Microformats: what are they, and why should we use them?

Dan Connolly
Overview

• Microformats in action
  – How to *(just a little)*

• Principles, process
  – A little history

• Beyond Microformats
  – Limitations: what doesn't work so well
  – Native XML
  – Semantic Web (RDF, OWL)
Acknowledgements

- microformats.org
- Tantek Çelik
The Personal Information Disaster

- Document exchange in the Web mostly works
- Not so for data
  - soccer schedule
  - flight itinerary
- Copy and paste one field at a time
  - Isn't that what computers are for?!
Microformats to the Rescue

• Integration with personal data tools
• calendar subscription with hCalendar
• contact import with hCard
Events on a Web Page

Microformats at the Web 2.0 Expo

The Web 2.0 Expo is taking place at Moscone West on 747 Howard Street in San Francisco, California from April 15th to 18th. Microformats will be well represented.

John Allsopp, author of the newly published Microformats book from Friends of ED, is scheduled to speak on Tuesday. John’s presentation is called Microformats, Much More Than Just Promise. The time is currently set for half past twelve. John will be looking at current implementations of microformats and is asking what applications remain unexplored.

There’s also a presentation called The Beauty in Standards and Accessibility on Tuesday at 3:45… I’m sure microformats will get slipped in there at some stage.

The schedule for the conference seems to be still in flux so keep your eyes...
hCalendar bookmarklet

• Bookmarklet = javascript powered bookmark

• Technorati service converts hCalendar to .ics format used by desktop tools
Rescued!

- No manual copy-and-paste
- Subscription with live update
John's presentation is called Microformats, Much More Than Just Promise. The time is currently set for half past one. John will be looking at current implementations of microformats as well as asking what applications remain unexplored.
hCalendar deployment

• hCalendar creator
• Dreamweaver plug-in
• iCal Extraction
• large and growing list of sites:
  – Eventful.com
  – LinkedIn
  – Yedda
  – Yahoo! Tech Reviews
hCalendar = iCalendar * HTML

- vevent {1}
  - category*
  - class?
  - description?
  - dtend (ISO date)?
  - dtstart (ISO date) {1}
  - duration?
  - location?
  - status?
  - summary {1}
  - uid?
  - url?
  - last-modified?
hCard = vCard * HTML

• vcard {1}
  – fn {1}
  – n?
    • honorific-prefix*, given-name*, additional-name*, family-name*, honorific-suffix*
  – sort-string?
  – title*, org*
  – email*, tel*, url*
  – adr*
    • type, street-address?, locality?, region?, postal-code?, country-name?
  – photo*
  – ...

...
Microformats Principles

• solve a specific problem
• simple as possible
  – evolutionary improvements
• humans first, machines second
  – presentable *and* parsable
  – adapt to current behaviors
• reuse from widely adopted standards
  – semantic (X)HTML, schemas from interoperable RFCs
• modularity / embeddability
• decentralized development, content, services
hCard Creator

- http://microformats.org/code/hcard/creator

### hCard-o-matic

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>given name</td>
<td>John</td>
</tr>
<tr>
<td>middle name</td>
<td></td>
</tr>
<tr>
<td>family name</td>
<td>Doe</td>
</tr>
<tr>
<td>organization</td>
<td></td>
</tr>
<tr>
<td>street</td>
<td>123 Main Street</td>
</tr>
<tr>
<td>city</td>
<td>Anytown</td>
</tr>
<tr>
<td>state/province</td>
<td>NY</td>
</tr>
<tr>
<td>postal code</td>
<td>USA</td>
</tr>
<tr>
<td>country name</td>
<td></td>
</tr>
<tr>
<td>phone</td>
<td></td>
</tr>
<tr>
<td>email</td>
<td></td>
</tr>
</tbody>
</table>

### code

```html
<div id="hcard-john-doe" class="vcard">
  <span class="fn">John Doe</span>
  <div class="adr">
    <div class="street-address">123 Main Street</div>
    <span class="locality">Anytown</span>
    <span class="region">NY</span>
    <span class="postal-code">USA</span>
  </div>
</div>
```

### preview

John Doe  
123 Main Street  
Anytown, NY, USA  

This hCard created with the hCard creator.
hCard sign up

- to fill out such basics as your name, your user icon
- use your existing hCard
  - from Cork’d, Last FM, Flickr, Technorati, Twitter, Yedda etc.
  - like dropping a business card
Satisfaction?

