PROPOSAL FOR DOCUMENTATION GUIDELINE RELATED TO DIGITIZATION OF CULTURAL HERITAGE

Tamako KITAOKA [a], Norio TOGIYA [b]

[a] Dept. of Comparative Social Studies, Ochanomizu University, Tokyo, Japan - g0770112@edu.cc.ocha.ac.jp
[b] Interfaculty Initiative in Information Studies, University of Tokyo, Japan - togiya.norio@iii.u-tokyo.ac.jp

KEY WORDS: cultural heritage, standardized digitizing, guideline, workphase, technical recommendation, documentation templates

ABSTRACT:

This proposal introduces guidelines for a standard digitization workphase, standard technologies recommended for adoption in each workphase and standard documentation templates for profiles given in each phase. The main focus of these guidelines is on the third of these, documentation, comprising information about what information should be gathered in each phase and where metadata should be stored. The workphase configuration includes not only digitization but covers the entire project including related activities. Organizing these workphases is to be able to design digitalization workflows. Also discussed is what relationship these documentation items have with existing international metadata standards such as PREMIS.

1. INTRODUCTION

1.1 Motivation and Aims

Many organizations are collecting, storing and using digital data of their cultural properties, calling this activity digital cultural heritage or digital archives. With the growth of digitization, many different guidelines have been created for the digitization of cultural heritage. However, there has been little progress in establishing guidelines for data formats describing about managing information and technical information that are particularly important in the digitizing process. Most digital archives that exist today were created for the purpose of using the created data for publishing or for managing cultural objects, and especially in the humanities field few were created for being used for research of the digital data as academic resources. Creating and managing digital data based the standards of today, this data will enable to be used for a variety of purposes including academic research. It is the most important to built a more valuable and useful cultural heritage which is the accumulation of this kind of data. For this reason, a guideline is proposed here for the creation of digital heritage with a focus on standards-compliant digitization and standards-based management.

1.2 Placing the Guideline

Below are two guidelines related to building digital cultural heritage.
PREMIS (May 2005) sets metadata standards for digital preservation. Information is managed from the three perspectives: administrative, technical and structural. Based on this structure, each digital data object is assigned metadata under the main categories of “object,” “event,” “agent” and “rights.” Consideration is made to ensure digital preservation by guaranteeing the viability, renderability, understandability, authenticity and identity.
NARA Technical Guidelines for Digitizing Archive Materials for Electronic Access (June 2004) proposes guidelines for digital image creation, a minimum level of metadata, file management and quality assurance. Of the two schools of thought for creating data for the purpose of digital preservation versus creating data to enable the use of cultural assets, these guidelines are defined from the latter perspective. These guidelines offer solutions to problems that arise when actually performing cultural heritage digitization.
The aim of the standardization guideline proposed here is to deliver more robust resource information that created digital data represent than under PREMIS. And it is to cover the entire cultural heritage digitization project planning better than the NARA-Technical Guidelines and to expand the utility of cultural assets discussing a aspect for digital preservation also.

2. STANDARDIZATION GUIDELINE

2.1 What is the Standardization Guideline

This guideline serve as a reference standard when creating digital heritage what digitizing both physical and nonphysical cultural properties for storage as well as usage. The standardization guideline was first envisioned in 2007 and has been developed since then as a joint research project of the University of Tokyo Interfaculty Initiative in Information Studies and Toppan Printing Co., Ltd., as Joint Research Related to Digital Heritage Standardization.
This guideline is envisioned for use in creating digital heritage for resources that could be used in both humanities research and natural sciences research. Based on the concept of “one source, multi-use”, images or audios are stored as source data files appropriate for the investigative analysis requirements of academic research, as well as presuming multiple use such as publishing for dissemination or educational use. For this reason, the guideline enables setting a direction in accordance with standards that will always meet the standard requirements in the

various places where data is handled, such as data creation and management.
The following standardization perspectives are considered.

- Data format
- Documentation format
- Colour management
- Monitor evaluation method
- Image evaluation method
- Data management method
- Backup method
- Dissemination method
- Metadata method

Digital heritage created in accordance with these guidelines provides digital sources for being used not only by organizations that create the digital heritage but also to share and use the heritage with others, including a digital source that can be inspected or analyzed for academic research, as well as digital sources for active use for various purposes, such as internet content, portable media content, printed publication or dissemination.
This guideline helps to set a direction for planning, designing, building and managing digital heritage and answers the needs of administrators or managers that handle digital heritage in organizations.

