Providing Metadata to Historical Material on Viewer Application iPalletnexus

Norio Togiya, Mitsuhiro Tsuda, Akira Baba

The University of Tokyo, Ipallet, JAPAN
{togiya. norio@iii, baba@ hi}.u-tokyo.ac.jp, mtsuda@ ipallet.org

Abstract:
The iPalletnexus development group has currently developed a viewer application, iPalletnexus. The application allows browsing of large volumes of digitized images such as pictorial diagrams, maps and paintings over networks. On iPalletnexus, it is necessary to provide metadata that describe information such as the title and the author of each resource, and the metadata must be adaptable to any type of resources since the object resources belong to various genres. However, as Japanese traditional historical materials are to be handled on iPalletnexus at the same time, the metadata also needs to contain detailed information that can be used for academic research as well. Furthermore, the metadata must also contain the digital data certification that is handled on iPalletnexus. Thus, this paper introduces novel methods for storing information unique to traditional historical materials and digital data certification in metadata while maintaining versatility by using Dublin Core.

Keywords*: metadata, historical material, cultural resource, Dublin Core, iPalletnexus , Sekisui-zu, versatility*

1 Introduction

Since browsing of various digitized historical materials on the Internet is becoming more common, how metadata should be provided to such digitized historical materials is a major issue in the future [1]. How information about various cultural resources should be stored in the metadata has been discussed in various organizations [2]. Up todate metadata has been provided to various kinds of historical materials based on different relative standards on the resource types such as art objects, books and ancient documents in Japan. For example, in libraries in Japan, JAPAN MARC [3] based on USMARC [4] is the representative of metadata element set for modern book cataloging. However, cataloging rules for Japanese and Chinese old books are diverse, and there is neither national standard for metadata element set nor cataloging rule in museums and archives. Each organization has built particular metadata element set or cataloging rule for historical materials they own.

However, from the 1990's researching about various international standards of metadata element sets or cataloging rules was started in Japan. Nevertheless, it is difficult to adopt international standard cataloging rules or metadata element sets because these standards are not necessarily adapted for traditional Japanese historical materials and specific research. Although, Historical material has various unique features in it, international standard cataloging rules seem not to adapt to such a unique point in details. Thus, these cataloging rules have not become popular and there are not standardized cataloging rules for historical material in libraries, museums and archives in Japan.

However, as digitized historical materials can be handled on a viewer application or a browser, irrespective of their resource types, range of users who can access it widen since digitalized data can be treated very easily. The unification of metadata to a certain degree is needed for any kind of historical material.

The iPalletnexus development group [5] has also pursued research into metadata designs that can be handled beyond the genres of resources and range of users by using the metadata function of iPalletnexus. iPalletnexus is an application

capable of browsing large volumes of digitized image data at high speed and analyzing them by utilizing functions such as transmission and comparison [6]. The historical materials that can be read by iPalletnexus must contain metadata describing their titles, creators and created years, etc. However, since the genres of historical materials handled on iPalletnexus vary from paintings, maps, pictorial diagrams, ancient documents, Ukiyoe (Japanese wood block prints), Japanese and Chinese books, to sculptures, the metadata describing them must be able to adapt beyond these genres.

Meanwhile, the metadata must store more detailed information, as well as information specific to the Japanese traditional historical materials, for academic research. In addition, since the metadata handles not only information on the physical object, but also on their digital data, they also need to contain information on specifications such as format, size and capacity of that digital data, etc. In consideration of these points, the development group has created elemental set, the metadata for iPalletnexus.

This paper explains the process of creating that metadata element set for iPalletnexus. In Section 2, the iPalletnexus application and role of the metadata are explained. In Section 3, a practical composition method of metadata is given. In Section 4, an example of actual application, Sekisui-zu, an antique map that is issued in the Edo period in Japan, is given. In Section 5, searching, editing and updating functions for created metadata are described. And finally in Section 6, accomplishments and future prospects are discussed.