Add your twitter user name
http://twitter.com/
factoryjoe

Screen Name
(for public display)

Email Address
We will never share your email with anyone. Period.

Password
7 characters or more

Repeat Password

User Icon
(optional)

Choose one of the following supporting sites
flickr
technorati
upcoming
last.fm
twitter
cork'd
other...
XFN – social networking

- <a href="http://jane-blog.example.org/
  rel="sweetheart date met">Jane</a>
- <a href="http://dave-blog.example.org/
  rel="friend met">Dave</a>
- <a href="http://darryl-blog.example.org/
  rel="friend met">Darryl</a>
- <a href="http://www.metafilter.com/">MetaFilter</a>
- <a href="http://james-blog.example.com/
  rel="met">James Expert</a>
XFN+hCard social network import

• instead of having to manually find and re-add everyone to yet another site...

• import your contacts from any other site that publishes your contact list with hCard for the people and XFN for your relationship(s) to them
XFN+hCard import: Dopplr.com June 2007

- travelplan sharing site
- currently invite-only

Fellow travellers

The following travellers are listed as contacts in http://www.twitter.com/factoryjoe.

- Andy Baio
  - San Francisco
  - Share trips

- Andre Charland
  - Vancouver
  - Share trips

- Brad Fitzpatrick
  - Share trips

- Ben Cerveny
  - Los Angeles

- Andrew Crow
  - San Francisco

- Amy Hoy
creative commons search with rel-license

• I need an image of a koala for my presentation...
• How do I know if I have rights to reuse all these photos?
rel-license support

• Creative Commons (cc) license chooser
• Yahoo! (cc) search
• Google "Usage Rights" search
• Trivial* for publishers:
  – Copyright © 2007 <a href="http://creativecommons.org/licenses/by/2.0/" rel="license">some rights reserved</a>
  – * assuming a license is chosen
More Microformats

- XOXO – outlines, blogrolls
- hReview – product reviews
- hListing – item, price, ...
microformats - just better markup?

• built on valid semantic HTML (POSH)
  – as with HTML, broadly human read/writable/editable
  – easier maintenance, updates, collaboration, team work
  – better default behaviors from more user agents
  – accessibility - screen readers
  – diverse device support - mobile
  – search engine optimization

• more than just "good class names"
  – principles keep things "micro"
  – process emphasizes getting real
  – community minimizes duplicates
Why not XML?

• Why encode this in HTML?
• Why not use an XML vocabulary for contacts and calendar information and ...?
Why not XML? (Ogbuji)

• "XOXO is far more inscrutable and harder to process than almost any XML you might design to replace it."

Thinking XML: Microformats the XML way
Uche Ogbuji, 15 May 2007
Validating Microformats (Walsh)

• The use of token lists in attributes “blows direct RELAX NG validation completely out of the water.”
  – Direct W3C XML Schema validation is “impossible too.”

• Schematron might work
  – But you have to express both what is and what isn't allowed
  – And you have to mess with all that token parsing

Validating microformats, Norm Walsh, Apr 2006
XML Calendar Vocabulary

• **Guideline for use of XML with iCalendar elements**
  – Internet Draft 2004-2005 by T. Hare
  – Some support in Mozilla Sunbird
    • via XSLT
  – Expired

• **Technically sound, but not widely adopted**
  – insufficient advocacy?
  – insufficient benefits?
  – *insufficient integration with Web architecture?*
“XML has had many successes. I just finished writing a 400-page book in DocBook XML. It is over one megabyte of well-formed XML, and I use off-the-shelf tools to transform it into HTML, PDF, and a number of other formats. This was one of the original use cases for XML, way back in 1997 when SGML ruled the Earth and dinosaurs like ISO-8879 were gospel. It works; I love it; I recommend it for any serious content creator.”