2.2 How should the Guideline be used

Using this guideline, we choose data formats for capturing, copying or digitizing data from provided formats list on each phase, and then decide on the method for creating this data. The guideline recommends standard formats appropriate to that type of data. By following the recommendations, users can apply the standard format and create data while ensuring usability and viability of the data source.
Next, quality requirements are considered. By understanding the data density (resolution, granularity, capacity) standards and color management requirements (saving RGB value information, input environment information), it is possible to promote the creation of data that is well suited to multi-use and data quality that is required for digital heritage.
The guideline's information item tables also serve as description templates for data profiles. By filling in these tables with values corresponding to the created data, this enables capturing a sufficient amount of attributes for that data and ensuring a correct understanding of the identity and attribution of that data object.

2.3 Structure and Components

The guideline is separated into seven phases of the digital heritage creation process.

A. Resource Content Research
B. Digitization
C. Evaluation
D. Data Management
E. Contents and Publishing
F. Metadata Design and Input
G.&H. Academic Research Use + Release

The guideline defines from step B through E.
Each phase specifies related standards, standardized formats, quality requirements and necessary information, which should be considered for the steps of that phase.

2.3.1 Phases

A. Resource Content Research

The cultural property is examined and its details are recorded. This is the information about the source for digital data kept in digital heritage. This guideline does not discuss which information items recorded here. Several metadata standards already widely used for recording and describing cultural material information and these are presented as lists sorted into "electronic resources," "libraries," "archives" and "museums" categories. Referring to these lists helps organize their data into an appreciate standard format. The types of cultural properties considered by the guideline are media objects recording intangible properties such as sounds or movements, flat shape objects such as maps or paintings, solid shape objects such as sculptures or fossils and film objects.

B. Digitization

Steps for creating master data digitized for the first time from the original cultural property are discussed. Input devices considered by the guideline include digital video cameras, scanners, digital still cameras, spectral radiographs, spectrophotometric colorimeters, colimeters, 3-D scanners and digital audio recorders. Digital data is created using the device type appropriate for the physical form of each original source object. Data documentation covers keeping information about not only common profiles for data creation, but also profiles for each input device and input environment and color management profiles.

C. Evaluation

This phase discusses evaluation of the created data. Only visual data is covered and auditory data is not here. Evaluation can be loosely divided into two stages: evaluation of the devices used to perform the evaluation and evaluation of visual data operated to input/output by those devices. A list of standard images or movies is provided for performing device evaluations. Observation criteria standards and evaluation perspective terms are defined for data evaluation for each item evaluated.

D. Data Management

This phase discusses standards for managing of primary digital data and secondary data created from master digital data. The following five processes are performed for data management: data format selection, storage media selection, storage environment selection, data storage documentation, data backup. Lists are provided for data format standards, recommended storage medias and standard backup methods and necessary conditions of storage environment are stipulated.

E. Content and Publishing

The publishing method is determined by selecting a method and format according to the purpose. A list of foreseeable methods is provided and work requirements that should be considered for each method are defined. Simple content types include motion picture movie file, still image file, specialized non-visible scanning data file, 3-D data file and audio file. Composite content types include virtual reality theatre formats, edited and compiled works such as packaged digital collections and printed materials.

F. Metadata Design and Input

This phase discusses standards for organizing and documenting all profile information as metadata, namely, data about the

digital data. Profiles gathered in each of the above phases. Information items about each profile must be documented and stored in a machine-understandable form and the schemas can be presented for RDF modeling or XML markup.

G. Academic Research Use + H. Release

By building a digital heritage through the steps up until the previous phases, it is possible to use the accumulated data for purposes requiring a stricter equivalence or authority, such as not only academic research but also conservation and repair of cultural properties. Reproducibility is thus ensured.