2 Relation Between iPalletnexus and Metadata

2.1 Outline of iPalletnexus

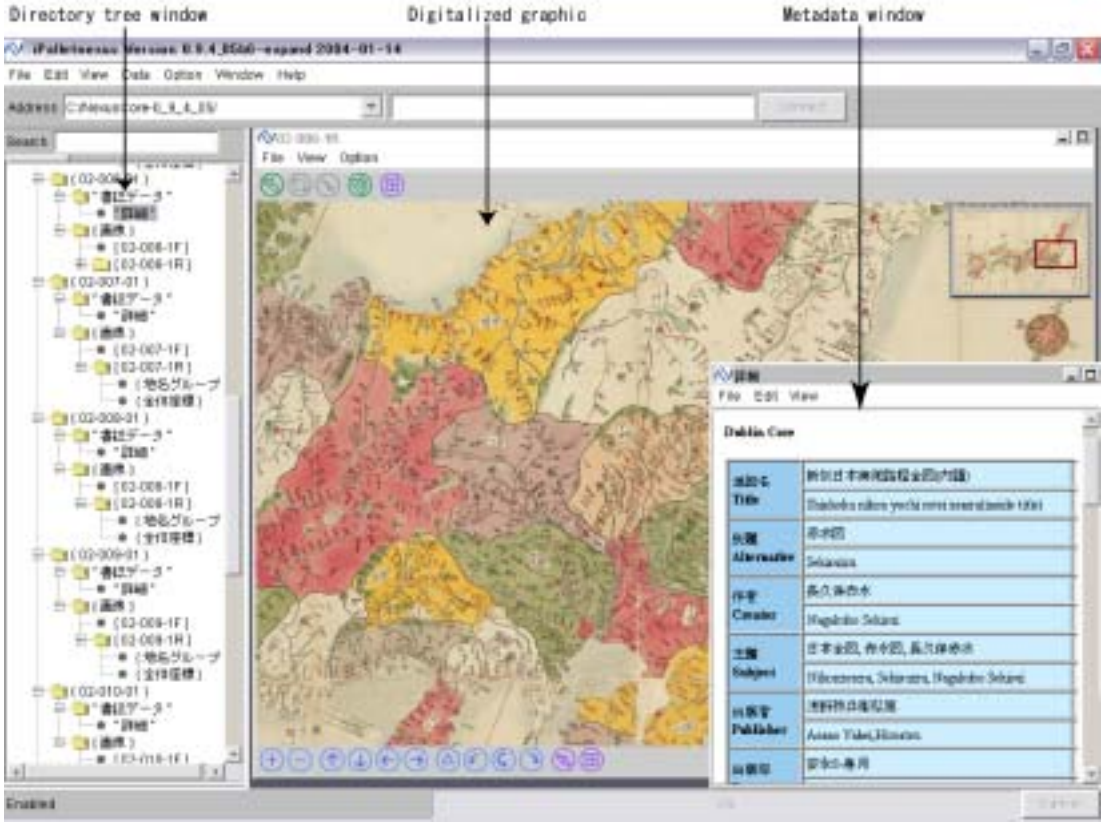
iPalletnexus is an application that aids in the research related to historical materials by reading digitized images such as pictorial diagrams, paintings, antique books and ancient documents and, allow browsing them while freely zooming in/out and attaching annotation where needed. An index file called Pal describes the relation between each image, metadata and annotation. By reading this index file, the application can display a list of image metadata and corresponding annotations in the left window in the form of directory structure (Figure1). By clicking on an icon in the directory, user can see corresponding images, and the metadata in the right window. The data handled on iPalletnexus uses a generic data format, such as JPEG format for images and XML for pal and metadata, providing some versatility. As the metadata is described in XML and displayed in XSL, they can be displayed on generic browsers, etc.

