



Tantek Çelik and Rohit Khare: The Progress and the Promise of Microformats

Published : July 20, 2006 in [Knowledge@Wharton](#)

The microformats movement was officially launched with the unveiling of the Microformats.org website one year ago at Supernova 2005. At that time, [Knowledge@Wharton spoke with Tantek Çelik](#), one of the founders of Microformats.org, about his vision for a more flexible worldwide web with content that can be easily interpreted, collected, and repurposed for other applications.



This is a single/personal use copy of Knowledge@Wharton. For multiple copies, custom reprints, e-prints, posters or plaques, please contact PARS International: reprints@parsintl.com
P. (212) 221-9595 x407.

Microformats are simple extensions to the standard HTML tags used to create web pages. By including the additional microformat markup, web pages go from merely presenting the visual display of content to embodying its meaning. When a traditional web page contains information about an event, for example, the HTML markup conveys little more than the formatting of the text describing the event. But the addition of microformatting can unambiguously identify the date, start time, end time, and venue for the event. With microformat extensions added to the HTML tags, software can add the event to a personal datebook, aggregate content from different web pages into a comprehensive calendar, or let people "mash up" the content in new ways such as adding events to online maps or other web pages.

The microformat movement has been gathering steam since launching one year ago. In this year's Supernova 2006 workshop on "Decentralizing Data," Çelik stated that the number of microformatted entities on the web has increased from a few thousand a year ago to tens of millions today.

At that same workshop, Yahoo! Inc.'s Andy Baio, announced that Yahoo Local is supporting microformats "in a big way," with all business listings, search results, reviews, and events marked up with microformats. According to Baio, Yahoo is now the biggest supporter of microformats on the web. Acknowledging the tendency of Internet companies to constantly outdo each other, Baio wondered aloud whether this would be the first round in a forthcoming "arms race" to microformat the web. His take on this possibility -- "I absolutely hope so. I hope that tomorrow Google, Microsoft, AOL, and eBay [include microformat content]. Because everybody benefits."

At Supernova 2006, Knowledge@Wharton spoke with two of the leading evangelists for microformats -- Tantek Çelik, chief technical officer of Technorati, and Rohit Khare, director of CommerceNet Labs -- on how microformats have progressed over the past year and the issues the movement faces going forward. An edited version of that discussion follows.

Knowledge@Wharton: You launched Microformats.org one year ago at Supernova 2005. What's happened in the last year?

Tantek Çelik: A year ago, we had drafts for a bunch of standards for interoperably exchanging information about contacts, events, reviews. Now those standards have reached the fairly reliable specification state. More importantly, we have numerous publishers publishing their content in microformats, such a Yahoo's recent announcement of [using microformats for] all of Yahoo Local. Flickr profiles are all hCards; EVDB and Upcoming's events and venues are all marked up with hCalendar, hCard. [The web site] Judy's Book marks up all their reviews with hReview and lets you blog your own reviews as well. And Edgeio just launched 465,000 hListing microformatted pages.

Rohit Khare: That last example is interesting because even an experimental microformat is now picking up adoption rapidly. hListing is an effort that was launched by some folks at [the classified ad search site] Oodle along with work at CommerceNet and Edgeio when they observed, "Hey, there's something going here because lots and lots of web sites use classified ads of one sort or another."

A second [capability enabled by] microformats are cases where you'd like to share and reuse that data. Microformats isn't just about saying *your* site should have more structured templates, it's that your site might want more structured templates so that it can better interoperate with other sites.

Knowledge@Wharton: Just so it's clear -- what difference does it make when a site uses microformats? What functionality is enabled by using microformats that wouldn't be possible if the site was just using HTML?

Çelik: There's a practical example for people here at the conference -- the Supernova site has published its pages marked up with microformats. The list of venues for Supernova -- both Wharton West and the Palace Hotel -- is marked up with hCard. So with a single click on the link that says "Add these venues to my address book," they get converted into the standard vCard format, which the computer can easily import into anybody's address book and then synchronize with their PDA or Blackberry or whatever.

So when you're walking around and you're [thinking], "Where's that venue?" -- it's in [your PDA] because it was in a microformat on the web page.

