
- Catalonia
- Executive Order 13769
- 19th National Congress of the Communist Party of China
- Diwali
- Spain
- Winston Peters

# RSS Reader – Browser



- In this example, the browser consumes each feed individually.
- The browser signals points news that have been read.
- In this case, there isn't any recombination of information.

# RSS Reader – forecast toolbar

universidade de aveiro



- URL
  - [http://rss.accuweather.com/rss/liveweather\\_rss.asp?metric=1&locCode=EUR|PT|PO001|AVEIRO](http://rss.accuweather.com/rss/liveweather_rss.asp?metric=1&locCode=EUR|PT|PO001|AVEIRO)



- Basic structure of a RSS document:

```
<rss>
  <channel>
    <item></item>
    <item></item>
    <item></item>
  </channel>
  <channel>
  </channel>
</rss>
```





- RSS 2.0 doesn't declare a namespace for its elements.
- `<rss>` is the root element and it must have the version attribute

```
4 <rss version="2.0">
5   <channel>
6     <title>Publico.pt - Geral</title>
7     <link>http://www.publico.pt</link>
8     <description>Últimas 10 notícias do canal "Geral" do PUBLICO.PT</description>
9     <copyright>© 2010 PÚBLICO Comunicação Social SA</copyright>
10    <language>pt-pt</language>
11    <lastBuildDate>Mon, 01 Nov 2010 14:33:02 GMT</lastBuildDate>
12    <category>Geral</category>
13    <ttl>5</ttl>
14    <managingEditor>geralPPT@publico.pt</managingEditor>
15    <webMaster>webmaster@publico.pt</webMaster>
16    <image>
17      <url>http://static.publico.pt/files/header/img/publico.png</url>
18      <title>Logo PÚBLICO</title>
19      <link>http://www.publico.pt</link>
20      <width>85</width>
21    </image>
22    <item>
23      <title>Juros da dívida invertem tendência e ultrapassam os seis por cento nos títulos a 10 anos</title>
24      <description>Os juros exigidos pelos investidores para comprar dívida portuguesa a 10 anos inverteram a tendê
25      <link>http://feedproxy.google.com/~r/PublicoRSS/~3/DIcrbec9ZFY/juros-da-divida-invertem-tendencia-e-ultrapass
26      <category>Economia</category>
27      <pubDate>Mon, 01 Nov 2010 14:10:00 GMT</pubDate>
28    </item>
```

# Channel – mandatory elements

universidade de aveiro



Element	Description	Examples
<b>title</b>	The name of the channel. It's how people refer to your service. If you have an HTML website that contains the same information as your RSS file, the title of your channel should be the same as the title of your website.	Publico.pt – Geral; Publico.pt – Mundo; Publico.pt – Ecosfera; Publico.pt – Política
<b>link</b>	The URL to the HTML website corresponding to the channel	<a href="http://www.publico.pt">http://www.publico.pt</a>
<b>description</b>	Phrase or sentence describing the channel.	Last 10 news from channel PUBLICO.PT – Geral.

# Channel – optional elements (i)



Element	Description	Examples
language	The language the channel is written in. This allows aggregators to group all portuguese language sites, for example, on a single page.	pt-pt
copyright	Copyright notice for content in the channel.	© 2010 PÚBLICO Comunicação Social SA
managingEditor	Email address for person responsible for editorial content.	geralPPT@publico.pt
webMaster	Email address for person responsible for technical issues relating to channel.	webmaster@publico.pt
pubDate	The publication date for the content in the channel. For example, the New York Times publishes on a daily basis, the publication date flips once every 24 hours. That's when the pubDate of the channel changes.	
lastBuildDate	The last time the content of the channel changed.	Mon, 01 Nov 2010 14:33:02 GMT
category	Specify one or more categories that the channel belongs to.	<category>Geral</category>

# Channel – optional elements (ii)



Element	Description	Example
generator	A string indicating the program used to generate the channel.	
docs	A URL that points to the documentation for the format used in the RSS file.	
cloud	Allows processes to register with a cloud to be notified of updates to the channel, implementing a lightweight publish-subscribe protocol for RSS feeds.	
ttl	ttl stands for time to live. It's a number of minutes that indicates how long a channel can be cached before refreshing from the source.	<ttl>5</ttl>
image	Specifies a GIF, JPEG or PNG image that can be displayed with the channel.	
textInput	Specifies a text input box that can be displayed with the channel.	
skipHours	A hint for aggregators telling them which hours they can skip.	
skipDays	A hint for aggregators telling them which days they can skip.	

