



Date 2<sup>nd</sup> version : 14/06/2006

## **New Innovative Access to Educational and Cultural Multimedia Contents**

### **Yuka Egusa**

Educational Resources Research Center, National Institute for Educational Policy Research  
6-5-22 Shimomeguro, Meguro-ku,  
Tokyo 153-8681, Japan  
[yuka@nier.go.jp](mailto:yuka@nier.go.jp)

### **Takashi Nagatsuka**

Tsurumi University, Dept. of Library,  
Archival and Information Studies Tsurumi 2-1-3,  
Tsurumi-ku,  
Yokohama 230-8501, Japan  
[nagatsuka-t@tsurumi-u.ac.jp](mailto:nagatsuka-t@tsurumi-u.ac.jp)

<b>Meeting:</b>	<b>97 Information Technology with Audiovisual and Multimedia and National Libraries</b>
-----------------	---

<b>Simultaneous Interpretation:</b>	<b>No</b>
-------------------------------------	-----------

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL  
20-24 August 2006, Seoul, Korea  
<http://www.ifla.org/IV/ifla72/index.htm>

## **Abstract**

This paper describes how we have been able to provide a new kind of innovative access to the invaluable content of old textbooks published in Japan before and during the World War II, the reports of research projects funded by Japanese Government, and also the classical Japanese ancient poems and maps at Tsurumi University. These contents were not available for public access before. Through these projects we were able to acquire and develop important skills in multi-media and hypertext presentation. These overall processes were very labor-intensive and also presented many difficulties even in the case of single digital content. There were substantially more difficulties in the case of multiple digital contents.

# **1 . Introduction**

The patrons of libraries are familiar with the capability of the World Wide Web (Web) to provide access to a wide variety of documents and information. Digital information is developing into a vast resource of information, and yet many users will be frustrated by the inability to find what they want and when they want it.

A digital library, a collection of information which is both digitized and organized, provides us with a power we never had with traditional libraries, nor with the Open or Surface Web.

What will be the content of a digital library? It can either be new material prepared digitally or old material converted to digital form. Content then needs to be stored and retrieved. Information is widely found in the form of text stored as characters. There are many old books and cultural materials in the world. These printed pages, as well pictures or photographs are scanned and stored. Recently, audio, video, and interactive material is accumulating rapidly in digital form.

Once stored, the content must be made accessible. Information retrieval systems are needed to let users find things not only for text but also for pictures, sounds, and video. The digital libraries must deliver the content to the users and also have an office of preservation of sorts.

This paper describes how we have been able to provide a new kind of innovative access to the invaluable content of old textbooks published in Japan before and during the World War II, the reports of research projects supported by Japanese Government financially, and also the classical Japanese ancient poems and maps at Tsurumi University. These contents were not available for public access before.

## **2 . Types of Information Sources on the Web**

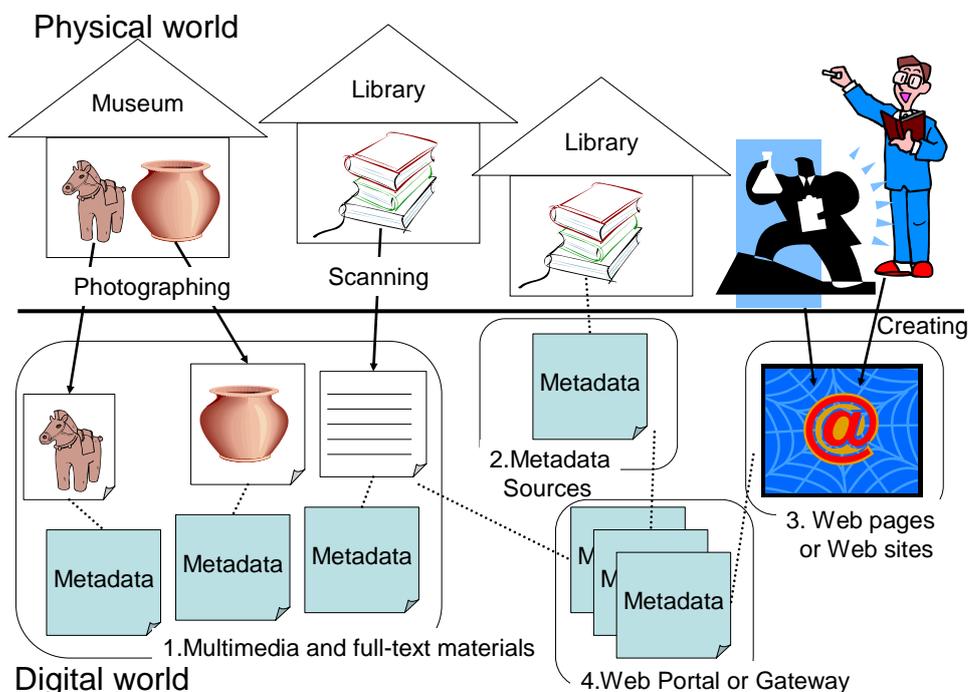
Many types of information sources are made available as part of the Web. We shall answer questions such as: What are the different types of information sources? Where do they originate?