Similarly, the workshop schedule and the events schedule are marked up in hCalendar. So everybody can easily add them to their calendar on their laptop or their PDA.

This was really easy for [Supernova] to do, because it is just in the web page, rather than having to create and maintain a separate file. They used a converter that is built on open source software that Technorati is hosting which converts hCalendar to iCalendar or hCard to vCard.

Knowledge@Wharton: Would it be fair to say that we're seeing a transition of the web from a viewable, presentation format to something you can interact with to extract meaningful pieces of information in a very granular way that you can then take and use to do other things with?

Çelik: That's exactly right. Instead of just having the "corporate brochure" web, where it's this nice, beautiful flyer you can look at and print out, we're now talking about this enhanced "data plus presentation" web, where not only is it beautiful in presentation, but the information is also available to computers to transform into different forms for you to use in different applications, like an address book or a calendar, or on different devices.

Khare: It's a bit like that maxim about "reduce, reuse, and recycle."

Microformats help reduce the effort versus having the separately publish [a parallel data format like] vCard files; they reuse the information by allowing users to cut and paste [the data]; and they recycle in that it becomes part of the input stream to the next round of innovation. Now that Yahoo Local is publishing hCards for businesses, you could imagine the equivalent of cutting them out of Yahoo and pasting them onto your calendar or taking those addresses and mapping where you're going to be on your next trip.

All those kinds of [functions] are where the frontier lies for the second year.

The granddaddy of these formats are hCard and hCalendar. They're useful precisely because they map existing ten- to fifteen-year old standards for cards and calendars into HTML. Microformats' hCalendar is not a new, *de novo* calendar format. It's a way of having your ordinary web page simultaneously encode information for applications like [IBM's Lotus] Notes and [Microsoft] Outlook and [Apple's] iCal on the Mac.

The new frontier includes things like hListing that build upon that further. A classified ad might have an hCard saying who's selling it, and an hCalendar-type entry that says how long the ad is good for, but it also has some of its own fields, like whether it's for sale or whether it's a wanted ad.

Being able to cut, paste, and recycle these larger units of information is this next frontier, because it may not be [just] that I want to add it to my date book but also [to enable] new kinds of services.

That's another thing I'd like to emphasize about the accomplishments in the first year. We're getting past the novelty of simply saying, "Hey, I can go to a single page, extract it, and add it to my notebook." I can

now have crawlers like Pingerati, Technorati Kitchens, and CommerceNet's own Microsearch Experiments crawl the web, aggregate [the content] in one place, and let you find addresses or reviews and, ultimately, do a better job of search than we've done before.

By [searching for] "New York Pizza" -- Is that going to get me a "New York" [style] pizza shop here in San Francisco or a pizza shop in New York?

Çelik: Because you might want either one. By marking up this information semantically with the hCard microformat, it allows search engines like Technorati's Microformats Search to be able to tell you, "Hey, this is a New York style pizza location in San Francisco versus a California Pizza Kitchen restaurant in New York."

And the amazing thing that's happening is that individuals can now publish information [like this] about restaurants they've visited. The restaurant itself doesn't need its own web site. The individual can just blog it and publish information themselves on their own [web] page.

What you have here is essentially the world's first decentralized, distributed address book. The potential of that is something that I don't think we've really grasped yet.

Imagine having the yellow pages and white pages of every single city in the U.S., in the *world*, at your fingertips. Imagine having anybody's notes and scribbles about every favorite restaurant that they've ever visited also at your fingertips. And you start to realize that, wow, this is starting to make it much easier to find things anywhere in the world.

Knowledge@Wharton: Doesn't this create a challenge for people who are publishing content and their business model assumes that the web page with the content is the destination page, and part of their goal is to get you to that page to see it in that context, with their branding. Isn't the functionality of microformats working against that for some publishers?

Khare: It's a bit like the relationship that content providers have with search engines. Enabling technologies like search ultimately helps users get to the content they need [and that has] value and will get compensated in the right ways. It doesn't always have to be an eyeball-page-view of a site.