2.2 Metadata in iPalletnexus

iPalletnexus can be used individually on a desktop computer, or it can also be adapted to a network environment. By providing an iPalletnexus server for storing Pal, various kinds of images and metadata can be accessed through the iPalletnexus on each client to this server via the network, and each image and metadata can be browsed via the network. Also, a user with editing authority can edit and add metadata on the network. Thus, researchers in distant places can work collaboratively by mutually deciding task sharing, searching for necessary information, or rewriting the metadata on the server. Particularly when a research is under development and the detailed information of the resources of a research subject are not fully clarified, the information can be stored in the metadata and update them as the research progresses. Thus, a function to update metadata through task sharing can be provided so that the progress of the latest research of the research group is always reflected in the metadata.

Currently, this iPalletnexus has been introduced to dozens of research institutes in Japan and it has been used in the research of various historical materials such as antique maps, classical books, pictorial diagrams, ancient documents and manuscripts.

Figure 1. iPalletnexus Interface



3 Application of Dublin Core to Metadata

3.1 Designing Metadata Element Set

The development group has pursued research as to what kind of structure for an element set of the metadata should be applied to each resource used on iPalletnexus. iPalletnexus is adaptable to various kinds of international standards and user's original metadata scheme.

However iPalletnexus must handle a variety of cultural resources such as images, documents, and 3D objects as images. If user adopted various metadata schemes for each resource genres, several kinds of metadata would be intermingled in one project. If several kinds of metadata are intermingled, user can not search and compare various resources beyond genres. Thus, development group decided to design a typical metadata element set adaptable to various kinds of historical material for iPalletnexus.

The development group examined what elements are needed as typical metadata model for iPalletnexus. After examination, the element set is divided roughly into three parts. In part A information on basic contents ,for example title, creator, date and rights are included. Part B is related to information about physical object, for example physical shape of the resource, type of paper/canvas, printing method, painting method, type of ink/paint, extent of damage, number of colors, and dimensions. Part C contains information on digital data, for example digital format, data size, data rights, digitalizing method, etc.

The development group examined existing metadata schemes for example MARC, ISAD (G) [7]/EAD [8], VRA Core [9], CDWA [10], Dublin Core, etc. However, most of the schemes are not adjustable to the entire elements set that the development group needed. And some schemes are adjusted to specific resources such as books, art or documents and designed for western cultural resources mainly. Therefore, it is difficult to adjust them to Japanese historical material information in detail. Thus, development group considered making completely original metadata element set for iPalletnexus. Nevertheless, a completely original metadata element set causes a barrier when dealing with other metadata databases. It is desirable to keep versatility for

cross-search and sharing information between databases in network circumstances, if possible. Hence, the development group decided to separate metadata element set into three parts and adapt international standard metadata elements set to Part A only as elements of part A is easy to be adapted to metadata for their simplicity. By considering various metadata schemes, element set of Dublin Core is the most adjustable to part A of metadata, because it has only minimum necessary elements and corresponds to part A in reality. In the parts B and C, elements are designed newly for physical object and digital data information. This metadata is described in XML, and Parts A, B and C each has their information described on different namespaces, which has been divided into three parts in the XML. By doing this, a single metadata file could concurrently contain three divisions, A. Information about contents of resource (=Dublin Core), B. information about physical object, and C. information about digital data.

3.2 Adopting Dublin Core Metadata Element Set to iPalletnexus

For adjusting the Dublin Core metadata element set to Part A of iPalletnexus, we adopted only the elements that can be adjusted and all the 15 elements were not used. Specifically, the adopted elements were title, creator, subject, description, publisher, contributor, date, identifier, source, language, relation coverage and rights. Other elements of type and format, were rejected. This is because, for "type", the recommended term, DCTYPE [11], had limitations and was not applicable to every historical material that is supposed to be used in iPalletnexus. Thus, "type" was rejected to describe more precise information, and a "description" prescribed in Part B was used regarding the detailed description of resource features. Also, "format" needs to be described for both the description of the physical paper format and the digitized data format. Since inputting both data to the same element would cause confusion, this element was not adopted either. Information about physical paper format is stored in a form element in Part B and information about digital data format is stored in a form element in Part C.