The digital information sources in libraries, museums or archives are various and are tailored according to the needs of the users and the aims of the organizations. The type of information you need will also change depending on the question you are trying to answer at a certain time.

Figure 1 shows an overview of the various types of information sources on the Web. These information sources could be classified in a variety of ways and we have chosen the following:

- \* Multimedia and full-text materials were digitalized from the physical objects or created in the digital formats at first known as born-digital contents.
- \* Metadata sources to cover catalogues, indexes and abstracts, or sources that are provided as “information about information”.
- \* Web pages or Web sites.
- \* Web Portal or Gateway provides direct online access to many kinds of information sources in a particular field.

These categories are based on the way to link the physical world entities with the digital world objects.



**Figure 1:** An overview of the various types of information sources on the Web

## 2.1 Multimedia and full-text materials

Multimedia and full-text materials digitalized from the physical objects in libraries, museums or archives are various and are tailored to meet the needs of the users and the aims of the organization.

There are many steps in the digitization process for multimedia and full-text materials digitalized from the physical objects in libraries, museums or archives. Some sources involve many steps such as the scanning process of page images by 2D or 3D scanner, converting the archival images to PDF files, and the texts encoding in a special tag set, etc.

During the 1990s both research and professional interest in digital libraries continued to grow rapidly. Many research projects were funded by the Japanese Government since the late 1990s. The earliest digital library experiments were essentially the domain of computer scientists. The experimental projects had little to say about libraries as understood by librarians. Some university libraries took an active part in the projects in Japan. Kyoto University Digital Library project (<http://ddb.libnet.kulib.kyoto-u.ac.jp/minds.html>) focused on developing a digital collection about old books, maps, scrolls and creating high-resolution images from their image collections such as maps.

The Modern Literature Digital Library project by The National Diet Library (NDL) since 2003 provides an image database (<http://kindai.ndl.go.jp/index.html>) of the books published in the Meiji era (1868-1911) and held by the NDL now. The books whose copyright had expired were selected from the collection for available to the public.

The Japan Center for Asian Historical Records (JACAR) (<http://www.jacar.go.jp/>) provides materials from the collections on the modern history of relations between Japan and other countries, primarily neighboring Asian countries which are held by some ministries and agencies of the Japanese Government.

Aozora Bunko (<http://www.aozora.gr.jp/>) is a digital library of Japanese literature works. Many volunteers work for the Aozora Bunko project to input the full-text data from the literatures similar to the Gutenberg Project.

## **2.2 Metadata Sources**

Metadata sources cover catalogues, indexes and abstracts, or sources that are provided as “information about information”. Bibliographic records have a tradition in this category and are in very common storing summary information. As for books in which there are a lot of OPAC databases on the Web today, there are location databases for museum materials, too.

Most university and public libraries create OPAC databases, and most of them are available on the Web. According to Japan OPAC list created by Agriculture, Forestry and Fisheries Research Council (<http://ss.cc.affrc.go.jp/ric/opac/opaclist.html>), 592 OPAC databases are available in Japan as of May 2006. There are two large scale OPAC databases in Japan, NDL-OPAC (<http://opac.ndl.go.jp/index.html>) and NACSIS

Webcat (<http://webcat.nii.ac.jp/>). On the other hand, as the samples of focusing the specific discipline, the Library of Education (<http://www.nier.go.jp/homepage/jouhou/toshokan/index.htm>) at National Institute for Educational Policy Research of Japan (NIER) provides an abstract service of educational research papers. The Educational Resources Research Center (<http://www.nier.go.jp/homepage/jouhou/jouhou.htm>) at NIER also provides educational databases, such as the Education Ministry's curriculum guideline, the entrance examination of high schools.

## **2.3 Web pages or Web sites**

For many people the only digital information sources they think about using are Web pages or Web sites. They do not use digital libraries or databases even though relevant information sources may be available. In a certain case, Web pages or Web sites are useful information sources for them. In the case of learning materials, teachers make their teaching materials and then put them on their homepages in order to provide access to them for every person. Most universities provide their faculty information and syllabuses of the lectures in this way. Some universities make and share their teaching materials in an open-courseware (<http://www.jocw.jp/>).

