Microdata & Cites

Finally, a Metadata Standard for HTML
Goals of a Good eDocs

- Human presentable format that is also machine readable.
- Documents that would be able to be identified and found with unique identifiers
- Document fragments that were easily addressable and usable
- Documents that were well structured
- Documents that could have portions that could be semantically addressed and used
- Documents with expressed and available schemas
- Collections of documents, documents and document fragments that could be seen as an object database
Microdata w/o Preconceptions

- Schema.org is not the only way to use Microdata, but it is great for SEO as Google take advantage of it.
- Microdata “items” are addressable in the DOM and by XQuery (if in well-formed HTML5)
- RDFa and RDFa Lite are not just like Microdata, in that they miss a couple features and depend on ontologies. Which is not to say that some prefer RDFa
- Microdata has a magical “itemid”
- Microdata schemas may not yet be like XML Schemas/XSD, but could be continue
Microdata w/o Preconceptions

continued:

- Microformats are not Microdata (see microformats) and miss some key features of Microdata
- Microdata makes composite documents more easily use different “schemas” without unwieldy DTDs or namespaced weirdnesses of XML Schemas
- Microdata does not replace all XML utility, but may allow XML languages to enhance HTML and perhaps allow full two way XSLT
Birth of Microdata

SGML → HTML

SGML → XML

Microdata
Birth of Microdata

Human Presentable Layer seen in a browser

Machine Processable Layer seen in view source

SGML → HTML

Microdata

SGML → XML
Precedents & Alternatives

- XML xslt’d into HTML in browser
- Microsoft Smart Tags in Word and Internet Explorer
- Microformats
- RDF/RDFa
- HTML Meta tag
- JSON
## What Comprises Microdata

<table>
<thead>
<tr>
<th>HTML Attribute</th>
<th>Optional</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>itemscope=&quot;&quot;</td>
<td>N</td>
<td>indicates the tag that is the item or object, can be whole document or nested fragment</td>
</tr>
<tr>
<td>itemid=&quot;URL&quot;</td>
<td>Y</td>
<td>points to the original item, including itself if it is the original one</td>
</tr>
<tr>
<td>itemtype=&quot;URL&quot;</td>
<td>Y</td>
<td>points to document that sets out the description, similar to a namespace in XML</td>
</tr>
<tr>
<td>itemprop=&quot;name&quot;</td>
<td>N</td>
<td>at least itemprop. Itemprop indicates a key pair, with the name and value (value may be within another attribute or between tag begin/end of the itemprop’ed tag)</td>
</tr>
<tr>
<td>itemref=&quot;id&quot;</td>
<td>Y</td>
<td>used like itemprop, but uses id attribute to find the key pair outside of the itemscope. This is mainly to deal with tables in HTML.</td>
</tr>
</tbody>
</table>
At human presentable, browser layer it is not seen except through CSS and Javascript.

In HTML, any element including <html> can include Microdata

<HTML5tag itemscope="" itemtype=""http://designatedschema.org/typeofitem"" itemid=""http://thesitethathasdefinitiveversion.org/theorginalitem.html">

<div itemprop="name">This is the <strong>name</strong>. </div>
<div><a href="http://url.org/url.html" itemprop="url"> Click here</a></div>
<p>This is a paragraph. <span itemprop="importanttext">Important stuff here. </span>and on. </p>
</HTML5tag>
- Buttermilk Pancakes: http://www.marthastewart.com/318689/best-buttermilk-pancakes
- My deck of cards and two hands: http://citizencontact.typepad.com/xquery/cards-microdata.html
Within a web document to be cited and cite to other documents:

- Allows for structure to be added that unlike HTML tags, class, data-attributes, can be self-documented with itemtype within document
- Might be useful to use HTML root tag to designate the whole document
- Allows document to “cite” other documents using more than link and/or poorly embedded metadata
  - Even makes HTML <cite> tag usable...finally
  - Use the <meta> tag to have several metadata key pairs embedded without being seen
Microdata for Citations

- Use the Microdata also for SEO/Rich Snippets and other Microdata itemtypes
- Finally, semantic capability using other “schemas” like USGS for geographical terms, etc, financial data/appropriations info piggybacking on XBRL
- Gives hints for searching, ingesting/scraping multiple documents and indexing entire collections
- Itemtype allows the schema to be available for aggregators, value add folks to extend and enhance versions, yet “cite” back to original
Microdata for Citations

- Itemid can make it easy to enforce copying link back/cite back, and “share” capabilities in combination with id attributes for document portions/fragments
- Allows much of the same analysis and use cases as an XML version, but with easily browsed and linked to capability, usability and accessibility
- Could easily allow embedding Zotero/CSL types and key pairs, AKN structure, Dublin Core metadata, FRBR info and linking other versions like PDF and XML
- Easy CSS and Javascripts for applications and design based on Microdata
Great Explanation of Microdata
http://diveintohtml5.info/extensibility.html

Chrome Browser Extension that shows Microdata “items”
https://chrome.google.com/webstore/detail/microdatareveal/olapakiakblfdaajcifqldandnikpdh

Specifications

Google Help Page and Rich Snippets Page
https://support.google.com/webmasters/answer/176035?hl=en
https://support.google.com/webmasters/answer/99170?hl=en

Schema.org explanation of Microdata and links to using Schema
http://schema.org/docs/gs.html
Daniel Bennett
daniel@citizencontact.com
@citizencontact