Çelik: The more you publish your content in such a way that it can be indexed by search engines the more traffic you get to your web site. That's a very well understood relationship. There's a whole industry called search engine optimization

Khare: [An issue] that we emphasized in this year's [Supernova] workshop on Decentralizing Data, is that we should share political control when you decentralize something.

The traditional model right now is that there's a public web. And, so, if not me, somebody else can put the money into spidering all possible reviews of all pizza restaurants and say, "Just come to our mega-portal site for all known 'pizzas are us' and we'll have the info."

But the real world has private conversations, too. We live in these blended overlapping circles of identity and trust. [There is] some data [public portals] won't ever have access to. They won't have that email from my brother-in-law saying, "Hey, I just visited San Francisco and I found this cool pizza restaurant." Being able to join that data -- not because it happens to have the keywords "pizza" and "New York" in it -- but because I see that the same hCard exists, that ability to use microformats as a bridge between the public web and private is, I think, part of the new frontier as well.

Decentralization means you're doing something that no search engine can do alone.

Knowledge@Wharton: What's the biggest challenge you faced over the last year? What didn't work out as well as you had hoped?

Çelik: I think the biggest challenge is keeping up with the level of interest and the number of different people wanting to try new and different things. There are so many different pieces of information that people want to create new microformats for. One of our biggest challenges has been shepherding folks through the microformats process, which works hard to insure that microformats stay "micro"; [to insure that the] formats themselves stay very small and simple, and the total number of formats stays quite small

as well, so that anyone who's looking to publish information doesn't need to worry about the fact that there a thousand microformats out there. Right now, there are probably a dozen or so that are solid enough for people to use on a regular basis. There are probably another two dozen that are in experimental development. And we like it that way, because that makes it possible for like the average web author to understand this stuff and to put it in their code -- which doesn't require a programmer.

If you think about it, there are hundreds, perhaps thousands, more people who understand how to make a little bit of HTML than there are programmers out there. And that's been another key to the success of microformats -- we designed them for adoption by a much broader base of people than just programmers and IT professionals.

Khare: I concur that the biggest challenge of the past year is around process. Not just educating people about microformats, but also educating them about why we have a process that's fairly conservative about creating new microformats. That, if what you want to talk about is a new kind of classified ad listing, you'll get a hearing and you can experiment with it. But if you come around and say, "Hey, I'd like to do a new microformat for electronic prescriptions and health records," you probably won't.

And that's actually part of this fine line.

Knowledge@Wharton: If we talk again in a year at Supernova 2007, where do you think you'll be then?

Khare: A problem we might face for 2007 is making sure that people who get enthusiastic about microformats have a way to participate and join in.

But, also, [communicating that] when people have an idea for a brand new microformat where they say, "Hey, look, at these six sites that are doing it", it doesn't always have to become a new microformat. The challenge will not be going from four or five to ten or twelve [microformat specifications] as we've done in the past year, but making sure that we go from ten or twelve to maybe twenty or twenty-five, but never to a hundred or two hundred.

Çelik: Let me give you another challenge we're going to face in the next year. Now that we've firmly established a really core set of microformats that are reliable and supported interoperably between a bunch of publishers and applications and search engines, I think that the next year our biggest challenge is going to be in improving the user interfaces so the average person can create their own microformatted information without even worrying about the fact that it's a microformat.

Your typical blogger doesn't worry about the fact that they're typing in HTML -- because they're not, they're just typing into a form field. It just works.

We want the same simplicity of publishing so individuals can create their own structured information and can share it with each other. That's a really important piece of it.

Some of the work Ray Ozzie's team is doing [at Microsoft] regarding the Live Clipboard is exactly the kind of thing we want to see deployed widely in the next year -- being able to go to any page that has microformatted information and you don't have to think, "Oh, it's got microformats." You just see, "Oh, it's got a picture," "It's got a contact," "It's got an event," and those somehow indicate that you can copy and paste them as an entirety into another web site or another desktop application.

And once we get to the point of that kind of copy-and-paste, publish-and-subscribe model that users can use with a couple of clicks here and there, that's going to be when we know that we've really connected with users, when anybody can just use the stuff and not have to worry about calling them microformats anymore. That will just be for us geeks.

Khare: We will have succeeded when no one cares.