## **2.4 Web Portal or Gateway**

Web Portal or Gateway provides direct online access to many kinds of information sources in a particular field. A lot of useful resources on the web are provided from different organizations or individuals, for that reason the resources in a particular field are not organized properly. Web Portal or Gateway will be useful tools for gathering specific topics from the web.

For example, the ULIS Digital Library project[1] created a subject gateway for the field of library and information science. Cultural Heritage Online (<http://bunka.nii.ac.jp/>) provides a feature of the gateway to heritage resources which originate from major museums in Japan. The National Information Center for Educational Resources (NICER) organizes and provides educational materials in Japan with metadata on the Web (<http://www.nicer.go.jp/>). The National Institute of Japanese Literature's cross-museum search DB project (<http://www.nijl.ac.jp/~kiban-s/>) creates the gateway of social science resources and provides the facility of cross-database search.

## 3. Library of Education's Collection

The Library of Education is one of the largest special libraries for educational resources in Japan, and aims to collect, organize and provide to access their collections for the members of institution and also make digital materials from the original sources available to the public on the Web.

The Library holds about five hundred thousand items including 148,000 for Japanese books, 39,000 for foreign books, 97,000 for textbooks, and 194,000 for others. The Library's collection consists of eight groups as follows: (1) books related to education, (2) magazines and journals related to education (including bulletins of colleges in Japan), (3) pre-war textbooks of Japan, (4) post-war textbooks of Japan, (5) foreign textbooks, (6) government document for post-war reform, (7) personal collections, and (8) research outcome reports.

Bibliographic information of the collection (1) and (2) is available on the Web. The Educational Research Paper Indexing Service (教育研究論文索引 in Japanese) provides article-level bibliographic records from the second group on the web. There is a plan to make a digital library. It is planning to create the metadata from the second to eighth groups and also the digital images from the groups of third, sixth and eighth by scanning digitally. It has been started the test version of these databases since October 2005 gradually. In the following section, details of this project are reported.

### 3.1 Cross-Collection Search

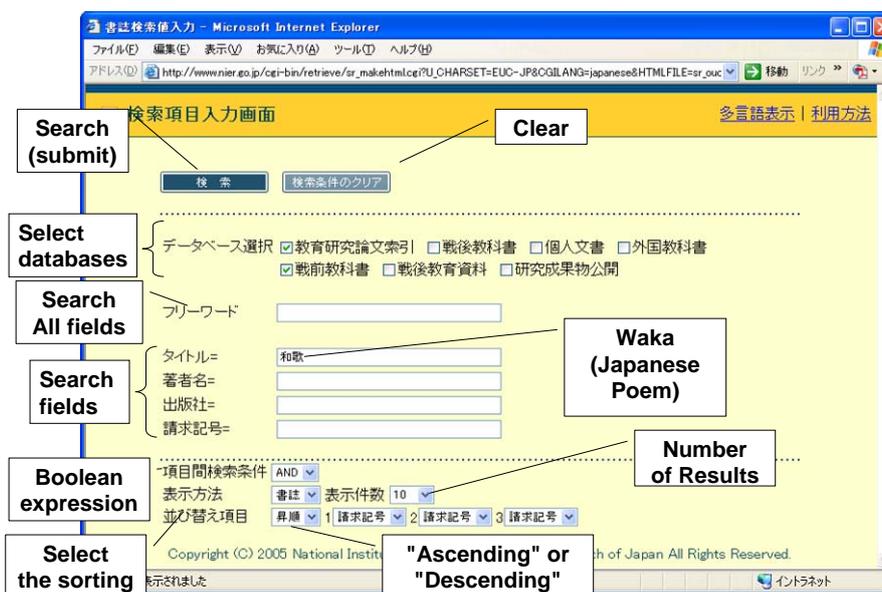
The Library holds a variety of collections such as old textbooks, educational reports, etc. Some users hope to search a whole set of collections seamlessly at the same time if need be. Others are interested in only part of the materials and need to specify their own needs toward a particular collection.

It was decided to provide two search interfaces to the users, Cross-Collection Search and Individual Collection Search. It can be searched using keywords with Boolean operators either across the entire collections or within specified collections. The function of Individual Collection Search has been restricted to search for a specified collection. Table 1 shows a cross-collection mapping of search fields in our system. It was selected four fields, Title, Author, Publisher, and Request No for the cross-search. These fields are common in all the collections, but have slightly different names in each collection. In addition, we also provide keyword search which allows users to search any words in the all fields which users can search individual collection.

