Study Report and NP Proposal on Semantic Metadata Mapping Procedure

2008. 11. 19
Tae-Hoon Lim and Tae-Sul Seo

taehoon@dpc.or.kr
tsseo@kisti.re.kr
Brief History

- 2007-07-19/New York, USA
  - Semantic harmonization of metadata
  - WG2N1042(SC32N1658)
  - Request for a new study period

- 2007-12-06/Seoul, Korea
  - Semantic harmonization of metadata
  - WG2-N1077, WG2-N1078
  - Study report

- 2008-05-xx/Sydeny, Australia
  - Semantic Metadata Mapping (SMM) Processes
  - WG2N1122
  - Request for extending the study period
Motivation

- A measure is needed to mediate between metadata sets already used.
- A metadata crosswalk has poor semantics because it is usually based on a simple one-to-one mapping.
- Therefore, a metadata crosswalk is required to be elaborated in order to have semantics.
- The idea of ISO/IEC 11179 can be helpful to improve metadata crosswalk semantically.
- This candidate standard includes a procedure called semantic metadata mapping procedure (SMMP), which deals how to identify metadata to be mapped and how to categorize the relationship of each pair of metadata.
Purpose

The purpose of this standard is to set up a procedure for making a metadata mapping table that conform to ISO/IEC 11179 standard, and thus, to improve semantic harmonization of metadata.
Semantic Metadata Mapping Procedure

Old Procedure

Collecting metadata sets

Grouping attributes

Finding common DECs

Mapping into a table

New Procedure

Identifying metadata sets

Grouping data elements

Semantic Mapping
Semantic Metadata Mapping Procedure

Main Processes

- Identifying metadata sets
- Grouping data elements
- Semantic Mapping

Sub-Processes

1. Identifying metadata sets to be mapped
2. Finding objects of all data elements
3. Grouping data elements by objects
4. Finding properties
5. Grouping data elements by properties
6. Finding data element concepts (DECs)
7. Mapping by data element concepts (DECs)
8. Recommending data element names
9. Giving notes according to types of heterogeneities (appendix A)
### Example of eBook

#### Identifying metadata sets

1. Metadata sets to be mapped: three metadata sets

<table>
<thead>
<tr>
<th></th>
<th>OpenEBPS</th>
<th>MODS</th>
<th>TEI header</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Description of Electronic Book</td>
<td>Description of Library resources</td>
<td>Encoding methods for machine-readable texts</td>
</tr>
<tr>
<td>Number of fields</td>
<td>15</td>
<td>About 60 (top level: 20)</td>
<td>Over 20</td>
</tr>
<tr>
<td>Sample data</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Authority</td>
<td>Open eBook Forum</td>
<td>Library of Congress</td>
<td>TEI Consortium</td>
</tr>
</tbody>
</table>
### Example of eBook

#### Grouping data elements

<table>
<thead>
<tr>
<th>OpenEBPS</th>
<th>MODS</th>
<th>TEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>title</td>
<td>title</td>
</tr>
<tr>
<td>subTitle</td>
<td>subTitle</td>
<td></td>
</tr>
<tr>
<td>partNumber</td>
<td>partNumber</td>
<td></td>
</tr>
<tr>
<td>partName</td>
<td>partName</td>
<td></td>
</tr>
<tr>
<td>nonSort</td>
<td>nonSort</td>
<td></td>
</tr>
<tr>
<td>creator(role)</td>
<td>name:role</td>
<td>title</td>
</tr>
<tr>
<td>creator(file-as)</td>
<td>name:namePart</td>
<td>seriesStmt:title</td>
</tr>
<tr>
<td></td>
<td>name:displayForm</td>
<td>seriesStmt:idno</td>
</tr>
<tr>
<td>subject</td>
<td>topic</td>
<td>keyword</td>
</tr>
<tr>
<td></td>
<td>classification</td>
<td>class</td>
</tr>
<tr>
<td></td>
<td>catographics</td>
<td>catRef</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
<td></td>
</tr>
</tbody>
</table>

#### Diagram

- eBook
  - title
  - creator
  - subject
## Example of eBook

### Semantic Mapping

<table>
<thead>
<tr>
<th>DEC</th>
<th>OpenEBPS</th>
<th>MODS</th>
<th>TEI</th>
<th>Recommended DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ebookTitle</td>
<td>title</td>
<td>title</td>
<td>title</td>
<td>H/dec</td>
</tr>
<tr>
<td></td>
<td>subTitle</td>
<td>H/dec</td>
<td>seriesStmt:title</td>
<td>H/dec, S/del</td>
</tr>
<tr>
<td>ebookAuthor</td>
<td>creator(role)</td>
<td>C</td>
<td>name:role</td>
<td>S/mis</td>
</tr>
<tr>
<td></td>
<td>creator(file-as)</td>
<td>D</td>
<td>name:namePart</td>
<td></td>
</tr>
<tr>
<td>ebookSubject</td>
<td>subject</td>
<td>topic</td>
<td>keyword</td>
<td>L/sim</td>
</tr>
<tr>
<td></td>
<td>classification</td>
<td>L/pre</td>
<td>class</td>
<td>L/pre</td>
</tr>
</tbody>
</table>

H – Hierarchical Differences  
D – Domain Differences  
L – Lexical Differences  
S – Syntactic Differences  
C – Complicated Differences
Types of Semantic Heterogeneities

- **Hierarchical Difference**; due to different level of detail
  - Generalization : Specialization, Composition : Decomposition

- **Domain Difference**; due to different context and culture

- **Lexical Difference**; is different appearance
  - Synonyms, Abbreviation, Acronyms, Case sensitivity, Language, Variation

- **Syntactic Difference**; due to different arrangement of parts
  - Ordering, Delimiters, Missing

- **Complicated Differences**; due to different policies
Summary

- The working draft has been made during the study period.
- The description system for mapping, the types of semantic heterogeneity, was elaborated.
- It is desirable to be proposed as an IS in the future.
Thank you!