Dublin Core Metadata in Corporate Environments

Work Item 3 –

CEN/ISSS Workshop on Dublin Core Metadata

CEN - the European Committee for Standardization
CEN/ISSS - Information Society Standardization System
WS/MMI-DC - Workshop on Metadata for Multimedia Information - Dublin Core
CEN Work Item 3 goal

- Identify & address corporate metadata needs to support modern business organizational functions like internal content reuse, federated search, & knowledge management.
What we did

- Compiled issues from listserv archives.
- Identified corporate metadata people.
- Developed questionnaire & conducted interviews.
- Articulated corporate metadata requirements.
- Validated these requirements with further interviews.
- Collected case studies & best practices.
Who we talked to

- Applied Information Technique
- AstraZenica
- BBC
- BellSouth
- Cisco
- Daimler Chrysler
- Giunti Labs
- GSK
- Halliburton
- HP
- IBM
- Intel
- John Wiley & Sons
- Lilly
- PeopleSoft
- Rohm Haas
- SAP
- Software AG
- Unisys
How is Dublin Core used in corporate environments?

- De facto: 57%
- Simple: 43%
- Access enabler: 43%
- Compliance: 29%
Dublin Core framework for corporate use

- Not just 15 elements
- A framework to enable cross-resource exploration and use

Dublin Core is framework for “integration metadata” at BellSouth
How Dublin Core is extended?

- Doc Types: 100%
- Products & Services: 86%
- Roles: 57%
- Inconsistent Encoding: 57%
## Business process document types or genres

<table>
<thead>
<tr>
<th>Source: Oil &amp; gas services company</th>
</tr>
</thead>
<tbody>
<tr>
<td>analysis, appraisals, assessments, forecasts, predictions</td>
</tr>
<tr>
<td>agendas, plans, designs, schedules, workflow</td>
</tr>
<tr>
<td>applications, proposals, requests, requirements</td>
</tr>
<tr>
<td>permits, consents, approvals, rejections, certificates</td>
</tr>
<tr>
<td>work orders, correspondence</td>
</tr>
<tr>
<td>auditing, compliance, testing, inspections, operations reports</td>
</tr>
<tr>
<td>lessons learned, after-action reviews, meeting minutes, FAQs</td>
</tr>
<tr>
<td>policies, procedures, training manuals, standards, best practices</td>
</tr>
<tr>
<td>research notes, journal articles</td>
</tr>
<tr>
<td>newsletters, bulletins, press releases</td>
</tr>
<tr>
<td>ads, brochures, data sheets, technical notes, case studies, price lists</td>
</tr>
<tr>
<td>checklists, templates, forms, logos, branding</td>
</tr>
<tr>
<td>software, database forms</td>
</tr>
</tbody>
</table>
Dublin Core encoding practices (fragment)

<table>
<thead>
<tr>
<th>DC Elements</th>
<th>Definition</th>
<th>Schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Content maker</td>
<td>dc.creator</td>
</tr>
<tr>
<td>Publisher</td>
<td>Entity responsible for making the resource available</td>
<td>dc.publisher</td>
</tr>
<tr>
<td>Format</td>
<td>File format of the resource</td>
<td>dc.format</td>
</tr>
<tr>
<td>Type</td>
<td>Content genre</td>
<td>dc.type</td>
</tr>
<tr>
<td>Coverage</td>
<td>Spatial locations</td>
<td>dc.coverage</td>
</tr>
<tr>
<td>Audience</td>
<td>Content audience</td>
<td>dcTERM.audience</td>
</tr>
<tr>
<td>Non DC</td>
<td>Products and services</td>
<td>local.products</td>
</tr>
<tr>
<td>Non DC</td>
<td>Business purpose</td>
<td>local.businessPurpose</td>
</tr>
<tr>
<td>Non DC</td>
<td>Technical expertise or competencies</td>
<td>local.expertise</td>
</tr>
<tr>
<td>Non DC</td>
<td>Market segments</td>
<td>NAICS.industries</td>
</tr>
</tbody>
</table>
Methods used to create & maintain metadata