**Table 1: A cross-collection mapping of search fields (translated in English)**

collection Fieldname	(2) Magazines and journals	(3) Pre-war textbooks	(4) Post-war textbooks	(5) Foreign textbooks	(6) government document for post-war reform	(7) personal collections	(8) research outcome reports
Any	Title of paper Journal title Volume Author(s) Transcription of author(s) Publication date ISSN Request No Keyword(s)	Title of book Other title Volume Edition Author(s) and editor(s) Publisher Request No Subject Sub-area of subjects Period and kinds of schools Audience (either for students or teachers)	Title of book Edition Author(s) and Editor(s) Publisher Publisher code Publisher code (in 2 or 3 letter code) Request No Added year in textbook catalog Used year Authorized year(in westan calendar) Authorized year(in Japanese era) Type of school Subject Sub-area of Subject Scholastic year grade Textbook mark Textbook No Audience (either for students or teachers)	Title of book Volume Author(s) and Editor(s) Publisher Country Publicationdate Request No Subject Grade category Audience (either for students or teachers)	Title of book Journal title Author(s) and Editor(s) Publisher Publication year(in westan calendar) Publication year(in Japanese era) Document code Keyword(s)	Title Journal title Collection name Author(s) and Editor(s) Publisher Request No Date	Title Type of research Principal researcher Publication date Fund ID Affiliation
Title	Title of paper Keyword(s)	Title of book Other title	Title of book	Title of book	Title of book Journal title Keyword(s)	Title Journal title	Title Type of research
Author(s)	Author(s) Transcription of author(s)	Author(s) and Editor(s)	Author(s) and Editor(s)	Author(s) and editor(s)	Author(s) and Editor(s)	Author(s) and Editor(s)	Principal researcher
Publisher			Publisher Publisher code Publisher code (in 2 or 3 letter code)	Publisher	Publisher	Publisher	
Request No	Request No	Request No	Request No	Request No		Request No	

It can be searched using keywords linked Boolean operators either across the entire collections or within the selected collections (Figure 2). An excerpt from the results of a search is shown in Figure 3.



**Figure 2: Cross-search interface for multi-collections**

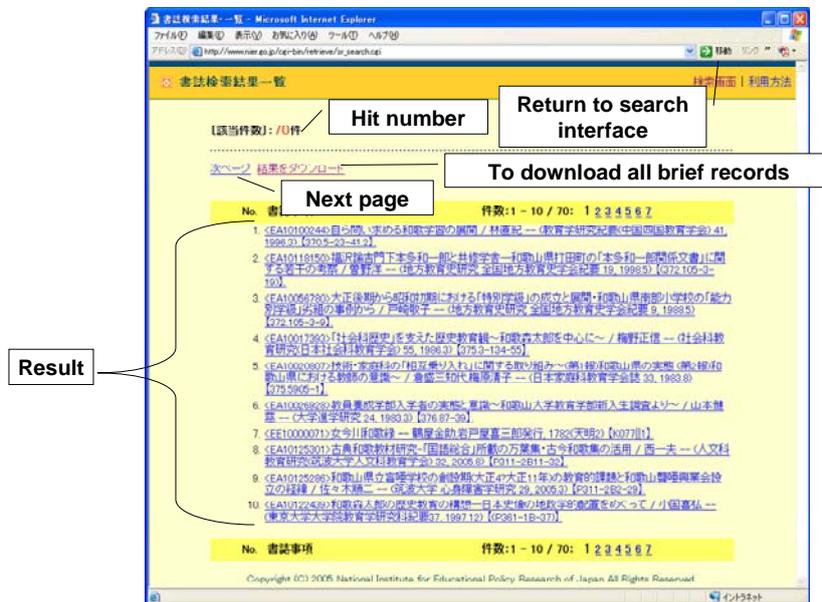


Figure 3: Cross-search result page

If a user finds an interesting item in the list, he/she can jump to the details of the item by clicking on it (Figure 4). If the image or full-text content is available for the item, it can be obtained by clicking the "Display this resource" to link into the image or full-text page. Figure 5 is a screenshot of the image content after clicking.

