Sources of news by country [2015]

<table>
<thead>
<tr>
<th></th>
<th>UK</th>
<th>Germany</th>
<th>Spain</th>
<th>Italy</th>
<th>France</th>
<th>Denmark</th>
<th>Finland</th>
<th>USA</th>
<th>Urban Brazil</th>
<th>Japan</th>
<th>Ireland</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TV</strong></td>
<td>75%</td>
<td>82%</td>
<td>82%</td>
<td>78%</td>
<td>80%</td>
<td>75%</td>
<td>75%</td>
<td>64%</td>
<td>81%</td>
<td>73%</td>
<td>76%</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Radio</strong></td>
<td>37%</td>
<td>50%</td>
<td>40%</td>
<td>23%</td>
<td>28%</td>
<td>50%</td>
<td>45%</td>
<td>26%</td>
<td>39%</td>
<td>17%</td>
<td>50%</td>
<td>41%</td>
</tr>
<tr>
<td><strong>Printed Newspapers</strong></td>
<td>38%</td>
<td>38%</td>
<td>47%</td>
<td>38%</td>
<td>19%</td>
<td>33%</td>
<td>49%</td>
<td>23%</td>
<td>33%</td>
<td>44%</td>
<td>49%</td>
<td>39%</td>
</tr>
<tr>
<td><strong>Online (inc. social media)</strong></td>
<td>73%</td>
<td>60%</td>
<td>86%</td>
<td>81%</td>
<td>71%</td>
<td>85%</td>
<td>90%</td>
<td>74%</td>
<td>91%</td>
<td>70%</td>
<td>83%</td>
<td>85%</td>
</tr>
</tbody>
</table>
Main source of news by age – young prefer online – all countries

Social media as a source of news – all countries

February’s Most Shared Sites

<table>
<thead>
<tr>
<th>Website</th>
<th>Facebook Shares</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Huffington Post</td>
<td>4,631,287</td>
</tr>
<tr>
<td>BuzzFeed</td>
<td>3,354,310</td>
</tr>
<tr>
<td>Fox News</td>
<td>3,043,063</td>
</tr>
<tr>
<td>The Guardian</td>
<td>2,905,813</td>
</tr>
<tr>
<td>The New York Times</td>
<td>2,805,191</td>
</tr>
<tr>
<td>NBC</td>
<td>2,607,382</td>
</tr>
<tr>
<td>BBC</td>
<td>2,565,922</td>
</tr>
<tr>
<td>Washington Post</td>
<td>2,432,556</td>
</tr>
<tr>
<td>Daily Mail</td>
<td>2,245,179</td>
</tr>
<tr>
<td>CNN</td>
<td>1,984,541</td>
</tr>
</tbody>
</table>

Based on content published in Feb. 2016 only.

Digital news metadata: Publisher and library perspectives

Frederick Zarndt
IFLA Governing Board
Digital Divide Data
frederick@frederickzarndt.com
@cowboyMontana
Digital news metadata: Publisher and library perspectives

Frederick Zarndt
IFLA Governing Board
Digital Divide Data
frederick@frederickzarndt.com
@cowboyMontana
Why digital news metadata?

"Once I know WHY, I can figure out what I can do about that ... or if I CAN do anything about that.”

Image from Monique Torres. “Why” is the Marketer’s Secret Sauce: 5 Buyer Persona Insights. [http://goo.gl/pAe78e](http://goo.gl/pAe78e)
Overview

- IPTC: The Methuselah of digital news
- Google News
- [schema.org](http://schema.org)
- Facebook’s open graph protocol and instant news
- RSS: Methuselah’s younger sibling
- Metadata crosswalk
- Issues vs. stories vs. websites
IPTC International Press
Telecommunications Council: Methuselah of Digital News Standards

Stained glass window in Canterbury Cathedral. Gveret Tered. 2008. https://creativecommons.org/licenses/by-sa/2.0/
Digital News, Enabled.

The IPTC is the global standards body of the news media. We provide the technical foundation for the news ecosystem.
I can't imagine being a professional photographer without using IPTC Photo Metadata. It makes my life easy: I can search my database efficiently and I don't lose track of photos.