- Forms: 71%
- Distributed Production: 57%
- Centralized Production: 43%
- Not Automated: 43%
Selected metadata tools used by companies

**Metadata tagging products**
- Documentum Content Intelligence Services
- Interwoven MetaTagger
- Microsoft SharePoint
- Stellent Content Categorizer
- Vignette Taxonomy & Advanced Search

**Auto-categorization tools**
- Autonomy
- Convera
- Entopia
- Entrieva (Semio)
- InXight
- Stratify
- Teragram
- Verity K2 Developer
- Verity Ultraseek

**Vocabulary editing tools**
- Interwoven Metatagger Studio
- Microsoft Excel, Visio, Word Outline
- MultiTes
- SchemaLogic
- Verity VCC
- Wordmap

**Guided navigation tools**
- Endeca
- Siderean
- Thunderstone
What controlled vocabularies are being used?

- ERP: 57%
- LDAP: 29%
- Business Process: 14%
- ISO 3166: 43%
## Common controlled vocabularies

<table>
<thead>
<tr>
<th>Facet</th>
<th>Definition</th>
<th>Example Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Products and Services</td>
<td>Names of products and services.</td>
<td>Local products and services from ERP system, etc.</td>
</tr>
<tr>
<td>Organization</td>
<td>Organizational structure.</td>
<td>Local organizational structure, etc.</td>
</tr>
<tr>
<td>Content Types</td>
<td>Structured list of the various types of content being managed or used.</td>
<td>AGLS Document Type, AAT Information Forms, records management policy, etc.</td>
</tr>
<tr>
<td>Industries</td>
<td>Broad market categories such as lines of business, life events, or industry codes.</td>
<td>SIC, NAICS, etc.</td>
</tr>
<tr>
<td>Locations</td>
<td>Place of operations or constituencies.</td>
<td>ISO 3166, local postal service, etc.</td>
</tr>
<tr>
<td>Functions</td>
<td>Functions and processes performed to accomplish mission and goals.</td>
<td>Enterprise Ontology, AAT Functions, etc.</td>
</tr>
<tr>
<td>Audiences</td>
<td>Subset of constituents to whom a piece of content is directed or intended to be used.</td>
<td>GEM, ERIC Thesaurus, IEEE LOM, etc.</td>
</tr>
<tr>
<td>Topics</td>
<td>Business topics relevant to mission and goals.</td>
<td>ERIC Thesaurus, ProQuest, etc.</td>
</tr>
</tbody>
</table>
What specific guidelines are needed for the use of specific elements?

![Bar chart showing usage percentages of different DC elements: dc.subject 57%, dc.publisher 43%, dc.type 43%, dc.date 57%, Schema registry 29%.]
Next steps

❖ Writing case studies for metadata practice areas.
❖ Iterating best practice document.
❖ Conducting workshop in Brussels for corporate metadata implementers & vendor product managers.
Content Management and Metadata – CEN/ISSS Workshop on Dublin Core Metadata

- Thursday, 13 January, 2004 – CEN/CENELEC Meeting Centre, Rue de Stassart 36, B-1050 Brussels

- Focus on automated metadata workflow & enterprise content management systems (ECMS)

**Experts**
- Martin White, Managing Director, Intranet Focus
  ... and others

**Vendors**
- Documentum
- Entopia
  ... and others

**Case Studies**
- Shell
- Nokia
  ... and others

**Q&A**
Contact Information

❖ Joseph A. Busch, Taxonomy Strategies (USA), jbusch@taxonomystrategies.com
❖ Makx Dekkers, AMI Consult SARL (Barcelona), mail@makxdekkers.com
❖ Kerstin Forsberg, AstraZeneca (Sweden), kerstin.l.forsberg@astrazeneca.com
Questions?