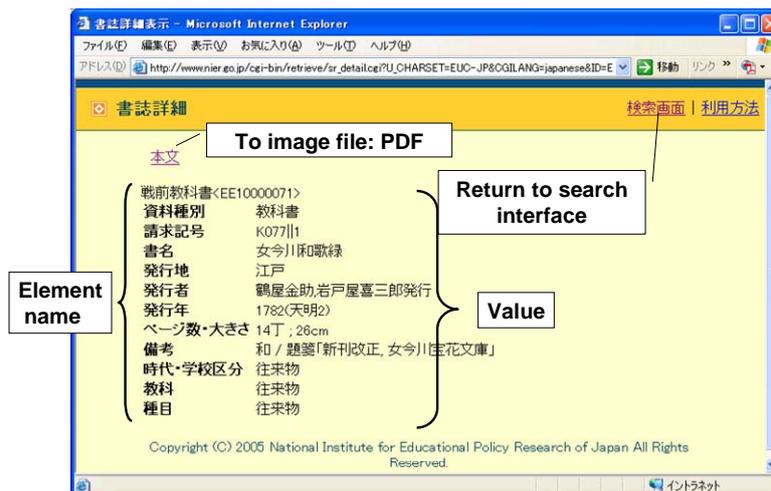


Figure 4: Bibliographic record page

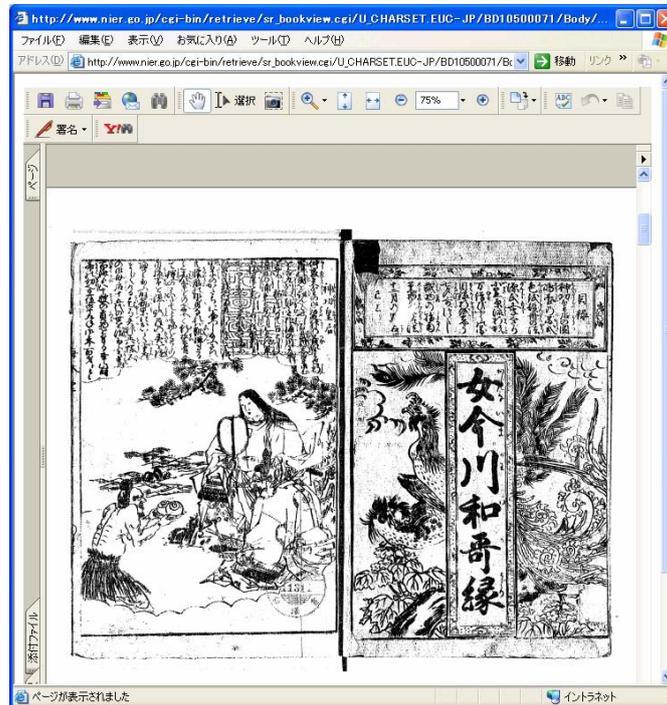


Figure 5: Scanned image of the original material

## 3.2 Pre-War Textbooks of Japan

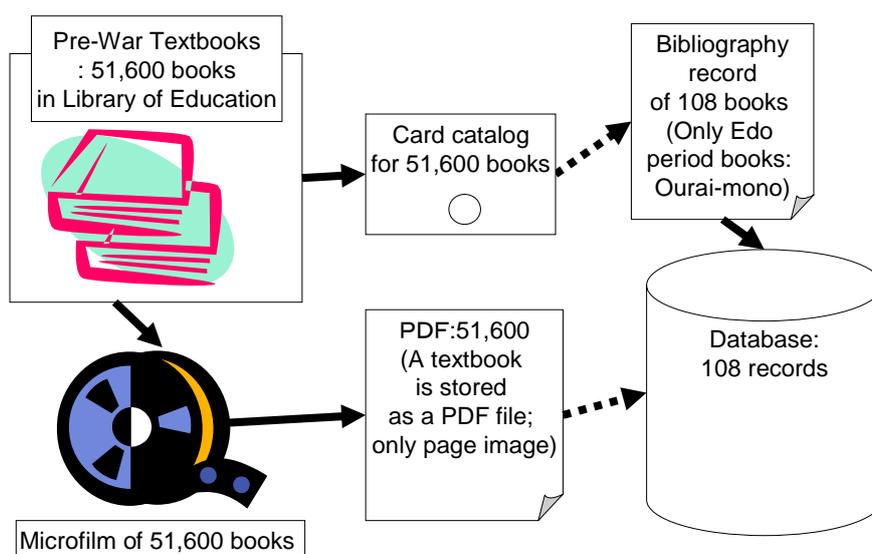
The Pre-War Textbooks of Japan are composed of about 51,600 old textbooks published in Japan from Edo-era (17th-19th centuries) to the end of the World War II (1945). Since the word of “pre-war” means before and during the World War II in Japan, this collection is called Pre-War Textbooks (戦前教科書 in Japanese).

