Metadata Models for Organizing Digital Archives on the Web
Metadata-Centric Projects at Tsukuba and Lessons Learned

Shigeo Sugimoto, Senan Kiryakos, Chiranhi Wijesundara, Winda Monika,
Tetsuya Mihara, Mitsuharu Nagamori
Faculty of Library, Information and Media Science
University of Tsukuba

Outline

• Background
• Project 1: Disaster Archive Metadata
• Project 2: Manga and Pop-culture Metadata
• Project 3: Cultural Heritage Metadata
• Summary
Background

• Research Domain: Digital Archive – a collection of digital resources, mainly of cultural and historical resources
• Metadata-Centric Projects at Tsukuba
  • Enhancing usability of digital archives of Great East Japan Earthquake by aggregating metadata within and across archives
  • A metadata model for aggregating Manga resources on the web
  • Modeling cultural heritage objects for digital archives – tangible and intangible cultural heritage
• Common Research Goals
  • Enrich values of digital archives by metadata aggregation within and across digital archives and linking institutional digital archives and web resources

Project-1: Great East Japan Earthquake Archive (1/5)

• Background – Great East Japan Earthquake and Tsunami
  • Great East Japan Earthquake (2011.3.11) caused serious damage in the North-Eastern Part of Japan, in particular the pacific coast by Tsunami
  • Regional governments, NPOs and universities started collecting materials about the disaster soon after the earthquake
  • National government started a program to help local sectors record the disaster as digital archives and National Diet Library (NDL) built a portal named “Hinagiku” to collect metadata from the archives and provide unified access
    • OAI-PMH + Standard metadata based on Simple Dublin Core
    • As of 2018.9, Hinagiku collects metadata from 47 archives/databases
    • Many photographs taken after the disaster and various resources
A screen shot of Hinagiku (in Japanese)

Search “津波 石巻” (tsunami Ishinomaki)
Ishinomaki: a name of a city near Sendai, Miyagi pref.

Found 10464 items incl. 4714 photographs, 197 sounds/videos

Search Results thumbnail and metadata elements – title, creator, date, source (contributor)
Project-1: Great East Japan Earthquake Archive (2/5)

- Research Problems
  - Quality of metadata, in particular those for photographs and videos
    - Basically, low quality because of financial and time limitations
  - Item-based metadata – simple but not easy find relationships among the items, e.g., grouping photographs taken during one event
  - What do users want to find – a particular photograph or a group of photographs of a particular event or place, linkage among items about a particular place and community across archives and web resources?

- Approach
  - Metadata Aggregation: Make a collection of related items using metadata – resource aggregation by temporal, special, agent and subject information
Project-1: Great East Japan Earthquake Archive (3/5)

1. **Metadata Aggregation of Photographs by Time, Location, and creator**
   - Location (longitude & latitude) → GeoHex
   - Time → Time gap between two consecutive photographs > 30 minutes
   - Creator → same agent (person or organization name)

2. **Metadata Aggregation by Subject Terms**
   - Clustering of all subject terms → Subject Term Clusters
   - Collect resources for each subject term cluster → metadata aggregation

3. **Ontology created using Types of Local Organization Names, e.g., government offices, public organizations, and private sectors**

4. **Change History of Local Geographic Entity Names**
   - Linking current and past resources
Project-1: Great East Japan Earthquake Archive (4/5)

Aggregation by Time and Location (longitude and latitude)

<table>
<thead>
<tr>
<th>Archives</th>
<th>Metadata</th>
<th>Sets</th>
<th>Size=1</th>
<th>Size&gt;99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aomori</td>
<td>48,338</td>
<td>6,571</td>
<td>1,599</td>
<td>102</td>
</tr>
<tr>
<td>Kuji-Noda-Fudai</td>
<td>72,894</td>
<td>8,531</td>
<td>2,023</td>
<td>189</td>
</tr>
<tr>
<td>Shinrokuden</td>
<td>96,441</td>
<td>8,188</td>
<td>962</td>
<td>59</td>
</tr>
</tbody>
</table>

Aggregation by Subject Terms

<table>
<thead>
<tr>
<th>Archives</th>
<th>Metadata</th>
<th>Sets</th>
<th>Size=1</th>
<th>Size&gt;99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aomori</td>
<td>68,032</td>
<td>3,596</td>
<td>1,060</td>
<td>1,332</td>
</tr>
<tr>
<td>Kuji-Noda-Fudai</td>
<td>127,383</td>
<td>8,290</td>
<td>2,240</td>
<td>609</td>
</tr>
<tr>
<td>Shinrokuden</td>
<td>124,552</td>
<td>7,910</td>
<td>2,047</td>
<td>4,612</td>
</tr>
</tbody>
</table>

Project-1: Great East Japan Earthquake Archive (5/5)

- **Findings**
  - Aggregation by temporal-spatial information and subject terms
    - Effective but not perfect – many small aggregated sets
      - A set of photographs taken at a festival and at a disaster location
  - Aggregation by subject terms with ontology
    - This is still under development
    - Location names play key roles
- **Current research**
  - Develop a geographic entity dictionary as a LOD dataset with name change history to connect Great East Japan Earthquake archives and reports created for Tsunami disasters happened from late 19th century to 20th century
  - Link the archives using the dictionary
Project-2: Metadata Model for Manga, Anime and Game (1/3)