# Channel / Image – sub-elements

universidade de aveiro



Element	Description	Exampe
url	URL of a GIF, JPEG or PNG image that represents the channel.	
title	Describes the image, it's used in the ALT attribute of the HTML <img> tag when the channel is rendered in HTML.	
link	Is the URL of the site, when the channel is rendered, the image is a link to the site	
width (optional)	Number. Maximum value for width is 144, default value is 88.	
height (optional)	Number. Maximum value for height is 400, default value is 31.	
descripton (optional)	Contains text that is included in the TITLE attribute of the link formed around the image in the HTML rendering.	



- A channel may contain any number of elements <item>.
- An item may represent a “story” – much like a story in a newspaper or magazine.
  - Element <description> must be a synopsis of the story.
  - Element <link> points to the full story.
- An item may also be complete in itself.
  - Element <description> contains all the text.
  - Elements <link> and <title> can be omitted.
- All elements of an item are optional, however at least one of title or description must be present.

# Channel / Item (sub-elements)

universidade de aveiro



Element	Description	Example
title	The title of the item.	Juros da dívida invertem tendência e ultrapassam os seis por cento nos títulos a 10 anos.
link	The URL of the item.	<a href="http://feedproxy.google.com/~r/PublicoRSS/...">http://feedproxy.google.com/~r/PublicoRSS/...</a>
description	The item synopsis.	Os juros exigidos pelos investidores para comprar dívida portuguesa a 10 anos inverteram ...
author	Email address of the author of the item.	
category	Includes the item in one or more categories.	Economia
comments	URL of a page for comments relating to the item.	
enclosure	Describes a media object that is attached to the item.	
guid	A string that uniquely identifies the item.	
pubDate	Indicates when the item was published.	Mon, 01 Nov 2010 14:10:00 GMT
source	The RSS channel that the item came from.	



The Internet Engineering Task Force (IETF) – December 2005

<http://tools.ietf.org/html/rfc4287>

# **ATOM – ATOM SYNDICATION FORMAT**



# What is ATOM?



- Atom is an XML-based document format that describes lists of related information known as "feeds".
- Feeds are composed of a number of items, known as "entries", each with an extensible set of attached metadata. For example, each entry has a title.
- The primary use case that Atom addresses is the syndication of Web content such as weblogs and news headlines to Web sites as well as directly to user agents.

Source: <http://tools.ietf.org/html/rfc4287>



- Very similar, but with subtle differences.
- ATOM has the attribute type which allows to differentiate between text contents, for example.
- Examples

```
<title type="text">  
  Less: &lt;  
</title>
```

```
<title type="html">  
  Less: &lt;em> &amp;lt; &lt;/em>  
</title>
```

```
<title type="xhtml" xmlns:xhtml="http://www.w3.org/1999/xhtml">  
  <xhtml:div>  
    Less: <xhtml:em> &lt; </xhtml:em>  
  </xhtml:div>  
</title>
```



## Atom - Feed

- author
- id
- title
- updated

- generator
- icon
- logo
- rights
- subtitle

- category
- contributor
- link
- **entry**

Mandatory

Optional - Unic

Optional - Multiple

## Atom - Entry

- id
- author
- summary
- title
- updated

- content
- published
- rights
- source
- summary

- category
- contributor
- link



- If they do the same and look the same, what is different?
- Internationalization
  - Atom allows the use of the xml:lang standard not only at the level of the feed, but also at the level of items and characteristics.
  - RSS defines the language only at the feed level.
- Modularity
  - Atom allows elements to be reused outside the context of an Atom feed document.
  - RSS doesn't allow it.
- Information
  - In Atom we can automatically know what kind of data is sent in a field, while in RSS it's not.

