Chinese-European Workshop on Digital Preservation

Beijing (China), July 14 – 16, 2004

Metadata-Registries

Dr. Heike Neuroth
Research & Development
Göttingen State and University Library (SUB)
neuroth@mail.sub.uni-goettingen.de
Table of Contents

- Metadata
- Application Profile
- Interoperability
- Metadata Registry
- Examples
- Semantic Web
- Metadata Registry & Semantic Web
Metadata

- Metadata are (structured) data about data
- They describe the content, quality, conditions (like rights, access), and other characteristics of (digital) objects
- For discovery, description, archiving, etc.
- Each metadata element has a detailed definition (semantic, syntax)
- Set of metadata elements = metadata element set (or schema)
Metadata is a language

- A metadata "sentence" might say:
  - "This book has **Author Johann Wolfgang von Goethe**, **Title Faust I**, and **Date Published** 1808

- Dublin Core was designed as a simple metadata language for the discovery of digital objects (Tom Baker: “pidgin language for a digital tourist”).

- In unqualified Dublin Core, the sentence above would say:
  - "This **resource** has **Author Johann Wolfgang von Goethe**, **Title Faust I**, and **Date 1808**
Application Profile

- Dublin Core Metadata Element Set
  - Too simple for most local application
  - Need to extend this set locally

- An application profile is a schema which consists of terms (metadata elements) drawn from one or more metadata schema optimised for a particular local application

- Examples
  - Dublin Core Library Application Profile
  - Renardus Application Profile
  - ...

Application Profiles and Interoperability

Contents:

• Keep precision in the meaning (semantics) of the terms used
• Make the sources of the terms explicit
• Spell out the local restrictions of use (value range, syntax, repeatability, obligation etc.)

Goal:

• Exchange of data with other providers: **Interoperability!**
Interoperability

- ... enables **different communities**, with different types of information and technologies, to achieve a general level of **information sharing** ...

- ... is defined as the ability of digital library components or services to be functionally and logically **interchangeable** ...

- ... different services and components can **communicate with each other** through open interfaces ...

---

*Eric Miller, W3C*
### Attribute Space
- e.g. DCMES, LOM

### Value Space
- e.g. DDC, LCSH

### Representation & Exchange
- e.g. METS, XML, RDF, SOAP, OAI – PMH

### Transport
- e.g. Z39.50, LDAP, HTTP

© based on White Paper of the DELOS Working Group on Registries
Metadata Registry

Need of registration of Application Profiles

- Clear declaration of rights holder
- Somebody is responsible for maintenance and updates
- Guarantees interoperability between metadata implementers
- Increase cooperation (metadata exchange, metadata sharing)
- Avoid double work by re-using already defined metadata elements
- ...

Examples

MetaForm (SUB Göttingen)

- Local "manifestations" of Dublin Core for specific projects introduce variations -- like "dialects"
- "Crosscuts": how are elements used in different implementations?
- Provides "mappings" and "crosswalks" between Dublin Core and other schemas of similar scope
- Demonstrates the sort of output one would want from queries to a distributed registry

http://www2.sub.uni-goettingen.de/
# Crosswalks

Database containing Dublin Core manifestations and other metadata formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Project description</th>
<th>Version</th>
<th>Mappings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dublin Core and its Dialects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGLS</td>
<td>Australian Government Locator Service</td>
<td></td>
<td>Mapping</td>
</tr>
<tr>
<td>BIBLINK</td>
<td>BIBLINK Project</td>
<td>Stand: 28.07.1998</td>
<td>Mapping</td>
</tr>
<tr>
<td>BSZ</td>
<td>Dublin Core in der Interpretation des BSZ (Bibliotheksservice-Zentrum Baden-Württemberg)</td>
<td>Draft Stand: 27.09.2000</td>
<td>Mapping</td>
</tr>
<tr>
<td>CEDARS</td>
<td>Metadata for digital preservation</td>
<td>Stand 2002</td>
<td>Mapping</td>
</tr>
<tr>
<td>CIC</td>
<td>Gesellschaft Deutscher Chemiker Keine Formatbeschreibung vorhanden</td>
<td></td>
<td>Mapping</td>
</tr>
<tr>
<td>DED</td>
<td>Denmarks Elektroniske</td>
<td>Stand: 2000</td>
<td>Mapping</td>
</tr>
<tr>
<td>CEDARS</td>
<td>Definition CEDARS</td>
<td>DC-Element</td>
<td>DC Definition</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Preservation Description</td>
<td>This sub-element contains the name of the publisher of the digital object.</td>
<td>DC.Publisher</td>
<td>An entity responsible for making the resource available</td>
</tr>
<tr>
<td>Information.Provenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Rights Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Information.Copyright Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name of Publisher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation Description</td>
<td>This sub-element contains the place of publication of this version of this digital object.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Provenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Rights Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Information.Copyright Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication-Place</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation Description</td>
<td>This sub-element contains the date of publication of this version of this digital object.</td>
<td>DC.Date.Created</td>
<td>Date of creation of the resource</td>
</tr>
<tr>
<td>Information.Provenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Rights Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Information.Copyright Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of Publication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation Description</td>
<td>This element specifies any other information objects which were judged, at the time of ingest, to be significantly related to the ingested digital object.</td>
<td>DC.Relation</td>
<td>A reference to a related resource</td>
</tr>
<tr>
<td>Information.Context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Related Information Objects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservation Description</td>
<td>This metadata section contains information relating to the intellectual property rights relevant to the digital object.</td>
<td>DC.Rights</td>
<td>Information about rights held in and over the resource</td>
</tr>
<tr>
<td>Information.Provenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information.Rights Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights Information.Copyright Statement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples

Schemas/CORES

- Registry of “all” Application Profiles and Metadata Element Sets
- “Rights holder”, maintainer
- Updates
- Search for Aps, single metadata elements etc.

...  

http://www.cores-eu.net/registry/
### Application Profiles

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orrnpaper Application Profile</td>
<td>1.0</td>
<td>Food and Agriculture Organization of the United Nations</td>
</tr>
<tr>
<td>Renardus Application Profile</td>
<td>1.0</td>
<td>Universidade do Minho</td>
</tr>
<tr>
<td>The British Library Application Profile</td>
<td>1.0</td>
<td>Renardus Consortium</td>
</tr>
<tr>
<td>The British Library</td>
<td></td>
<td>The British Library</td>
</tr>
<tr>
<td>The IEEE LOM Application Profile</td>
<td>1.0</td>
<td>The British Library</td>
</tr>
<tr>
<td>The IEEE LOM Metametadata Application Profile</td>
<td>1.0</td>
<td>Metadata for Education Group</td>
</tr>
<tr>
<td>The Qualified Dublin Core Application Profile</td>
<td></td>
<td>Metadata for Education Group</td>
</tr>
<tr>
<td>The Qualified Dublin Core Application Profile</td>
<td></td>
<td>合格的Dublin核心应用配置文件</td>
</tr>
<tr>
<td>The RDN Record Sharing (rdn_dc) Application Profile</td>
<td>1.0</td>
<td>Resource Discovery Network</td>
</tr>
<tr>
<td>The RSLP Collection Description Application Profile</td>
<td>1.0</td>
<td>Research Support Libraries Programme</td>
</tr>
<tr>
<td>The Simple Dublin Core Application Profile</td>
<td></td>
<td>The Dublin Core Metadata Initiative</td>
</tr>
<tr>
<td>The Simple Dublin Core Application Profile</td>
<td></td>
<td>The Dublin Core Metadata Initiative</td>
</tr>
</tbody>
</table>
Examples

Dublin Core Registry

- Defines Dublin Core schemata (terms, controlled vocabularies etc.)

- And relation between them

- Translations of elements, refinements and their definitions in different languages

http://dublincore.org/dcregistry/index.html
The Dublin Core Metadata Initiative's (DCMI) Metadata Registry is designed to promote the discovery and reuse of exiting metadata definitions. It provides users, and applications, with an authoritative source of information about the Dublin Core element set and related vocabularies. This simplifies the discovery of terms and related definitions, and illustrates the relationship between terms.

The reuse of existing metadata terms is essential to standardization, and promotes greater interoperability between metadata element sets. The discovery of existing terms is an essential, and prerequisite, step in this process. This application promotes the wider adoption, standardization and interoperability of metadata by facilitating its discovery, and reuse, across diverse disciplines and communities of practice.

This application was developed by the OCLC Office of Research, in cooperation with the Dublin Core Metadata Initiative Registry Working Group. It was developed, and is distributed, as an open-source project, built entirely upon open-source/open-standards software.
The Dublin Core Metadata Registry

Search / Browse

Display: Controlled Vocabulary Terms

Summary of DCMI Controlled Vocabulary Terms

Collection
Dataset
Event
Image
InteractiveResource
MovingImage
PhysicalObject
Service
Software
Sound
StillImage
Text

Items Found: 12
DCMI Roadmap for Development of Vocabulary Management and Registry Systems
(H. Wagner, R. Heery, S. Weibel)

Chinese-European Workshop on Digital Preservation, Beijing July 14 – 16 2004
Metadata Registries: Services

Advanced Registries will provide services for metadata exchange and enhancements:

- Map from one data element to another
- Adjust value range (e.g. notation -> caption)
- Normalize data (e.g. date format)
- Support standardization (e.g. by the use of authority files)
Semantic Web

A vision

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Metadata Registries & Semantic Web

- Metadata provides well-defined meanings to information
- Registries help to maintain this well-defined meanings
- Increase interoperability, also in (semi-) automatically ways (metadata exchange, metadata sharing, metadata re-use)
- **But**: all registries so far are very complicated, maintenance is very time-consuming (depending on metadata developers)
- **Vision**: more simple, concentrating on core set
Thank you very much for your attention

Dr. Heike Neuroth
Research & Development
Göttingen State and University Library (SUB)
neuroth@mail.sub.uni-goettingen.de