Permanence and Temporal Interoperability of Metadata in the Linked Open Data Environment

Shigeo Sugimoto\textsuperscript{1}, Chunqiu Li\textsuperscript{1}, Mitsuharu Nagamori\textsuperscript{1}, Jane Greenberg\textsuperscript{2}

1. University of Tsukuba, 2. Drexel University

sugimoto@slis.tsukuba.ac.jp
Risks in Long-term Use of Metadata

30 years after creation of a metadata database,
• Metadata schema of this metadata is unknown...
• Data creation guidelines of this database is lost...
• A controlled vocabulary is used in this database but nobody knows the vocabulary...
• URIs are obsolete...
• ..........
Goal of this Study

• To share issues of long-term maintenance and use of metadata in the Linked Data environment on the Web

• To clarify facets for long-term maintenance and use of metadata from a viewpoint of interoperability, i.e. temporal interoperability
Background and Motivation (1/2)

• Digital Preservation is important --- a well-known issue
  • Preservation is a long-term maintenance process of the preserved resources with their metadata

• Metadata has crucial roles for preservation of any objects --- physical objects, analog resources, digital resources
  • e.g., Preservation Description Information (PDI) of Open Archival Information System (OAIS)

• OAIS is a well-known standard but it is not designed for metadata preservation in the Linked Open Data environment
Open Archival Information System (OAIS)

- **Ingest**
- **Archival Storage**
- **Data Management**
- **Preservation Planning**
- **Access**
- **Administration**

**PRODUCER**
- SIP
- AIP

**CONSUMER**
- DIP

**IP**: Information Package
**SIP**: Submission IP
**AIP**: Archival IP,
**DIP**: Dissemination IP
Open Archival Information System (OAIS)

Packaging Information

Information Package

Information Object

Preservation Description Information (PDI)

Description about package
Open Archival Information System (OAIS)

Information Object

Preservation Description Information (PDI)

Packaging Information

Description about package

Information Package

Data Object + Representation Information
Open Archival Information System (OAIS)

Information Object → Preservation Description Information (PDI) → Description about package

Packaging Information

Information Package

Information Object ← Data Object + Representation Information
Open Archival Information System (OAIS)
• OAIS is well known as a standard for digital preservation.
• Metadata is a key component in this model.
• We need to keep metadata interoperable over time.
• We can use OAIS to preserve metadata and metadata schemas in a package but ....
In conventional environments, metadata Schemas are preserved as a document, which may be preserved in an information package. This would not fit to the LOD environment where metadata schemas may be transferred and shared on the Web, i.e., First Class Object. Need to keep metadata instances consistently interpretable not only by humans but also by machines. Need to understand requirements for metadata longevity and to build a model to keep metadata and their schemas consistent over time.
Metadata Longevity as Temporal Interoperability of Metadata

• Long-term maintenance of metadata is to keep metadata interoperable over time, i.e. temporal interoperability

• DCMI’s Application Profile has contributed to improve metadata interoperability
  • Clear separation of metadata constructs – vocabularies, structural features, implementation syntax

• There are models based on DCMI AP that help understand metadata interoperability
  • Layered Model for Metadata Interoperability by DCMI
  • Simple Layered Model for Metadata Schemas by the authors at DC-2004
Basic Aspect (1): DCMI AP and Interoperability Models

Application Profile

- Functional Requirements
- Domain Model
- Usage Guidelines
- Metadata Vocabularies
- DCMI Abstract Model
- Syntax Guidelines and Data Format
- Community Domain Models
- RDF/S
- RDF

Domain standards

Foundation standards

Clear Separation of metadata constructs in layers
**DCMI AP and Interoperability Models**

<table>
<thead>
<tr>
<th>Description Set Profile Interoperability</th>
<th>Implementation Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared formal vocabularies and constraints in records</td>
<td>Concrete syntax</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description Set syntactic interoperability</th>
<th>Structural Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared formal vocabularies in exchangeable records</td>
<td>Abstract syntax</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Formal semantic interoperability</th>
<th>Metadata Vocabularies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared vocabularies based on formal semantics</td>
<td>Term Semantics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shared term definitions</th>
<th>A Layered Model of Metadata Schema (Nagamori, Sugimoto, 2004)</th>
</tr>
</thead>
</table>

**Interoperability Levels of DCMI** (Nilsson et al, 2008)
## DCMI AP and Interoperability Models

### Description Set Profile Interoperability
- Shared **formal vocabularies** and constraints in records

### Description Set syntactic interoperability
- Shared **formal vocabularies** in exchangeable records

### Formal semantic interoperability
- Shared **vocabularies** based on formal semantics

### Shared term definitions
- **Shared vocabularies** defined in natural languages

### Implementation Syntax
- Concrete syntax

### Structural Constraints
- Abstract syntax

### Metadata Vocabularies
- Term Semantics

---

**Interoperability Levels of DCMI** (Nilsson et al, 2008)

---

**A Layered Model of Metadata Schema** (Nagamori, Sugimoto, 2004)
DCMI AP and Interoperability Models

<table>
<thead>
<tr>
<th>Description Set Profile</th>
<th>Interoperability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared formal vocabularies and constraints in records</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description Set</th>
<th>syntactic interoperability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared formal vocabularies in exchangeable records</td>
<td></td>
</tr>
</tbody>
</table>