2.3.2 Profiles and Metadata sets

In each phase, profile information related to the digital data on each stage is obtained. Each profile element is arranged using a template. Template of elements has four contents, such as "element ID", "element name", "form", and "discussion". Approximately 90 items are defined as profile elements in total, and of these, the three items: "record ID" in phase B, "control ID" in phase D and "content ID" in phase E, are mandatory.
For the information managed as digital heritage, primary digital data created from cultural properties is recorded profiles in each phase: B (Digitization), C (Evaluation), D (Data Management), and secondary digital data is recorded profiles in C and D phases. An identifier on phase D, "control ID" is only one item that every digital data must have. In phase A (Resource), each individual countable cultural property as a source of some digital data has a single ID. In phase E (Contents & Publishing), each packaged content has a single ID. By each ID in A or E, profiles in phase A or E would be associated with profiles of a digital data in phase D.

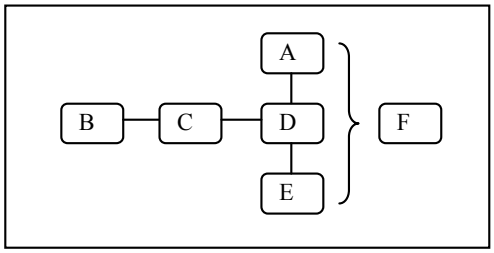

Figure 1. Correlation of metadata sets in phases

In the phase B (Digitization), there is special emphasis on obtaining profile information about a situation of digital data scanning, such as input devices, device positioning and settings, and input environment. This enables determining information about the condition of original source, namely cultural property, that condition cannot be determined solely from attributed value of digital data that created from the original. For example, information about the type of light source and the input device settings can be valid when attempting to restore the actual shape of a solid object from shadows left in the 2D image data created from the original.

3. FURTHER DEVELOPMENT AND ISSUES

3.1 XML scheme

For Phase F (Metadata) of this guideline, details regarding the XML describing method will be determined in the future. (An updated version including this information is planned for release in December 2009. Please refer from the Tokyo University Multimedia and Socio-information Studies Archive website (http://www.center.iii.u-tokyo.ac.jp/).) Ensuring consistency and uniformity in the description format will help to achieve the goal of a digital heritage in accordance with this guideline. Using this guideline will also enable exchanging and using object records from other database systems.
However, there are already several standards for machine understandable formats such as RDF models and XML descriptions as well as models based on RDFa and XHTML. It will be necessary to select the appropriate models in the future based on the perspective and objectives of this guideline.

4. CONCLUSION

It is considered that institutions that have cultural properties promote to digitize their heritage and use digitized resources not only for object management but also using material to promote, educate, research or restore. But to urge to make a digital heritage in-house is often difficult for organization and management systems today in Japan museums. Staffs feel they are preoccupied with the object management in their daily conventional operations, and they feel uneasy to make a digital heritage based on many conventions. For them to join the task of building a digital heritage means to increase the amount of work rather than arrange and streamline their operations. Such a way of getting a digital heritage should be altered. By improving the operating procedures to synchronize the process and built it more affordable for the daily conventional operations and digital heritage. It will contribute to the context where the digital heritage could be promoted and actually made. The guideline could play a role in combination with NARA Guidelines. In addition, taking into account its interaction with the UK standard SPECTRUM, the guideline will need to deepen the cooperation collection/object management. Clarifying the position of phase A of the guideline, the usefulness and practicality of the guideline will be enhanced.

References

Okumoto, M., 2007, The present condition of digital archives in Japanese museums, *Bulletin of Japan Museum Management Academy*, (11), pp.1-7.
PREMIS, 2005, *Data Dictionary for Preservation Metadata*, http://www.loc.gov/standards/premis/v2/premis-2-0.pdf (accessed 23 July 2009)
Puglia, S., 2004, *NARA Technical Guidelines*, http://www.archives.gov/preservation/technical/guidelines.pdf (accessed 23 July 2009)
Togiya, N., 2009, *Study of the model for integrating with the basis of objects on Digital Archives*, Tokyo : Bensei shuppan.
Togiya, N., 2009, 75th IFLA general conference and council, "The Digitization of Cultural Resources: An Exploration of Current Issues and the Future Outlook", http://www.ifla.org/files/hq/papers/ifla75/146-togiya-en.pdf (accessed 23 July 2009)