In Part B, various kinds of elements were provided to store detailed information on physical object that could not be stored in Part A. First, a form element was provided to store physical information such as the physical shape of the physical object resource, type of paper/canvas, printing method, painting method, extent of damage, dimension, quantity and detailed information about contents of resources. The information is large in volume and the described content covers a wide range. Thus, in order to store detailed information on the physical object, a single element that could describe a wide variety of information was created instead of multiple elements that could describe singular information. This eliminated the need for various elements per resource and, since the framework is free, all information including extremely detailed information and specialized information that are not shown elsewhere could be covered. Also, not just physical shape, but for content description, a comment element was provided to store information that could not be fully described in the description element in Part A. In addition, we decided to describe information such as the copyright of resource in Part A, However, if there is a separate ownership, a right element is provided in Part B and the information is stored therein. Regarding the identification numbers, in Part A, a digital image is assigned an identification number but may be assigned a different identification number when the resources are physically separated for the management purposes in reality. So, when an identification number is assigned based on physical shape, an identifier element is provided in Part B and the information is stored therein.

Finally in Part C, an element to store information on digital data is provided mainly. First, a producer element is built for storage of information on the person responsible for digitized data of physical resources. Next, a data element is provided to store information by year, month and date when resource data is digitized. Also, a specification is made to store information such as the size and format of the digital data. Furthermore, an identifier element is built to store an identification number of digital data, and a comment element is provided for store other information on the digital data, and a rights element is supplied for storing information on the digital data and the person responsible for disclosure. Using these elements, information on digital data storage is determined (**Table 1**).

Regarding language entry, we decided that only the Dublin Core section of Part A would contain concurrent English notations to add English tags in XML [12]. This is because all major basic information is covered in Part A, and detailed information such as that in Part B contain many terms that are difficult to translate into English, which would complicate metadata creation process. Therefore, Japanese and English tags are prepared for all elements in Part A in order to allow description in both

languages.

Through this procedure, information about contents of resource, physical object and digital data can be stored in a single metadata file by applying the Dublin Core format. This allows iPalletnexus to handle various historical materials by storing in separated divisions in a single metadata file.

Table1. Metadata Element Set in iPalletnexus

A: Information about contents of resource (Dublin Core)		
Element name (English)	Element name (Japanese)	Element descrption
title	名称	Title of the resource
creator	作者	Name of the creator of the resource
subject	主題	Theme of the resource
description	解説	Simple explanation of the resource contents
publisher	出版者	Publisher and editor of the resource
contributor	寄与者	Contributor to the creation of the resource
date	出版年月日	Year the resource was created
type	—	Not used on iPalletnexus
format	—	Not used on iPalletnexus
identifier	識別番号	Identification number of the resource
source	来歴	Source of resource
language	言語	Language(s) used on the resource
relation	関係	Relationship with other resources
coverage		
spacial	地域	Location handling the resource
temporal	時代	Time period handled on resource
rights	権利	Rights of resource

B: Information about physical object		
form	形状	Detailed information on physical shape of the resource
comment	所見	Detailed explanation of contents of the resource
rights	所有に関する権利	Holder of the resource
identifier	現物管理番号	Identification number to manage the corresponding to physical shape of the resource

C: Information about digital data		
producer	デジタルデータ作成者	Person responsible for digitization of the resource
date	デジタル化年月日	Date of digitization
specification	デジタルデータ仕様	Information on digital data specifications
identifier	デジタルデータ識別番号	Identification number of the digital data
comment	デジタルデータ所見	Commentary information on the digital data
rights	デジタルデータ権利	Rights and person in charge of the digital data and its disclosure.

4 Application to Japanese Antique Map, Sekisui-zu

The previous section explained the basic structure of the metadata on iPalletnexus. This section uses an actual example of a Japanese antique map to explain how the prescribed features of the metadata of digitized historical materials are actually used on iPalletnexus.

The Sekisui-zu digital archive project which is carried out by The University of Tokyo, has digitized data of a map called "Kaisei-nihon-yochi-rotei-zenzu" (hereafter referred to as "Sekisui-zu"), created in latter half of the Edo period by Japanese geographer NAGAKUBO Sekisui (1717-1801), and we are conducting an examination and research on it using iPalletnexus. This Sekisui-zu is a large map, approximately 120 cm in height and 80 cm in width, and was photographed and digitized to allow browsing using iPalletnexus. The photographed image data was digitized and compressed so that it can be readable on iPalletnexus.