Hardy Klahold, Photographer and Owner, Hardy Klahold Photography
The IPTC had significant impact and leadership in the development of the W3C ODRL specifications that is the underlying framework for RightsML. Their insight and experience has driven both standards to be world-leading in the domain of rights management.

Renato Iannella, Semantic Identity and Chair, W3C ODRL Community Group
Our success is largely connected to the worldwide adoption of the IPTC Standards, upon which our products are heavily dependent. They allow us to connect to and integrate with anyone and anything - a fundamental requirement for succeeding with software products.

Christopher Frenning, CEO of FotoWare
IPTC was instrumental in developing the rNews standard, now part of the Schema.org standard, and Parse.ly uses this markup as the basis for our content analytics platform.

As a result, major news publishers globally have integrated semantic data into their websites, and the entire web is better for it!

Andrew Montalenti, Co-Founder & CTO, Parse.ly
The IPTC brought openness, flexibility, tenacity and deep domain expertise to their participation in the Schema.org effort, and our cooperation set the standard for all others that followed.

R.V. Guha, Google Fellow and Co-founder, Schema.org
# IPTC Standards

<table>
<thead>
<tr>
<th>Standard</th>
<th>Released</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPTC 7901</td>
<td>1979</td>
<td>1995</td>
</tr>
<tr>
<td>IPTC Information Interchange Model (IIM)</td>
<td>1991</td>
<td>2014</td>
</tr>
<tr>
<td>IPTC NITF [XML]</td>
<td>early 1990s</td>
<td>2012</td>
</tr>
<tr>
<td>IPTC Photo Metadata</td>
<td>early 1990s</td>
<td>2014</td>
</tr>
<tr>
<td>IPTC Subject Codes / Media Topics [taxonomy]</td>
<td>~1400 terms</td>
<td>early 1990s</td>
</tr>
<tr>
<td>IPTC RightsML [XML]</td>
<td>2013</td>
<td>2013</td>
</tr>
<tr>
<td>IPTC rNews [HTML]</td>
<td>2011</td>
<td>2013</td>
</tr>
<tr>
<td>IPTC ninjs [JSON]</td>
<td>2014</td>
<td></td>
</tr>
</tbody>
</table>

IPTC standards are released under Creative Commons Attribution 3.0 ([https://creativecommons.org/licenses/by/3.0/](https://creativecommons.org/licenses/by/3.0/))
IPTC Photo Metadata

Data and information is entered into the image file by users or by automated capture from cameras or scanners.

- There are 31 IPTC Core metadata fields, 46 Extension fields
- IPTC photo metadata is primarily descriptive, administrative, and rights
- Metadata is stored in two main places
  - Internally – embedded in the image file in formats such as JPEG or TIFF.
  - Externally – outside the image file in a digital asset management system (DAM) or by a “sidecar” file, such as XMP, or an external XML-based news exchange format file.
IPTC NITF
News Industry Text Format

• Facilitates the exchange of text news. Multimedia “attachments” are supported
• Published as a XML-based standard in 1998
• Released version 3.4 in May 2007 as DTD and XML schema
• Supports identification and description of news
  • Who owns the copyright to the item, who may republish it, and who it's about.
  • What subjects, organizations, and events it covers.
  • When it was reported, issued, and revised.
  • Where it was written, where the action took place, and where it may be released.
  • Why it is newsworthy, based on the editor's analysis of the metadata.
IPTC rNews

... news providers have created feeds to supply news using IPTC formats such as NITF and NewsML-G2. ... there are an increasing number of consumers of news who only want to work with "pure" web technologies, i.e. HTML rather than XML. ...IPTC has been looking at the two major paths to represent metadata in HTML - microformats and RDFa.

- **hNews** is the microformat for news that was adopted by the [news] community in late 2009. It builds upon hAtom by adding a few news-specific fields.

- **rNews** is a proposal for a semantic markup vocabulary for news. ... rNews and hNews are similar in intent (news-specific metadata in HTML) but somewhat different in approach. Whereas hNews went through the microformats process, a semantic markup vocabulary can be created by anyone. The IPTC recently adopted rNews 1.0, based somewhat on the NewsML-G2, NITF and hNews models but extending beyond those standards ...

