DITA, Open Semantics, and XHTML Microformats - Opening remarks

Michael Beaver and John Hunt
Lotus Information Development Center, IBM
For presentation to OASIS DITA for Web Sub-committee
January 13, 2010
Why DITA?

Darwin Information Typing Architecture

DITA Fundamentals:

- **Topics** – a content object
- **Maps** – a collection of topics
- **Metadata** – available to both topics and maps
- **Specialization** – built-in extensibility for topics, maps, domains, and metadata
Specialization provides semantic content types

- Concept
  - Tutorial
- Task
  - Command
- Reference
- Learning
- Troubleshooting
  - Event announcement
    - Insurance claim form
    - Use case specification
      - ...

Use the right topic type for your content

- API
  - C++ API
  - Java API
Open Semantics

We currently suffer a tremendous amount of semantic loss when DITA-sourced information is published.

- This is partly due to the exclusion of those semantics in the output, but also due to the lack of support within information hosting environments to exploit those semantics.

Looking to the world of microformats, there might exist a way to strike a balance between “on the glass” and “on the class”

- Portable and distributable semantics
- Larger universe of tools and users can consume your structured content with the meaning still intact
- Brings the semantic power of DITA to a larger community
How about an example? Meet hCard

- Makes it easy to include and share contact information with humans and machines.
- Includes the semantics about contact information in existing XHTML building blocks.
- Derived from the IEEE vCard standard.
- In an environment that is not hCard aware, the “on the glass” still works and displays as intended to the user. However, when placed into an environment that is hCard aware, interesting things may happen thanks to the semantics included “on the class”.

```html
<div class="vcard">
  <a class="fn org uri" href="http://www.ibm.com">IBM</a>
  <div class="adr">
    <span class="type">Work</span>
    <div class="street-address">4205 S. Miami Blvd</div>
    <div class="extended-address">Building 502, Office M210</div>
    <span class="locality">Durham</span>
    <span class="region">North Carolina</span>
    <span class="postal-code">27703</span>
    <span class="country-name">USA</span>
  </div>
</div>
```
A DITA Task and a microformat “hTask” equivalent

DITA Task

```xml
<task id="task-lesson-serving">
  <title>Serving it up</title>
  <taskbody>
    <steps>
      <step><cmd>Place…</cmd></step>
      <step><cmd>These:</cmd>
        <choices>
          <choice>PB & J sandwich</choice>
          <choice>Cookie</choice>
        </choices>
      </step>
      ...
    </steps>
  </taskbody>
</task>
```

XHTML Task (hTask)

```xml
<div class="task" id="task-lesson-serving">
  <h2>Serving it up</h2>
  <div class="taskbody">
    <ol class="steps">
      <li class="step">
        <span class="cmd">Place…</span></li>
      <li class="step">
        <span class="cmd">These:</span>
        <ul class="choices">
          <li class="choice">sandwich</li>
          <li class="choice">Cookie</li>
        </ul>
      </li>
    </ol>
  </div>
</div>
```
What’s This Got To Do With DITA?

- We have a need to be able to natively represent information typing and semantics within XHTML (think wikis, client-side filtering, better search)
- DITA provides an existing, widely-adopted standard for authoring reusable, semantically-rich information
- Microformats provide "semantics based on the HTML class attribute" and DITA also uses the class attribute to track the semantic structure and ancestry of each DITA element.
- Currently we already publish from DITA to XHTML, losing the semantics along the way
- If we can represent DITA semantics natively within XHTML, then it should become easier to capture XHTML-sourced information as proper DITA topics
- If we can leverage DITA from within XHTML, then the other 99.9% of the web can start writing more reusable, semantically-rich information without giving up their existing tools, skills, and infrastructure.
Opportunity: a simple, DITA-aware browser extension