In Japan, many museums, libraries and archives possess various types of Japanese old maps. However, forms and rules of metadata vary across organizations. There are currently no standard rules or styles for them [13]. The Sekisui-zu digital archive project decided to adopt metadata element set introduced in section 3 to Sekisui-zu. It has various description methods, printing methods, and color usage, depending on its edition, and such detailed information must be stored in metadata from the map research standpoint. Thus, the following 24 information points were given as necessary items for storage as the metadata of Sekisui-zu.

(1) Name of the Map, (2) Name of creator, (3) Keywords, (4) Publisher, (5) Year the map was published, (6) Identification number, (7) Era and region covered, (8) Languages, (9)Rights, (10) Kind of paper, (11) Printing method (12) Extent of damage to the map, (13) Size, (14) Number of colors, (15)Other detailed information about map, (16) Identification number for controlling the physical map, (17)Holder of the map, (18) Creator of the digital data, (19) Year, month and date of digitization, (20) Size of the digitized image data of the map, (21) Digital data volume, (22) File format of the digital data, (23) Identification number of digital data, (24) Holder of the digital data

Of these items, information in Part A is applied to corresponding Dublin Core element. Results are presented in **Table2.**

Table2. Element Set in Part A

Information	Element (DC)
(1) Name of the map	Title
(2) Name of the creator	Creator
(3) Keywords	Subject
(4) Publisher	Publisher
(5) Year the map was published	Date
(6) Identification number	Identifier
(7) Era and region covered	Coverage
(8) Languages used on the map	Language
(9) Rights	Rights

Meanwhile, the contributor, description, resource and relation were not used at this time since there were no applicable items in the Sekisui-zu project case.

Next in Part B, the descriptions of information of the physical map are applied to elements as shown in **Table3**.

Table3. Element Set in Part B

Information	Element
(10) Kind of paper	Form
(11) Printing method	
(12) Extent of damage	
(13) Size	
(14) Number of colors	Comment
(15) Other detailed information about map	
(16) Identification number for controlling the physical map	Identifier
(17) Holder	Rights

Finally in Part C, mainly digital data and disclosure information is stored. The information was applied respectively to corresponding elements in Part C as shown in **Table4**.

Table4. Element Set in Part C

Information	Element
(18) Creator of the digital data	Producer
(19) Year, month and date of digitization	Date
(20) Size	Certification
(21) Volume	
(22) File format	
(23) Identification number of digital data	Identifier
(24) Holder of the digital data	Rights

By using each of the above elements from Part A to Part C, all metadata required for research on Sekisui-zu, the Japanese traditional map, could be stored. Information in each of these parts were described in XML and each part was described with a different namespace on XML as described in Section 3. Also, as described in Section 3, only the Dublin Core part used both Japanese and English languages. Since Sekisui-zu is an antique map, it partially used old Kanji (Chinese) characters; these were described after being replaced with other Kanji characters that could be displayed in the UTF-8 character code. Through this procedure, both basic information based on the versatile standard and the detailed information required for academic research of Japanese antique map metadata, successfully coexist in a single XML file (**Figure 2**).