Figure 6 shows the overview of the digitalization process of Pre-War Textbooks. The 108 digital bibliographic records of all Ourai-mono (往来物 in Japanese), textbooks published in Edo-era(17th-19th centuries), were created from paper card catalogues selected from the whole collection. To create the metadata of Pre-War Textbooks is expensive to undertake, and the project has only been able to proceed relatively slowly in capturing its large holdings in this way. One of the reasons for the higher cost of making digital bibliographic records from card catalogs is that indexing policies change from time to time.

Some catalogues are composed of several bibliographic records. The others possess the different levels or formats of descriptions. The staff of the Library planned to

digitize the whole collection at first, but it was eventually recognized it would require a rather long time to change all catalogues of the collection into the digital format.

Bibliographic elements of the collection are composed of Record ID, Title of book, Other title, Volume, Edition, Author(s) and editor(s), Publication location, Publisher, Publication year, Request No, Subject, Sub-area of subjects, Period and kinds of schools, Page numbers & physical size, Notes, Audience (either for students or teachers), and Record creation date.



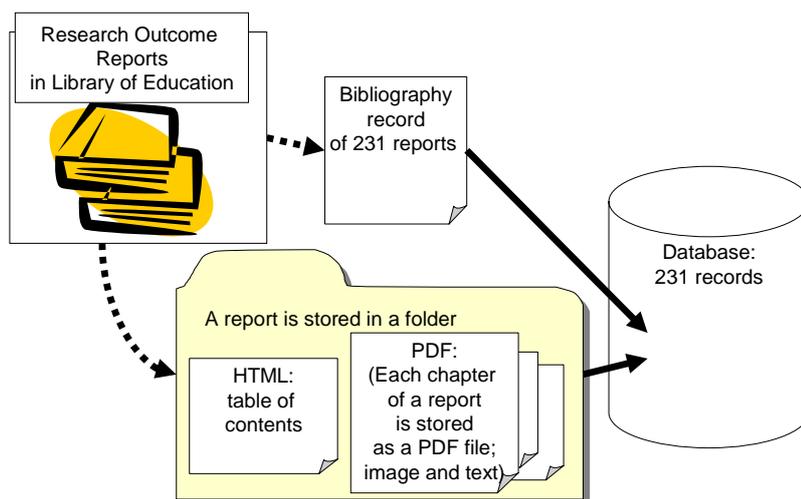
**Figure 6: The digitalization process of Pre-War Textbooks**

All Pre-War Textbooks were scanned and then were converted into microfilms and PDF files. Because microfilm can be preserved for a long time, it was chosen as a primary media format. The PDF file was made by scanning the microfilm at 400dpi monochrome which makes it perfectly readable for screen display. The filename of the PDF file is named after Request No of the card catalogue. It takes a long time to make digital bibliographic records of the whole collection. It will be important for users to be able to access the PDF file at least through the Request No of the card catalogues.

### 3.3 Research Outcome Reports

Figure 7 shows the overview of the digitalization process of Research Outcome Reports. Research Outcome Reports (研究成果物公開 in Japanese) comprises the digital content of the reports of research projects funded by the Japanese Government.

The Research Outcome Reports are composed of bibliographic records, images and full-text of the reports published from 2000 to 2003 by National Institute for Educational Policy Research of Japan. The 231 records made by Educational Resource Research Center are available on the database as of May 2006.



**Figure 7: The digitalization process of Research Outcome Reports**

The digital contents of the reports were created by the Center. The Center plans to expand the records of the reports published from 1989 to current. The image contents as the PDF format and full-text data inputted manually can be searched. One report is composed of several chapters. Each chapter is stored separately as the different PDF file which is linked from the HTML pages of the table contents.

## 4. Tsurumi Collection

The collaboration between Chen's group and Tsurumi University began with the development of a prototype collection, based on images included in two publications of the Tsurumi University Library - the *Eighty Selections of Waka Poems and Tales from the Classical Japanese Literature* and the *Japanese Maps in the Old Age*[2]. Global Memory Net (*GMNet*) directed by Prof. Ching-chih Chen of Simmons College is intended to be an effective gateway to world-wide cultural, historical, and heritage image collections from selective collaborative institutions throughout the world. Much of these unique, valuable education and research collections are not currently accessible due to distance, form, and technical barriers. *GMNet's* goal is to create and share the tools necessary for users to access and exploit these significant collections

via a global network. *GMNet* aims to be an effective gateway or digital portal to world culture and heritage. It can also be viewed as a comprehensive multimedia digital library which can offer instant access to materials of global culture, history, and heritage [3].

The Japanese contents are not yet sufficiently accessible by people in the world due to language and other factors, and that *GMNet* can provide a platform to make it possible for global sharing and distribution of some selective Japanese resources. *GMNet*'s great potential for "real" collaboration with different countries in the world not only in the "content" areas, but also the needs for involving subject specialists in those countries in order for *GMNet* to provide more in-depth knowledge on the significant cultural and historical "memories" of those countries.