  - **rNews** is an approved standard for using semantic markup to annotate news-specific metadata in HTML documents.

  - **rNews** specifies the terminology and data model required to embed news specific metadata into HTML documents

  - **rNews** uses RDFa triples or Microdata
IPTC rNews

- **rNews** provides a basic semantic markup of the content of a web page.

- **rNews** apply to the portion of feed workflows where publishers target primarily the end user but also any business which makes use of the public presentation of content on the web [search engines].

- **rNews** provides no or only a very limited set of metadata for managing content. [Managing content] is the domain of the G2-Standards NewsML-G2 and EventsML-G2, of NITF and NewsML 1.

- **rNews** is not made for delivering packages of structured news items in a business-to-business context, but it can reflect the composite content of a web page.
Google News RSS feeds are available in RSS 2.0 format.
• Mark up your articles using the following properties of the https://schema.org/Article type. Be sure to mark up your page with the most specific applicable schema.org type. For example, a news article should be marked up as https://schema.org/NewsArticle and a blog post as https://schema.org/BlogPosting.

• Be unique. Each of your pages that display an article's full text needs to have a unique URL.

• Be permanent. ... In order to ensure that our links to articles function properly, each article on your news site needs to be associated with one unique URL, and that URL must be permanent (i.e., it can't be recycled).

• Additionally, do not republish previously published articles under a new URL.

• Google News RSS feeds are available in RSS 2.0 format.

Adapted from https://support.google.com/news/publisher
schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies...

Adapted from https://schema.org/
### Thing > CreativeWork > Article > NewsArticle

#### Properties from **NewsArticle**

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dateline</td>
<td>Text</td>
<td>The location where the NewsArticle was produced.</td>
</tr>
<tr>
<td>printColumn</td>
<td>Text</td>
<td>The number of the column in which the NewsArticle appears in the print edition.</td>
</tr>
<tr>
<td>printEdition</td>
<td>Text</td>
<td>The edition of the print product in which the NewsArticle appears.</td>
</tr>
<tr>
<td>printPage</td>
<td>Text</td>
<td>If this NewsArticle appears in print, this field indicates the name of the page on which the article is found. Please note that this field is intended for the exact page name (e.g. A5, B18).</td>
</tr>
<tr>
<td>printSection</td>
<td>Text</td>
<td>If this NewsArticle appears in print, this field indicates the print section in which the article appeared.</td>
</tr>
</tbody>
</table>

#### Properties from **Article**

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>articleBody</td>
<td>Text</td>
<td>The actual body of the article.</td>
</tr>
<tr>
<td>articleSection</td>
<td>Text</td>
<td>Articles may belong to one or more ‘sections’ in a magazine or newspaper, such as Sports, Lifestyle, etc.</td>
</tr>
<tr>
<td>pageEnd</td>
<td>Text or Integer</td>
<td>The page on which the work ends; for example “138” or “xvi”.</td>
</tr>
<tr>
<td>pageStart</td>
<td>Text or Integer</td>
<td>The page on which the work starts; for example “135” or “xiii”.</td>
</tr>
<tr>
<td>pagination</td>
<td>Text</td>
<td>Any description of pages that is not separated into pageStart and pageEnd; for example, “1–6, 9, 55” or “10–12, 46–49”.</td>
</tr>
<tr>
<td>wordCount</td>
<td>Integer</td>
<td>The number of words in the text of the Article.</td>
</tr>
</tbody>
</table>

#### Properties from **CreativeWork**

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>about</td>
<td>Thing</td>
<td>The subject matter of the content.</td>
</tr>
<tr>
<td>accessibilityAPI</td>
<td>Text</td>
<td>Indicates that the resource is compatible with the referenced accessibility API (<a href="https://schema.org/accessibilityAPI">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityControl</td>
<td>Text</td>
<td>Identifies input methods that are sufficient to fully control the described resource (<a href="https://schema.org/accessibilityControl">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityFeature</td>
<td>Text</td>
<td>Content features of the resource, such as accessible media, alternatives and supported enhancements for accessibility (<a href="https://schema.org/accessibilityFeature">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityHazard</td>
<td>Text</td>
<td>A characteristic of the described resource that is physiologically dangerous to some users. Related to WCAG 2.0 guideline 2.3 (<a href="https://schema.org/accessibilityHazard">WebSchemas wiki lists possible values</a>).</td>
</tr>
</tbody>
</table>

From [https://schema.org/NewsArticle](https://schema.org/NewsArticle)
As of April 2016, NewsArticle ...