# ATOM vs RSS (ii)



RSS 2.0	Atom 1.0
rss	-
channel	feed
title	title
link	link
description	subtitle
language	-
copyright	rights
webMaster	-
managingEditor	author or contributor
pubDate	published (in entry)
lastBuildDate (in channel)	updated
category	category
generator	generator
docs	-
cloud	-
ttl	-

# ATOM vs RSS (iii)



<b>RSS 2.0</b>	<b>Atom 1.0</b>
image	logo
-	icon
rating	-
textInput	-
skipHours	-
skipDays	-
item	entry
author	author
-	contributor
description	summary and/or content
comments	-
enclosure	-
guid	id
source	-
-	source

# ATOM – Example (i)



```
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <title type="text">dive into mark</title>
  <subtitle type="html">
    A &lt;em>lot</em> of effort
    went into making this effortless
  </subtitle>
  <updated>2005-07-31T12:29:29Z</updated>
  <id>tag:example.org,2003:3</id>
  <link rel="alternate" type="text/html"
    hreflang="en" href="http://example.org/">
  <link rel="self" type="application/atom+xml"
    href="http://example.org/feed.atom"/>
  <rights>Copyright (c) 2003, Mark Pilgrim</rights>
  <generator uri="http://www.example.com/" version="1.0">
    Example Toolkit
  </generator>
  <entry>...</entry>
</feed>
```

# ATOM – Example (ii)

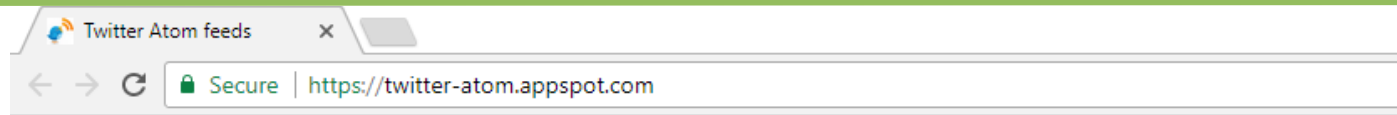


```
<entry>
  <title>Atom draft-07 snapshot</title>
  <link rel="alternate" type="text/html"
    href="http://example.org/2005/04/02/atom"/>
  <link rel="enclosure" type="audio/mpeg" length="1337"
    href="http://example.org/audio/ph34r_my_podcast.mp3"/>
  <id>tag:example.org,2003:3.2397</id>
  <updated>2005-07-31T12:29:29Z</updated>
  <published>2003-12-13T08:29:29-04:00</published>
  <author>
    <name>Mark Pilgrim</name>
    <uri>http://example.org/</uri>
    <email>f8dy@example.com</email>
  </author>
  <contributor>
    <name>Sam Ruby</name>
  </contributor>
  <contributor>
    <name>Joe Gregorio</name>
  </contributor>
  <content type="xhtml" xml:lang="en" xml:base="http://diveintomark.org/">
    <div xmlns="http://www.w3.org/1999/xhtml">
      <p><i>[Update: The Atom draft is finished.]</i></p></div>
    </content>
</entry>
```



# Atom Feeds Site Example

universidade de aveiro



## Twitter Atom feeds


Want an Atom feed of your [Twitter stream](#), or a [Twitter list](#), that you can subscribe to and read in your favorite RSS feed reader? This site can provide one for you!

First, [click here to get a Twitter app key](#). (Why?) Enter [details like these](#), click the [checkbox and button](#), and copy and paste the [consumer key and secret](#) below. Hover over these links for help.

Consumer key

Consumer secret

Second, choose either your full stream or just the people on a list. For a list, use the format `username/list`.

 **Get my feed**

OR

 **Get this list's feed**

When you click one of these buttons, Twitter will ask you if your app should be allowed to read your Twitter stream. [Click Authorize app](#) and you'll be redirected back here to get your new feed.