As the project developed, coinciding with the technology development of *GMNet* in bilingual retrieval as well as with being able to hear the poems spoken, the inclusion of audio files for each of the *Waka* selection was considered a very desirable feature since *Waka* poems are generally only readable by very small number of specialists.

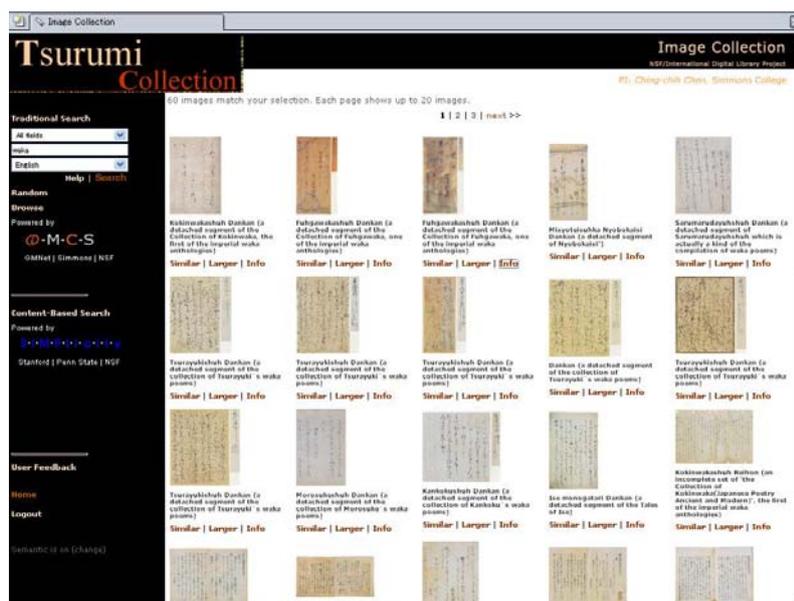


Figure 8: A part of Tsurumi's Image Collection in GMNet

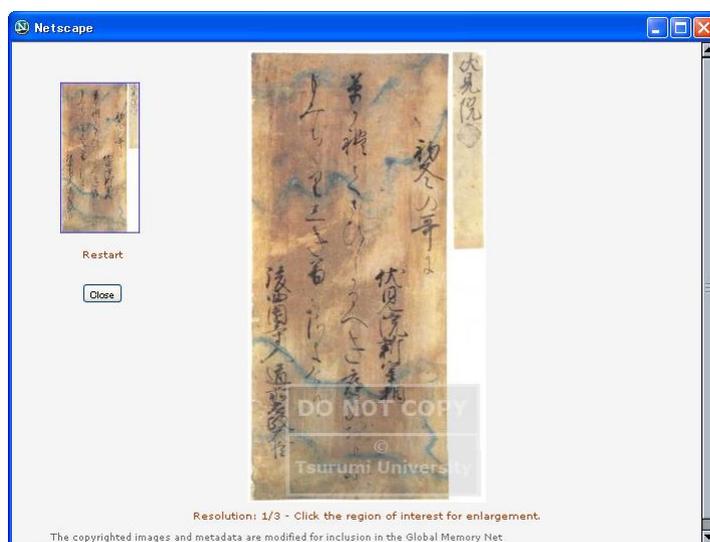
*Waka* poems and tales are rare and important classical Japanese literature, thus an invaluable part of Japanese cultural heritage. Hundreds of original texts, translations, studies, and electronic texts of classical Japanese literature are available on the Web, and many are in Japanese only. As to the *Waka* literature, there are several useful Web sites as well, such as the Japan 2001 *Waka* Website by the University of Sheffield which translates some Japanese texts of the imperial *Waka* anthologies to English (<http://www.shef.ac.uk/japan2001/>), and the site of the Japanese Text Initiative of the

University of Virginia Library's Electronic Text Center which displays classical Japanese literature including *Waka* poems in Japanese characters (<http://etext.virginia.edu/japanese/>). Yet, none has introduced the original look of the handwritings of the poems, presented sound files of the contemporary Japanese reading, and permitted bilingual retrieval capabilities. These are the features of our proto-type in *GMNet*.

When one enters the Tsurumi "image collection," the screen as shown in Figure 8 will show up. Where a user can decide to find what they want by searching in any of the metadata field or all. Here, one can browse images of the collection, or can randomly access the image collection. When the images are displayed, simple titles of these are also shown (Figure 8).