- ... has 5 unique (not inherited) properties
- ... has 83 properties inherited from Article, CreativeWork, and Thing
- ... may appear as an AssociatedArticle property on MediaObjects.

From https://schema.org/NewsArticle
From http://open.blogs.nytimes.com/2012/02/16/rnews-is-here-and-this-is-what-it-means
On 22 March 2016, the Washington Post published this story on its website @ https://goo.gl/dEoPTR

The simple, brilliant plan to change macaroni and cheese that duped us all

Last April, Kraft announced that it would eventually change its famous macaroni and cheese product. The plan was to get rid of artificial preservatives and replace artificial dyes with a trio of spices: paprika, annatto and turmeric. But there was no timetable — at least none shared publicly.
The simple, brilliant plan to change macaroni and cheese that duped us all
Follow schema.org

Welcome to Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use Schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies to power rich, extensible experiences.

Schema.org is sponsored by Google, Microsoft, Yahoo and Yandex. The vocabularies are developed by an open community process, using the public- schemaorg@w3.org mailing list and through GitHub.

A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefit for their efforts. It is in this spirit that the sponsors, together with the larger community have come together, to provide a shared collection of schemas.

We invite you to get started!

View our blog at blog.schema.org or see release history.
Follow schema.org on Github
Facebook is the No.1 Social Traffic Source for News Websites
Percentage of website traffic coming from social media sites in the past three months

<table>
<thead>
<tr>
<th>Website</th>
<th>Facebook</th>
<th>Twitter</th>
<th>Reddit</th>
<th>Stumbleupon</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>WashingtonPost.com</td>
<td>9.7%</td>
<td>3.7%</td>
<td>4.5%</td>
<td>4.3%</td>
<td>1.8%</td>
</tr>
<tr>
<td>HuffingtonPost.com</td>
<td>8.7%</td>
<td>2.9%</td>
<td>4.1%</td>
<td>3.5%</td>
<td>2.7%</td>
</tr>
<tr>
<td>NYTtimes.com</td>
<td>8.3%</td>
<td>3.2%</td>
<td>4.0%</td>
<td>3.9%</td>
<td>2.8%</td>
</tr>
<tr>
<td>WSJ.com</td>
<td>6.6%</td>
<td>2.2%</td>
<td>3.5%</td>
<td>4.4%</td>
<td>1.7%</td>
</tr>
<tr>
<td>LATimes.com</td>
<td>6.6%</td>
<td>2.3%</td>
<td>3.5%</td>
<td>4.5%</td>
<td>1.6%</td>
</tr>
<tr>
<td>USATODAY.com</td>
<td>6.1%</td>
<td>2.1%</td>
<td>3.5%</td>
<td>4.6%</td>
<td>1.7%</td>
</tr>
<tr>
<td>ABCNews.go.com</td>
<td>5.6%</td>
<td>2.0%</td>
<td>3.5%</td>
<td>4.8%</td>
<td>1.7%</td>
</tr>
<tr>
<td>CNN.com</td>
<td>4.9%</td>
<td>1.8%</td>
<td>3.5%</td>
<td>5.2%</td>
<td>1.7%</td>
</tr>
<tr>
<td>FoxNews.com</td>
<td>2.9%</td>
<td>1.2%</td>
<td>3.5%</td>
<td>5.5%</td>
<td>1.7%</td>
</tr>
<tr>
<td>NBCNews.com</td>
<td>2.4%</td>
<td>1.0%</td>
<td>3.5%</td>
<td>6.0%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

Source: SimilarWeb

Facebook Is a Must for Brands Around the World
Percentage of brands present on the following platforms (June 2014)

- Facebook: 100.0%
- Twitter: 96.3%
- YouTube: 94.2%
- Instagram: 90.8%
- Google+: 79.1%
- Pinterest: 75.4%
- Weibo: 59.7%
- WeChat: 44.0%
- Youku: 39.5%
- Tumblr: 29.3%

n=382 brands
Source: L2

**TOP SOCIAL NETWORKS**

- Facebook: 40%
- YouTube: 16%
- Twitter: 11%
- Google+: 5%
- Reddit: 4%

*used weekly for news

**DIGITAL PARTICIPATION**

32% share a news story via email or social media. US is 4th out of 12 in overall participation index.