- **Background – Linking metadata across domains**
  - Manga Metadata Research since 2009 at Tsukuba and Media Art Database (MADB) project by Agency for Cultural Affairs (Bunka-cho) of the Japanese Government
    - MADB: a database development project for Japanese Popular Culture – Manga, Anime, Video Game and Media Arts – Four component databases & four metadata schemas
    - Manga (M), Anime (A) and Game (G) are different domains but related each other
    - Metadata interoperability across the four databases
  - Linking institutional metadata and fan-created web sites to enrich metadata
    - Institutional metadata: Item-based bibliographic data
    - Fan-created web sites: Work-oriented descriptions, often across M, A and G

Project-2: Metadata Model for Manga, Anime and Game (2/3)

- **Research Goal – develop a metadata model to enhance metadata interoperability across the domains**
  - Metadata aggregation to link different metadata sources

- **Research Problems**
  - What entities of Manga should be described – a monograph (Item), a story presented in a monograph(s) or a magazine(s), a series of monographs, a title of a story series, etc.?
  - Is FRBR WEMI useful to link Manga, Anime and Game (MAG) domains?
  - Define a metadata model for linking MAG – institutional metadata and fan-created descriptions
    - *Superwork* as a concept to bridge M, A and G
Superwork as a Franchise entity connecting Works in different Domains

Superwork as an entity aggregating Works in different media created under a single franchise, e.g., Gundam, Dragon Ball, One Piece and Draemon.
Project-2: Metadata Model for Manga, Anime and Game (3/3)

• Findings
  • Need Class-based (or Description-based) mapping before Property-based mapping is required to make the MAG databases interoperable
    • Identification of objects for metadata description are inevitable to define interoperable metadata in each domain (i.e., Manga, Anime and Game)
    • Defining relationships of the objects across the domains
  • *Superwork* as an entity representing a Franchise, which is not included in FRBR but is crucial to link *Work* entities across the domains, i.e., M, A and G

• Current Research
  • Aggregation technologies: linking institutional Item-oriented metadata (e.g., the MAD database) and Work-centric non-institutional metadata (e.g., Wikipedia and fan-created sites)

Project-3: Metadata Model for Digital Archives – Tangible and Intangible Cultural Heritage (1/4)

• Background
  • Metadata model for collecting and aggregating institutional and non-institutional cultural heritage information to build digital archive for South and Southeast Asian countries where cultural digital archives are not well developed yet
    • Intangible cultural heritage as a connector across regions and among tangible objects
  • Conventional Item-centric metadata may not be good for intangible cultural heritage
Project-3: Metadata Model for Digital Archives – Tangible and Intangible Cultural Heritage (2/4)

• Research Goal
  • Metadata model for cultural heritage objects for digital archives, both tangible and intangible cultural heritage
    • Metadata of conventional cultural digital archives are Item-centric like metadata for conventional collections at memory institution
    • Modeling metadata to enhance reuse of exhibitions and organizations of cultural heritage as an intellectual creation
  • Metadata aggregation

Project-3: Metadata Model for Digital Archives – Tangible and Intangible Cultural Heritage (3/4)

• Research Problems
  • Metadata model for digital archives of tangible and intangible cultural heritage
  • Metadata model for building a digital archive by aggregating metadata collected from heterogeneous sources
  • Metadata model for cultural heritage exhibitions and organizations as a findable and reusable entities

• Approach
  • Identify a single performance of intangible cultural heritage as a physical instantiation which may be recorded in information media
  • Use One-to-One Principle of Metadata as an underlying model for metadata aggregation
  • Identify a single exhibition as an Item and develop a model using FRBR WEMI
Basic Idea for Building Digital Archives in S/SE Asia – Digital Archive by Collecting and Aggregating Metadata
Framework for aggregating CH metadata

Project-3: Metadata Model for Digital Archives – Tangible and Intangible Cultural Heritage (4/4)

• Findings
  • A base model for cultural heritage digital archives
    • Cultural Heritage in Digital Environment (CHDE)
  • A base framework to apply CHDE to metadata aggregation
    • One-to-One Principle and DCMI Application Profile (Description Set)
  • A base model to apply FRBRoo to (digital) exhibitions – Event & WEMI
    • Organizing metadata about exhibitions based in FRBR would enhance findability and reusability of exhibitions – Exhibitions as Intellectual Creation

• Current Research
  • Elaboration of the modular model of metadata for building digital archives
  • Elaboration of “Exhibition as Intellectual Creation” model
Some Lessons Learned

- Linking institutional digital archives metadata to web resources to enrich digital archives and enhance their usability
  - Link disaster photographs to historical documents and to regional web pages to create linked records as a community memory
- Metadata aggregation to create new access points
  - Photograph groupings, Superwork
- One-to-One Principle as a base model for metadata aggregation
  - Clear identification of objectives of metadata description helps aggregation
- FRBR-based view of digital archives
  - Digital archives and exhibitions as intellectual products by curators

Summary

- The objectives of these three projects are non-conventional digital archives in the cultural and historical domains
- The common issue among the three projects is metadata aggregation, although the aggregation types are different
- Data model level mapping before property level mapping for metadata aggregation across archives is crucial
- Linking institutional metadata and web resources, e.g., Wikipedia and fan-created sites, is crucial for enhancing the usability of institutional digital archives and for enriching archived contents, which is a common issue among these three projects – a challenging issue