

# What on Earth are... Microformats?

Want to embed even more information into your HTML? Of course you do! **Brian Suda** gives the lowdown on a set of data formats to make content richer.

» **Just when the world needed another Web 2.0 buzzword, I've been hearing about something called 'microformats'. Can you explain to me what they are?**

Microformats give additional meaning to simple, widely-published types of data on the web. For instance, they can be used to add more structure to contact information about people and places, and to calendar information for events and venues. While the term 'microformats' does get lumped in with other Web 2.0 buzzwords, the idea of encoding more meaning into HTML has been around since before 1997.

» **Are microformats another language I'm going to have to learn? I'm already busy with C# and Tcl...**

No, microformats are human-readable and work within HTML 4.01 and XHTML. They are built

» **Tell me a little about how they started.**

There is a long history of microformats before they were called microformats. In 2002 several smart folks – Tantek Çelik, Matthew Mullenweg and Eric Meyer – started work on XHTML Friends Network (XFN). Much of their work is available at Global Multimedia Protocols Group (<http://gmpg.org>), which has the goal of "Connecting people through incremental simplicity". This idea of simple, incremental additions of semantics later became much of the basis of microformats.

» **Why is it so important that the data be human-readable?**

Data that is human-readable is very important for several reasons. What is out of sight, is out of mind. Data you see through the browser window every day is more likely to be kept current than data in a flat file on the server that you have to download

wanted a vCard for them. I would need to use some application to convert your XML files to vCards. See? Without the information being available right there in the HTML, your back-end system would need to start publishing 30 different formats to keep all of your customers happy! By adding microformats, it is possible to extract the data from the HTML into any format the user wants – even ones you don't think are important or are not aware of. It's about open data.

» **Are there different 'microformat' formats?**

Certainly. When microformats started, the creators began by looking at what people were

**"That all sounds great, but does anyone own the microformat technology? Has Google bought it yet?"**

around existing standards, so the only things you have to know are how to code HTML and some of the element attributes.

» **So microformats are just CSS then?**

Not really – microformats make use of the class attribute in HTML, and so does CSS. And that's pretty much where the similarity ends. CSS is a set of rules for how to *style* content. Microformats are all about giving more meaning to already-existing content. The class attribute is a general-purpose attribute that the W3C added into HTML to allow for new ideas (such as microformats) to express metadata.

» **How does one go about adding microformats into HTML?**

There are three major places where you can encode additional meaning: the meta element, class attribute and the rev/rel attributes. Because microformats are pushing for content to be visible, they focus on the last two places where semantics can be added, the class and rev/rel attributes.

and import into some additional application to view. If the data is visible, it will be fresher and more useful to everyone.

» **Well, all my data is powered from a back-end content management system, so it's as fresh as the last publish. Why should I care about using microformats?**

It's true that a lot of people use database or XML back-ends and publish their sites regularly. This does solve the problem of keeping non-HTML files, like RSS, vCards and iCalendar files, all up to date with the HTML, but it doesn't give any added value to your HTML. The average web surfer probably doesn't have access to your back-end database or your source XML files unless you provide some sort of API. By adding microformats to your HTML, the page itself become the API.

» **My site doesn't do anything exciting – I can't see why would I need an API.**

Let's say your site published an XML file for each staff member and my importing application



publishing on the web. Some of the more popular things people publish is information about people and places, and events. From that, two of the more popular microformats emerged. They are hCard, which models vCards and describes people, places and organisations; and hCalendar, which models iCalendar and is used to describe events. Other formats handle CVs ('hResume'), reviews (for products, places, websites, films etc), citations, syndication and much more.

## » Are RSS and Atom microformats?

No, they are plain old XML formats. There is a microformat, hAtom, that is used to add meaning into HTML so that it can be converted into a corresponding RSS or Atom feed. Microformats encode more semantic structure into the existing HTML that is published over most of the internet.

## » OK, so all those sites that only publish older versions of RSS – you're saying that if they had marked up their HTML I could convert that to Atom 1.0?

Now you're getting it. Those site owners probably paid some freelance developer on a short contract to create the site and implement a feed. At the time Atom might only have been at version 0.3, so

that's what the developer would have used. Now Atom has reached 1.0, the developer has long since completed their contract and there is no money in the budget to upgrade, so the site is stuck on 0.3. Microformats could have prevented this problem by adding the necessary meaning so that the data could be extracted into just about any format you need, even at Atom 0.3.

## » Ah, so it is about adding more semantics! I've been wanting to model my 'stamp collection' online, how do I do that?

One of the key points of microformats is that they are not infinitely flexible. Their goal is not to solve all of the world's problems. They tackle very specific problems that already exist on the web today and try to give additional structure to the data. The initial formats that are available solve as much with as little effort as possible. They are based on what millions of sites are already publishing – no need to reinvent the wheel, just structure it a bit more so we can all interoperate.

## » Who actually uses microformats?

Lots of websites do. By tweaking an HTML template slightly and republishing, many sites can create thousands, if not millions of instances of microformats in just a few minutes. Sites like Yahoo Local, Yahoo Tech, Flickr, Meet-up, Upcoming, Eventful, Edgeio, Technorati, Ice Rocket, Apple's .Mac mail and others all use various microformats.

## » That all sounds great, but does anyone own the microformat technology? Has Google bought it yet?

No one owns microformats. A large group of volunteers and a wider community keep the discussion about new formats going. They also contribute to the upkeep of existing formats and documentation, all of which is under a very liberal licensing scheme. Since no company or individual owns the technology, microformats are not hampered by corporate decisions or red tape. It's just about an open community producing open data formats, which make the internet a better place for everyone to communicate.

## » Very inspirational, I must say! Where can I learn more?

The official website for microformats is <http://microformats.org>. There you will find a wiki, mailing list, blog, links to the IRC channel and much more. There are also several independent sites dedicated to microformats – <http://microformatique.com>, [www.whymicroformats.com](http://www.whymicroformats.com), and other sites that aggregate sites talking about the topics such as <http://del.icio.us/tag/microformats>, <http://technorati.com/tags/microformats> and <http://ma.gnolia.com/tags/microformats>. Even Microsoft's new Live Clipboard (<http://snipurl.com/10so0>) and Live Writer (<http://windowliveswriter.spaces.live.com>) use microformats. **LXF**