**TOP SOCIAL NETWORKS**

- Facebook: 23%
- YouTube: 13%
- WhatsApp: 9%
- Google+: 6%
- Twitter: 4%

*used weekly for news

Germans are less interested in news-related participation via social media than people in other countries. Facebook is still the biggest network for news while Twitter attracts media coverage but has struggled to appeal to the wider public. WhatsApp has been growing fast in Germany over the last few years and some local newspapers have been experimenting with it for distributing their stories.
Open Graph Protocol (OG)

- OG enables any web page to become an object in a social graph.
- News articles, blog posts, videos, audio clips, websites, and more can be described by OG.
- OG is used or recognized by Facebook, Google, Pinterest, LinkedIn, Twitter, WordPress, and others.

Adapted from http://ogp.me/
Basic OG Tags

The four required metadata properties for every OG object

- **og:title** - The title of the object as it should appear within the graph, e.g., "The Rock".

- **og:type** - The type of the object, e.g., "video.movie". Depending on the type specified, other properties may also be required.

- **og:image** - An image URL which should represent the object within the graph.

- **og:url** - The canonical URL of the object. This will be used as its permanent ID in the graph, e.g., "http://www.imdb.com/title/tt0117500/".

Adapted from [http://ogp.me/#metadata](http://ogp.me/#metadata)
Other Tags

Other metadata properties that may be attached to an OG news story

- **article:author** - Array of Facebook profile URLs or IDs of the authors for this article
- **article:expiration_time** - A time when the article expired (or will expire)
- **article:modified_time** - A time when the article was last modified
- **article:published_time** - A time when the article was published
- **article:publisher** - A Facebook page URL or ID of the publishing entity
- **article:section** - The section of your website to which the article belongs, such as 'Lifestyle' or 'Sports'
- **article:tag** - An array of keywords relevant to the article

Adapted from [http://ogp.me/#type_article](http://ogp.me/#type_article)
FaceBook Instant News

- An Instant Article is a HTML5 document optimized for fast mobile performance, rich storytelling capabilities, branded design and customized visual display.

- Every article published as an Instant Article must be published on a news publisher's website as well. That means when someone shares an Instant Article by email or on Twitter, they also share a link to the publisher's website.

From [https://developers.facebook.com/docs/instant-articles](https://developers.facebook.com/docs/instant-articles)
On 22 March 2016, the Washington Post published this story on its website @ https://goo.gl/dEoPTR

Last April, Kraft announced that it would eventually change its famous macaroni and cheese product. The plan was to get rid of artificial preservatives and replace artificial dyes with a trio of spices: paprika, annatto and turmeric. But there was no timetable — at least none shared publicly.
As shared on Facebook, the story had the properties below (and many others, some OG, some schema.org, some HTML5) https://goo.gl/dEoPTR

<meta property="og:description" content="Macaroni and cheese has changed--and nobody noticed.”/>

<meta property="og:type" content="article"/> <meta property="og:site_name" content="Washington Post"/>

<meta property="og:title" content="The simple, brilliant plan to change macaroni and cheese that duped us all”/> 

<meta itemprop="image" property="og:image" content="http://www.washingtonpost.com/blogs/wonkblog/files/2016/03/15426457201_3f7f18a168_o.jpg"/>

<meta property="article:publisher" content="https://www.facebook.com/washingtonpost">

The simple, brilliant plan to change macaroni and cheese that duped us all

By Roberto A. Ferdman

Last April, Kraft announced that it would eventually change its famous macaroni and cheese product. The plan was to get rid of artificial preservatives and replace artificial dyes with a trio of spices: paprika, annatto and turmeric. But there was no timetable — at least none shared publicly.