XML on the Web Has Failed  Mark Pilgrim, xml.com, 2004
Why not XML? (St Laurent)

• “XML has occasionally found its way to the Web, but it's hard to remember now that once upon a time, XML was supposed to be directly on the Web, the files people loaded and manipulated...

• “The particular XML Web described by Bosak and Bray never happened. (It still could, but hasn't.)”

The Next Web?
Why not XML? (Çelik)

• XML formats in the long run are *no better than proprietary binary formats*.
  – XML, both in technology (namespaces...) and as a "technical culture" is too biased towards Tower of Babel outcomes.
  – A few XML formats may survive and converge (RSS, maybe Atom)
    • but for now XHTML is the only longterm reliable XML format
    • That has more to do with it being based on HTML than it being XML.

Open data formats, longevity, and microformats
June 2006
Tantek Çelik
Living without XML Validation

• Simplicity
  – view source, eyeball it
  – lots of peers to ask

• Testing
  – manual testing with consuming tools
  – heuristic validator
    • beta
    • x2v “pre-flight”
  – automated tests

• ... Cross that bridge when we come to it
Limits of Microformats

• The microformats process applies to common problems...
  – ... and *only* common problems.

• Mixing works in some cases
  – but don't push it
  – arbitrary mixing runs counter to “solve a specific problem”

• Validation is more art than science

• Query tools are yet to mature

• *So what if my problem is not so common?*
Web Architecture scales down as well as up

- Microformats covers 80%
- Web Architecture scales
  - up to billions of documents * users
  - down to family photo albums
Web Architecture for Documents

- Simple Hypertext
- + URIs for global naming
- - link consistency guarantee
- = viral growth
Web Architecture for Data

• Spreadsheets and Databases
• + URIs for global naming
• - some consistency guarantees
• = viral growth?
To a computer, then, the web is a flat, boring world devoid of meaning...This is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them.

- Tim Berners-Lee, WWW1994
Semantic Web Building Block

- `<schedule#e1> cal:organizer <staff#ted>`.
Semantic Web Technologies

• Technologies Maturing
  – Web Ontology Language (OWL)
    • Medical Ontologies, drug discovery, ...
  – SPARQL Query Language and Protocol
    • Wikipedia query, ...

• Based on Resource Description Framework (RDF)
  – URI-based data merge is built-in
RDF is written in XML

• Too constrained for some
  – Tries to look like Book/author/title metadata but has subtle constraints with striping, rdf:about, rdf:resource, etc.

• Not constrained enough for others
  – doesn't work with DTDs nor W3C XML Schemas; works awkwardly as input to XPath and XSLT
RDF Calendar

• Oct 2002: Workshop on the Semantic web and Calendaring, Bristol, UK
• Sep 2005: RDF Calendar - an application of the Resource Description Framework to iCalendar Data W3C Interest Group Note, Connolly and Miller
• Supported by a handful of tools
• Adoption prospects colored by RDF/XML syntax
Microformats as Semantic Web Data

- hCalendar is XML
  - at least: if you use XHTML
- RDF Calendar is XML
- XSLT does XML to XML pretty well
  - glean-hcal.xsl
- GRDDL profiles link microformat documents to XSLT transformations
Semantic Web Mash-Ups

- Merging is automatic
Summary

• Microformats provide simple solutions to some common problems
  – calendar, contacts, ...

• XHTML is a good place to start
  – XML tools
  – HTML shared vocabulary

• XML vocabularies can be just as proprietary as binary formats

• Web Architecture scales down as well as up
  – URIs, namespaces

• Semantic Web data merges automatically