Figure 2. Metadata in XML

Part A

```
<?xml version="1.0" encoding="UTF-8"?>
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:original="http://www.hi.u-tokyo.ac.jp/personal/baba/meta/original/"
xmlns:digital="http://www.hi.u-tokyo.ac.jp/personal/baba/meta/digital/">

    <dc:title xml:lang="ja">新刻大日本程路全図(外題),新刻大日本全図(内題)</dc:title>
    <dc:title>Shinkoku dainihon rotei zenzu(outside title),
              Shinkoku dainihon zenzu(inside title)</dc:title>
    <dcterms:alternative xml:lang="ja">赤水図</dcterms:alternative>
    <dcterms:alternative>Sekisuizu</dcterms:alternative>
    <dc:creator xml:lang="ja">長久保赤水</dc:creator>
    <dc:creator>Nagakubo Sekisui</dc:creator>
    <dc:subject xml:lang="ja">日本全図, 赤水図, 長久保赤水</dc:subject>
    <dc:subject>Nihonzenzu, Sekisuizu, Nagakubo Sekisui</dc:subject>
    <dc:description xml:lang="ja"/>
    <dc:description/>
    <dc:publisher xml:lang="ja">山城屋佐兵衛, 加賀屋善蔵, 藤屋徳兵衛, 伊丹屋善兵衛, 河内屋太助</dc:publisher>
    <dc:publisher>Yamashiroya Sahei,Kagoya Zenzou,Hujiya Tokubei,Itamiya Zenbei,Kauchiya Tasuke,Kouchiya</dc:publisher>
    <dc:contributor xml:lang="ja"/>
    <dc:contributor/>
    <dc:date>1864</dc:date>
    <dc:date xml:lang="ja">元治2</dc:date>
                                                    ≀
```

Part B

```
<original:identifier>O-02-011-1</original:identifier>
    <original:form xml:lang="ja">形状：一紙(畳), 寸法：720mm×1007mm,畳寸法：244mm×169mm,地図本体彩色:6色</original:form>
    <original:comment xml:lang="ja">内容構成：表紙,序文,凡例,廻沙考証部,方格線,方位盤</original:comment>
    <original:comment xml:lang="ja">出版者詳細：山城屋佐兵衛, 加賀屋善蔵(角印陽刻１顆),
                                    藤屋徳兵衛(長方印陽刻１顆), 伊丹屋善兵衛(角印陽刻１顆),
                                    河内屋太助(角印陽刻１顆), 河内屋吉兵衛(角印陽刻１顆、黒印),
                                    美濃屋庄兵衛(角印陽刻１顆)</original:comment>
                                                    ≀
```

Part C

```
<digital:identifier>D-02-011-1</digital:identifier>
    <digital:producer xml:lang="ja">赤水図デジタルアーカイブプロジェクト(イバレットネクサス開発グループ,情報学環,
                                    情報学環歴史情報論研究室,株式会社 堀内カラー)</digital:producer>
    <digital:date>2003-03</digital:date>
    <digital:specification xml:lang="ja">ピクセルサイズ:横20000pixel×縦13400pixel(周辺余白含む),データ量:50.1MByte,
                                         解像度:359.8pixel/inch, 色空間:RGB, 色深度:各色8bit,
                                         カラープロファイル:sRGB IEC61966-2.1(作業空間),
                                         閲覧フォーマット:FreeZoom Pack(iPalletnexus専用
                                                          /内部形式JPEG圧縮品質80</digital:specification>
                                                    ≀
    </metadata>
```

5 References and Search of Metadata

5.1 Displaying metadata

By creating an XML-format metadata as described in Section 4 and correlating it with a
Pal file image as described in Section 2, relevant image and a list of metadata are displayed in the left window of iPalletnexus. By clicking on an icon on the left window, user can open the metadata window on the right side, with metadata stored by the Dublin Core of Part A being displayed in Japanese and English (**Figure 3**). Also, by clicking on the detail icon on the left tree window, "B. Information on the object" and "C. Detailed information on digital data" will be displayed additionally in the metadata window. Displaying only the Dublin Core part in Part A is sufficient for general use. However, if information for academic research or for digital data control is required, displaying Part B and C allows covering from the basics to specialized use of a single metadata file.

Figure 3. Metadata Window

5.2 Search Window

The metadata information can be browsed not only by selecting from the left window but also by searching metadata using specific terms in the search window. For searching, first you have to open a search window using the iPalletnexus menu bar. (**Figure 4**) The Search window contains forms such as Title, Publisher, Date and Identifier that correspond to Dublin Core format. Then you have to enter a keyword in this form and press the search button. A list of metadata containing relevant information is displayed in a new window. You can find a map image by clicking on the displayed metadata. Also, if you want to search about Part B and C, enter the keywords in the Keyword form. The keyword form and the forms such as Title, Publisher, date and Identifier can be conducted simultaneously "and" searched leaving a high accuracy for refinement of resources. A term can be searched for in either the Japanese or the English language.