Kraft’s diabolical konspiracy unveiled!

The simple, brilliant plan to change macaroni and cheese that duped us all

Macaroni and cheese has changed—and nobody noticed.

WASHINGTONPOST.COM | BY ROBERTO FERDMAN

On the Washington Post website

On Facebook

From Washington Post 22 March 2016: https://goo.gl/dEoPTR
Our latest Freakonomics Radio episode is called “The Economics of Sleep, Part 1.” (You can subscribe to the podcast at iTunes or elsewhere, get the RSS feed, or listen via the media player above.)

The gist: Could a lack of sleep help explain why some people get much sicker than others?

Below is a transcript of the episode, modified for your reading pleasure. For more information on the people and ideas in the episode, see the links at the bottom of this post. And you'll find credits for the music in the episode noted within the transcript.

* * *

This week we’re bringing you an episode from our archives. It’s called, “The Economics of Sleep, Part 1.” And yes, that means that there's a Part 2, which you’ll hear next week. We thought it was time to replay these episodes because they are two of our most popular episodes ever. Now, why is that? I think it may be because as much as people tend to focus on nutrition and exercise as the vital inputs in maintaining the human machine, sleep often gets overlooked. So, let's stop overlooking it, yeah? Hope you enjoy, and I hope you learn as much about sleep as we did in making this episode.

* * *
**RSS**

*Really Simply Syndication* or *Rich Site Summary* (RSS) is a family of standard ([http://www.rss-specifications.com/](http://www.rss-specifications.com/)) web feed formats to publish frequently updated information: blog entries, news headlines, audio, video. An RSS document (called "feed", "web feed", or "channel") includes full or summarized text, and metadata, like publishing date and author's name.

RSS was 1st used by Netscape in 1999. RSS is now administered by the RSS Advisory Board ([http://www.rssboard.org/](http://www.rssboard.org/))

Atom is a similar, competing syndication format which addresses some of the perceived problems with RSS. According to IETF RFC 4287

*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.*

RSS definition from [https://en.wikipedia.org/wiki/RSS](https://en.wikipedia.org/wiki/RSS)

RSS Channel Elements

The three **required** elements for every RSS channel

- **rss:channel:title** - Name of the channel. If the publisher has an HTML website that contains the same information as its RSS file, the title of the publisher’s channel should be the same as the title of its website.

- **rss:channel:link** - URL to the HTML website corresponding to the channel.

- **rss:channel:description** - Phrase or sentence describing the channel.

A RSS channel may also have a number of **optional** elements such as **copyright, pubDate, category, language, ttl**, etc.

Adapted from [https://cyber.law.harvard.edu/rss/rss.html#requiredChannelElements](https://cyber.law.harvard.edu/rss/rss.html#requiredChannelElements)
Adapted from [https://cyber.law.harvard.edu/rss/rss.html#optionalChannelElements](https://cyber.law.harvard.edu/rss/rss.html#optionalChannelElements)
RSS Item Elements

The elements for that can be used to describe each item published on the RSS channel. Multiple items may be published on a channel; for each item at least title or description must appear.

- **rss:item:title** - Title of the item. For each RSS item at least title or description
- **rss:item:link** - URL to the HTML website on which the item is published.
- **rss:item:description** - Synopsis of the item.
- **rss:item:author** - Email address of the author of the item.
- **rss:item:guid** - A string that uniquely identifies the item.

A RSS item may also have a number of other optional elements such as category, pubDate, source, etc

Adapted from [https://cyber.law.harvard.edu/rss/rss.html#requiredChannelElements](https://cyber.law.harvard.edu/rss/rss.html#requiredChannelElements)
RSS 2.0

- Many, many news publishers, blogs, etc, publish RSS channels.
- RSS 2.0 is published under Creative Commons Attribution-ShareAlike Generic CC-BY-SA-1.0 (https://creativecommons.org/licenses/by-sa/1.0/).
- The RSS 2.0 specification is owned and maintained by the Berkman Center for Internet & Society at Harvard Law School (http://www.rss-specifications.com/).
- W3C RSS Reference is at http://www.w3schools.com/xml/xml_rss.asp.
A Simple RSS Feed