When "Info" of a specific image is clicked, additional descriptive information of that image can be obtained as shown in Figure 10. When "larger" is clicked, the image can be zoomed and enlarged several times to enable the seeker to view the details of the handwriting of the *Waka* poem. To protect the ownership of the image, dynamically generated digital watermark can be seen in the lower right corner of the enlarged image (Figure 9).

The retrieval of the Tsurumi collection in both English and Japanese can be possible on *GMNet*. If a user can input the Japanese characters from his/her PC with a bilingual



**Figure 9: The chosen image can be enlarged several times**

word processing software, it will be able to access to the Tsurumi Collection in Japanese. After the user has retrieved and received the output in English, he/she may like to read a more detailed description in Japanese. In this case, he/she can retrieve the descriptive information in Japanese. Figure 10 shows the detailed descriptive information in Japanese.



Figure 10: The bilingual retrieval of Tsurumi Collection in both Japanese and English in *GMNet*



Figure 11: The sound presentation of Tsurumi Collection in *GMNet*

It was considered that the inclusion of sound files for the Waka selection is truly a desirable feature since few people, even Japanese, are able to read the old and ancient

Japanese poems. Thus, the sound files for most writings were created by recording the readings digitally and then converted to flash files for the web presentation.

When a user retrieves materials in English, the descriptive record will be shown as Figure 11. By clicking on the Audio button locating the upper part of the record, the sound file will then narrate the transcript in Japanese.

## 5. Discussion and Conclusions

Through these projects, we have gained valuable experiences. These overall processes were very labor-intensive and also have many difficulties even in the case of single digital content such as in the process of preparing the metadata from the paper card catalogues and also the raw English descriptive data. See before, there were much more difficulties in the case of multiple digital contents.

A feature of the multi-database search with simple interface in Educational materials was developed. In the case of the multi-database search, it is very important to carry out mapping among several collections located within the organization for Cross-Collection Search. The search interface should be developed to adapt Z39.50-based protocol for the inter-library cooperation and cross-institute search. In addition to Z39.50 standard, OAI-PMH will be necessary for some collection such as Research Outcome Reports.

In the case of Tsurumi's collection, the old Waka poems and tales, as well as ancient maps of the Tsurumi University not only can be shared with those scholars and learners who are interested in classical Japanese culture and heritage, but also greatly enhanced their capabilities in obtaining needed multimedia information quickly and effectively in a way which was simply not possible before.

There are numerous digital library projects for preserving cultural heritage in Japan, but the great majority of them are mainly targeted at the Japanese professional or non-specialist in Japan. Obviously, to create on English access capability is essential for international sharing and information distribution, but it also presents great challenges at the moment. Our experience or Tsurumi University's participation in *GMNet* not only highlights this problem, but also suggests that there is a great and urgent need to explore wider possibilities in meeting these challenges.

Tsurumi's pilot project is an important step in international collaboration for introducing the Japanese cultural heritage, especially the classical Japanese Literature, to the world.

It is very important to focus on we create a detailed descriptive information in each digital collection. If we create the detailed descriptive information from a specific content, the users can select more appropriate content for the whole collection. One of

the authors proposed a database system of adding the descriptive information by the users' community[4].

We now know that we have much work ahead of us if we are serious in pushing our digital library initiatives.

## Acknowledgements

We greatly appreciate Prof. Ching-chih Chen directing GMNet Project and the efforts of Chen's GMNet group at Simmons College. We also thank to the staffs at The Library of Education for helpful comments on an early draft of this paper.

## Bibliography

- 1.Hiroshi Hiraoka, Takayuki Manaka, Toshiaki Yokoyama, Tetsuo Sakaguchi, Shigeo Sugimoto, Koichi Tabata. Digital Library System at University of Library and Information Science. *Journal of Information Processing and Management*. Vol.42, No.6, 1999, p.471-479. (in Japanese).
- 2.Takashi Nagatsuka and Ching-chih Chen. Global Memory Net Offers New Innovative Access to Tsurumi's Old Japanese Waka Poems and Tales, and Maps. In the Proceedings of the eighth International Conference on Asian Digital Libraries (ICADL2005), Springer, LNCS3815, pp.149-157, 2005.
- 3.Ching-chih Chen. Thai's Invaluable Memory Celebrated via Global Memory Net. In the Proceedings of the eighth International Conference on Asian Digital Libraries (ICADL2005), Springer, LNCS3815, pp.496-508, 2005.
- 4 .Yuka Egusa, Masao Takaku, Hidehiro Ishizuka. Network Community Oriented Information Sharing System for Databases. International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society DLKC'04. Tsukuba, 2004-03, p.64-71. <<http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp64.pdf>>.