Metadata development & deployment: What software business practices apply?

William L. Anderson

Praxis101

20th International CODATA Conference

Beijing, China

23 - 25 October 2006
What we’ll cover

• Metadata development & deployment challenges
• Software product development challenges
  – How businesses are responding
• What software product development practices could help with metadata deployment?
• Questions & discussion
The metadata problem

• Data are almost useless without description
• All data are local and situated
• But sharing requires shared descriptions
• The primary requirement for metadata:
  – Locally useful and usable
  – Globally understandable and interoperable

So, how to deliver on both needs?
Software product development problem

• Product requirements in conflict
  – Rapid delivery of new features & updates
  – Product reliability & ease of use

• Development processes need to flexible and dependable

• New practices demand change from developers and users
A computer systems challenge

- Products, procedures, and standards
  - Presume activity that is orderly, step-wise and predictable
- What’s enacted in the world:
  - Activity that is situated and responsive to the needs of the moment
- The challenge is working across the gap
  - the key to usefulness and adoption
Software development responses

• User experience as the guide for features
• Deliver value to the customers & users
• Deliver less, but deliver sooner
• Let products evolve as use defines requirements
• Develop an architecture that adapts to changes (THIS IS THE HARD PART)
Software engineering changes

<table>
<thead>
<tr>
<th>Product-oriented engineering</th>
<th>Process-oriented engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design software</td>
<td>Design work processes &amp; software</td>
</tr>
<tr>
<td>Process information by rules</td>
<td>Interleave work, learning &amp; communication</td>
</tr>
<tr>
<td>Consider system static</td>
<td>Consider system dynamic &amp; evolving</td>
</tr>
<tr>
<td>Consider the environment to be predefined</td>
<td>Consider the environment to be tailored by users</td>
</tr>
<tr>
<td>Use document driven methods</td>
<td>Use iterative and agile methods</td>
</tr>
</tbody>
</table>
Metadata development/deployment ideas

• Respect the work practices of the users
  – Researchers and cataloguers

• Deliver simple schemas

• Focus on ease of use

• Start small:
  – Support sharing as it evolves

• Need a standards and deployment process that adapts to user experience and learning (THIS IS THE HARD PART)
MARC Content Designation and Utilization Study

Some preliminary data*

- 7,595,887 LC-created records in dataset
- Type of Record: Book, Pamphlets, and Printed Sheets
- Total number of unique fields occurring: 167
- 14 fields account for 80% of occurrences: (8.3%)
- 21 fields account for 90% of occurrences: (12.6%)
- Approximately 110 fields (66%) occur in less than 1% of all records

[* William E. Moen, School of Library and Information Sciences
Texas Center for Digital Knowledge, University of North Texas: http://www.mcdu.unt.edu ]
Consequences for software & metadata development

• Requirements cannot be completely known
  – Do not try to solve the general problem

• Develop and deploy anyway
  – Focus on what can be done soon

• Find out what users are doing
  – Be empirical
  – Learn from experience

• Design for change
  – Take smaller steps (mistakes less costly)
Thank you.

Bill Anderson
band@praxis101.com