<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
    <channel>
        <title>RSS Title</title>
        <description>This is an example of an RSS feed</description>
        <link>http://www.example.com/main.html</link>
        <lastBuildDate>Mon, 06 Sep 2010 00:01:00 +0000 </lastBuildDate>
        <pubDate>Sun, 06 Sep 2009 16:20:00 +0000</pubDate>
        <ttl>1800</ttl>
        <item>
            <title>Example entry</title>
            <description>Here is some text containing an interesting description.</description>
            <link>http://www.example.com/blog/post/1</link>
            <guid isPermaLink="true">7bd204c6-1655-4c27-aeee-53f933c5395f</guid>
            <pubDate>Sun, 06 Sep 2009 16:20:00 +0000</pubDate>
        </item>
    </channel>
</rss>
# Metadata Crosswalk

<table>
<thead>
<tr>
<th></th>
<th>MODS</th>
<th>rNews</th>
<th>NITF</th>
<th>schema.org</th>
<th>RSS 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publication title</strong></td>
<td>MODS:title</td>
<td></td>
<td></td>
<td>CreativeWork:publication</td>
<td>RSS:channel:title</td>
</tr>
<tr>
<td><strong>Publication ID</strong></td>
<td>MODS:identifier (ISSN)</td>
<td>NewsItem:identifier</td>
<td>NITF:pubdata</td>
<td>Thing:url</td>
<td>RSS.channel:link</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>MODS:publisher</td>
<td>NewsItem:sourceOrganization</td>
<td>NITF:distributor</td>
<td>CreativeWork:producer</td>
<td>RSS:channel:description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NITF:bytti</td>
<td>CreativeWork:publisher</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CreativeWork:sourceOrganizaon</td>
<td></td>
</tr>
<tr>
<td><strong>Place of publication</strong></td>
<td>MODS:place</td>
<td>Organization:address</td>
<td></td>
<td>NITF:pubdata</td>
<td>RSS.channel:link</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NewsItem:dateline</td>
<td></td>
<td>NewsArticle:dateline</td>
<td></td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>MODS:dateIssued</td>
<td>NewsItem:dateCreated</td>
<td>NITF:issue</td>
<td>CreativeWork:dateCreated</td>
<td>RSS:channel:pubDate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NewsItem:datePublished</td>
<td>NITF:release</td>
<td>CreativeWork:dateModified</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NewsItem:dateModified</td>
<td>NITF:expire</td>
<td>CreativeWork:datePublished</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NITF:dateline</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Article title</strong></td>
<td>MODS:title</td>
<td>NewsItem:headline</td>
<td>NITF:title</td>
<td>CreativeWork:headline</td>
<td>RSS.item.title</td>
</tr>
<tr>
<td><strong>Article author</strong></td>
<td>MODS:name MODS:role</td>
<td>NewsItem:creator</td>
<td>NITF:byline</td>
<td>CreativeWork:author</td>
<td>RSS.item.author</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NITF:tagline</td>
<td>CreativeWork:creator</td>
<td></td>
</tr>
<tr>
<td><strong>Article identifier</strong></td>
<td>MODS:genre</td>
<td>NewsItem:contributor</td>
<td>NITF:doc-id</td>
<td>CreativeWork:contributor</td>
<td>RSS.item.guid</td>
</tr>
<tr>
<td></td>
<td>MODS:subject</td>
<td>NewsItem:about</td>
<td>NITF:identified-content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NewsItem:genre</td>
<td>NITF:abstract</td>
<td>CreativeWork:genre</td>
<td></td>
</tr>
<tr>
<td><strong>Article subject</strong></td>
<td>MODS:subject</td>
<td></td>
<td></td>
<td></td>
<td>RSS.item.category</td>
</tr>
</tbody>
</table>
Digital News Metadata Standards Are Alive