Figure 4. Search Window

6 Conclusion and Prospects

6.1 Conclusion

Through the measures described in this paper, a metadata that balances versatility and expertise use on iPalletnexus could be constructed. iPalletnexus makes the handling of various historical materials possible, and the metadata of all resources can maintain a commonality by using the Dublin Core in Part A for basic items. Also, all necessary information, including resource-specific information, as well as information on digital data that cannot be stored in Part A, can even be covered just by using Part B and C. This allows more detailed information to coexist in metadata while maintaining a primary simple element set of the Dublin Core. And, by using the metadata of this structure, it is not necessary to remake an element set of metadata for each resource; necessary information can be stored as is in the default element set.

These element sets were created with the assumption that they will basically be used within iPalletnexus, however, since they are in a highly versatile XML format, this basic structure could be applicable to other web sites disclosing digitized historical materials in libraries, museums and archives. In this way, Part B and C structures may be changed according to the conditions and user target of each site. However, if a pure Dublin Core is applied in Part A, any site or database handling a resource can maintain a certain commonness. This should simplify the cross-search, etc., among multiple sites and databases.

6.2 Prospects

The metadata model described in this paper can be used not only on iPalletnexus but also on other sites or databases that exhibit digitalized Japanese historical material. We think it is possible to use the element sets of Dublin Core on web sites or in databases that handle historical materials with information that have already been digitized as metadata in the libraries, museums, archives and research organization in Japan. The metadata model does not deny metadata element set or cataloging rules for physical object, but comprehend existing metadata models. So far, metadata element sets or cataloging rules for various Japanese historical materials have not been standardized. But if the concept described in this paper is adopted for the digitalized historical material, barriers of exchanging metadata between the various organizations in Japan will be removed to a cretin extent. In order to accomplish this, promotion of the concept in this paper is necessary through the full use of the characteristics of the current metadata.

Acknowledgment

Special thanks to Isao Yokoyama,
Chandrasiri N.P,
Masataka Yoshida, Akiyoshi Tani,
Shinji Yamane,
HORIUCHI COLOR LTD,
Center for Research on Pictorial Cultural Resources and Historiographical Institute in The University of Tokyo.

References

[1] Simon C.Lin, Ya-ning Chen;
A Metadata Case study for the FRBR Model Based on Chinese Painting and Calligraphy at the National Palace Museum in Taipei,
Proceeding of the International Conference on Dublin Core and Metadata Applications, 2001.
[2] Nicholaos Mourkoussis, Martin White, et al.;AMS-Metadata for Cultural Exhibition using Virtual Reality, *Supporting Communities of Discourse and Practice-Metadata Research &Applications*,2003.
[3] JAPAN MARC
http://www.ndl.go.jp/jp/library/
[4] USMARC
http://www.loc.gov/marc/
[5]iPalletnexus development group
http://www.ipallet.org
[6] Kou Hida, Akira Baba, Mitsuhiro Tsuda;
iPalletnexus: Development of Research Support, *Proceeding of the Nara Symposium for Digital Silk Roads*, 2003.
[7] ISAD (G)
http://www.ica.org/
[8] EAD
http://www.loc.gov/ead/
[9] VRA Core
http://www.vraweb.org/vracore3.htm
[10] CDWA
http://www.getty.edu/research/conducting_research/standards/cdwa/
[11] DCTYPE
http://dublincore.org/documents/type-element/index.shtml
[12] Wonsook Lee, Shigeo Sugimoto, et al.;
A subject gateway in Multiple Languages: a Prototype Development and Lessons Learned, *Supporting Communities of Discourse and Practice-Metadata*

Research & Applications, 2003.
[13] S.Suzuki; Description of Catalogs about the Japanese Old Map, *Journal of Japan Indexers Association Vol.13, No.2*,1984.