Interoperability Levels of DCMI (Nilsson et al, 2008)

- Structural Constraints
  - Abstract syntax
  - Concrete syntax

- Metadata Vocabularies
  - Term Semantics

DCMI AP and Interoperability Models

Clear Separation of metadata constructs in layers

Implementation Syntax

Description Set Profile (Structural Constraints)

Vocabularies

Interoperability Levels of DCMI (Nilsson et al, 2008)

Basic Aspect (2): Metadata Schema as Meta-Metadata

• Metadata schema which defines semantics, structural constraints, syntactic features, etc. is metadata about metadata → meta-metadata

• There is a scheme to define metadata schemas → meta-meta-metadata

• This is an endless iteration of meta, but all of these are metadata which may be shared on the Web and should be maintained over time
Basic Aspect (3): Metadata as a First Class Object

• **General Question:** Where is an metadata instance instantiated?

• In conventional system environment, metadata instances are instantiated as a database record and retrieved through API/UI.

• On the Web, metadata instances which may or may not be instantiated as a database record can be realized as a (part of) XML resource transferrable and shareable on the Web.

• In the Web environment of today, not only metadata instances but also meta-metadata and meta-meta-metadata instances may be transferred and shared on the Web.
Metadata Entities and Preservation Options

• Conventional metadata preservation strategies
  • Metadata is stored in a database → maintain the database consistently
  • Metadata schemas are documented → maintain the documents consistently
  • Metadata schemas are designed based on standards → keep the standards information updated

• Three preservation options for preserving metadata and meta*-data
  • Preservation as an Information Package
  • Preservation as a database
  • Preservation as a first class object
Metadata Entities and Preservation Options

- Preservation as an information object
  - associated document-like objects (annotation)
  - I.P.
  - Meta-Schema
  - Metadata Schema Registry

- Object Instance
  - archived
  - RDF instances a First Class Object
  - annotated
  - Provenance Description

- Metadata Instance
  - deposited
  - archived
  - Metadata Schema Registry
Preservation as an Information Package

- Preservation as an information object
- Associated document-like objects (annotation)
- Object Instance
- Metadata Instance
- Metadata Schema
  - Application Profile
  - Metadata Vocabulary
- Meta-Schema

Long-term use and Preservation as a Database

- RDF instances a First Class Object
- Provenance Description
- Deposit
- Archived

Metadata Schema Registry
Preservation as a Database

- Preservation as an information object
- Associated document-like objects (annotation)
- Object Instance
  - Metadata Instance
  - Metadata Schema
    - Application Profile
    - Metadata Vocabulary
  - Meta-Schema

Long-term use and Preservation as a Database

- Provenance Description
- Archived
- Deposited
- Archived
- Annotated
- Metadata Schema Registry

RDF instances a First Class Object
Preservation as a First Class Object

- Long-term use and preservation as a First Class Object
- RDF instances as a First Class Object
- Metadata Schema
  - Application Profile
  - Metadata Vocabulary
- Meta-Schema
- Associated document-like objects (annotation)
- I.P.
- I.P.
- I.P.

Packaged

Archived

Deposited

Annotated

Metadata Schema Registry

Provenance Description

Archived
Discussion - Facets in Metadata Preservation

• Facets of Metadata Preservation
  1. Formats • • • Database, XML, Documents, etc.
  2. Meta Levels (meta*-metadata) • • • meta-, meta-meta-, meta-meta-meta-…
  3. Metadata Schema Components • • • Singapore Framework entities
  4. Dynamic Entities • • • mapping tables used to combine vocabularies
  5. Documentation • • • documents about contextual information

• Need to clarify requirements for preservation in each facet
Discussion – Risk Factors in Metadata Longevity

• Metadata Instances
  • Instances transferred to other sites from their primary sites have high risks

• Application Profiles
  • Document-oriented Components, e.g. usage guidelines: preservation as a document
  • Structural definitions, e.g. description set profiles, preservation as a machine
    interpretable object

• Metadata Vocabularies
  • Revision history should be recorded in machine interpretable forms, e.g., RDF

• Other Entities
  • URI: Open issue
  • Other entities such as mapping tables and contextual Information should to be
    maintained with metadata schemas
Concluding Remarks

• It is well recognized that metadata plays crucial roles in digital preservation. However, metadata preservation (=long-term maintenance of metadata) in the LOD environment has not been explored well

• Provenance is a very important aspect for metadata longevity

• There is no panacea but we need to understand requirements and to develop models for metadata permanence

Thank you very much for your attention!

Shigeo Sugimoto