# A Brief Standards Release History

<table>
<thead>
<tr>
<th>Standard</th>
<th>Released</th>
<th>Last update</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPTC 7901</td>
<td>1979</td>
<td>1995</td>
</tr>
<tr>
<td>IPTC Information Interchange Model (IIM)</td>
<td>1991</td>
<td>2014</td>
</tr>
<tr>
<td>IPTC NITF [XML]</td>
<td>early 1990s</td>
<td>2012</td>
</tr>
<tr>
<td>IPTC Photo Metadata</td>
<td>early 1990s</td>
<td>2014</td>
</tr>
<tr>
<td>IPTC Subject Codes / Media Topics [taxonomy]</td>
<td>early 1990s</td>
<td>2015</td>
</tr>
<tr>
<td>IPTC rNews [HTML]</td>
<td>2011</td>
<td>2015</td>
</tr>
<tr>
<td>IPTC RightsML [XML]</td>
<td>2013</td>
<td>2013</td>
</tr>
<tr>
<td>schema.org</td>
<td>spring 2012</td>
<td>late 2015</td>
</tr>
<tr>
<td>Open Graph Protocol / API / Instant News</td>
<td>2010</td>
<td>fall 2015</td>
</tr>
<tr>
<td>RSS</td>
<td>1997</td>
<td>2003</td>
</tr>
<tr>
<td>MODS</td>
<td>2002</td>
<td>spring 2015</td>
</tr>
</tbody>
</table>
Feb 24 : Keep Metadata of Photos Alive and Intact

Metadata is of high value for many parties in the photo business: photographers, libraries, agencies, archives. But many consider this only in a very limited context, not across many transitions in a supply chain or for a longer period.

The IPTC Photo Metadata Conference on 26 May 2016 in Zagreb (Croatia, Europe) will address how to avoid losing information when images are moved from one person or system to the next one or if they are kept in an archive for a long term.

All interested parties are welcome to participate: free-lance photographers, small and large picture agencies and libraries, and trade associations from the photo business.

April 7: Google Grant Supports Developing EXTRA Classification System

The International Press Telecommunications Council (IPTC) will use a grant from the first round of Google’s Digital News Initiative Innovation Fund to build and freely distribute an initial version of EXTRA: The EXtraction Rules Apparatus, a multilingual open-source platform for rules-based classification of news content.

EXTRA will be a classification system for annotating news documents with high-quality subject tags. Such tags will allow publishers to deliver a variety of valuable services including content recommendations, improved advertising targeting and subject-specific content streams, such as alerts and topic pages.

“By creating a freely available rules-based classification engine, IPTC will help publishers to enhance their content with all sorts of metadata services, including enriched search, intelligent recommendations and precise analytics,” said Stuart Myles, chairman of IPTC.

EXTRA will provide news publishers with several key capabilities: the ability to automatically categorize documents by subject (for example, terrorism, sports, names of celebrities); the ability to author classification rule sets tailored to existing taxonomies; and the ability to classify documents using the industry standard IPTC Media Topics taxonomy. Taxonomies are used by many news organizations to classify their content. Classification is used in various ways, including improved online news navigation by grouping and linking, to organize editorial workflows and to enrich search.

The 15th International Semantic Web Conference

http://iswc2016.semanticweb.org/
• IPTC schemas (rNews, NITF, NewsML, SportsML, etc), schema.org, and the Open Graph Protocol are story oriented.

• IPTC schemas NITF, NewsML-G2, SportsML-G2, Subject Codes, and others facilitate the exchange of stories between news publishers.

• rNews (and its predecessor hNews), ninjs, schema.org are designed for web publishing and for web discovery.

• The Open Graph Protocol OGP enables any object to become part of a social graph.

• Library metadata for digitized historical news(papers) is primarily issue oriented (some libraries do divide issues into articles).

• Libraries that harvest digital news seem mostly to collect websites, not individual articles (exception: National Library of Sweden collects RSS feeds).
References

• Facebook Instant Articles: https://developers.facebook.com/docs/instant-articles
• IPTC Standards: https://iptc.org/standards/
• https://schema.org
• MODS schema and documentation: https://www.loc.gov/standards/mods
• The Open Graph Protocol. http://ogp.me/
• RSS Specifications etc. http://www.rssboard.org/
Frederick Zarndt
IFLA Governing Board
Digital Divide Data
frederick@frederickzarndt.com
@cowboyMontana