Digitization – Basics and Beyond workshop

Interoperability of cultural and academic resources
New services for digitized collections

Muriel Foulonneau (mfoulonn@uiuc.edu)

Grainger Engineering Library
University of Illinois at Urbana-Champaign

Digital Imaging Lab.
Feb 2005
Sharing content

- Digitization produces digital content

- Share it with “services”
  - My Website
  - My internal system
  - Declare it to search engines
  - Allow other services to access it (possibly from different areas)

⇒ this is all about standards for content and services
1 collection – multiple services

Google

OA1ster

Institute of Museum and Library Services

C.I.C

mfoulonn@uiuc.edu
University of Illinois at UC
Integrated Access to CIC Metadata

http://cicharvest.grainger.uiuc.edu/

mfoulonn@uiuc.edu
University of Illinois at UC

Feb, 2004
Thematic access to resources
Why new services?

- Bring your content to attention of new users outside your immediate community
  - 37% of visits to images of the State Library of New South Wales came from the PictureAustralia portal in 2002/3

- ‘Union catalog’ approach to sharing

- Access to the hidden Web.....
Digitization and interoperability

- Building services

=> New services need content with similar features
What is interoperability

- Interoperability is the capacity for different systems to talk to each other

- I need
  - A standard language
  - An interpreter

- “01-04-04”
  - this is a month
  - 01=“Jan”
Various types of interoperability

- Technical
  - Protocols, hardware, … Mac/PC, Netscape/IE …

- Organizational

- Content – related = metadata
  - What do you talk about? The “item” = Granularity and nature of the object
  - Semantic : date…. Created? Published?
  - Syntactical : 04 January 2004
  - Linguistic : 04 Enero 2004
Metadata

- Are used to
  - Define rights and conditions of use
  - Describe structure
  - Manage
  - Provide information
  - Retrieve
  - Preserve

⇒ Descriptive
⇒ Administrative
⇒ Structural

(subject) Upper portion rather fragmentary
(subject) two small fragments left apart as their appartenance to t
(subject) <dct:spatial xsi:type="ui:ISO3166" ui:code="EG">Egypt</dct:spati
(subject) probably Karanis, Herakleidou meris, Arsinoite nome, pr
(subject) <subject>Karanis section (now research box, Locker 3, 1/2)</subject>
(subject) List of names followed by a number (1 or 2), under the f
(subject) <subject>published</subject>
(subject) <subject>Pap</subject>
(subject) 22
Interoperability layers

**Organization:**
Data sources have a similar role

**Protocol:**
Computers can talk together

**Content-related:**
Content is meaningful when viewed or managed together
Sharing metadata: Federated search

- My user wants “mills”… Whatever that comes from
  - Federated search

Example metadata:
- <title>My resource</title>
- <date>04

Eg. Z39.50, SRU/SRW, WAIS
Interoperability is not the last stage

- Digitization project
  - Selection of content
  - Technical Features
  - Masters storage + mgt
  - Metadata creation /capture

- Sharing content
  - Disclosure of resources
  - Re-use and re-purposing
Multiple services use different features

Metadata

Collection descript.

Full text

Metadata AND resources

Metadata

Metadata AND resources
Selection of material

- Selection criteria for a specific project
  - Content quality
  - Copyright
  - Preservation of originals
  - Popularity
  - Etc

- But also for other potential uses
  - Complementary to other institutions
  - Interest in other contexts
    - Potential service providers
Technical features

- Technical features are defined for a specific project
  - Definition
  - Size
  - Format

- But also for other potential uses
  - Definition
  - Size …
  - Master used to generate new versions for other purposes
Storage and management of digital masters

- Will be the basis of future uses
  - Be able to retrieve them
  - Be able to re-use them = generate relevant versions for other types of use
- Do not digitize twice
  - Db of digital masters
- Preservation
  - Interoperability with future systems
  - Formats: readability in the future
  - Data integrity
Metadata creation/capture

- Metadata are created for a purpose
- Re-usability features
  - Documenting creation
  - Completeness
  - Include context ....

Record 2 of 8

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Author/Creator</td>
<td>Richard Lewis, c.1700-1734</td>
</tr>
<tr>
<td>Publisher</td>
<td>Printed by Hall and Sellers</td>
</tr>
<tr>
<td>Type</td>
<td>text</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://name.umdl.umich.edu/AM1315">http://name.umdl.umich.edu/AM1315</a></td>
</tr>
<tr>
<td>Collection</td>
<td>American Poetry Database</td>
</tr>
</tbody>
</table>
Quality criteria for
- Accessibility
- Usability
- Maintenance

Etc

But also FAME
- Discoverability
  - Through engines, peer referencing, registries …
- Disclosure
Disclosure

- Multiple access to the content
  - Website: dynamic
  - OAI repository
  - Multi-search?

- Metadata
  - In META TAGS
  - Collection level descriptions
Sharing content - Political issues

- Not all sectors have the same “philosophy”
  - Preservation
  - Access
  - User experience

⇒ Not all sectors have the same attitude towards sharing content

- Everybody agrees on the necessity to share / aggregate if…
  - New services for my user
  - Gaining new users
  - Not yet for professional and intermediary services

⇒ But not necessarily with who + not a priority if no political incentive and value-added services