

The Status of Compliance with OAIS Reference Model in the National Diet Library Digital Collections

Tsukasa Kimezawa
Digital Information Services Division,
Digital Information Department, National Diet Library
1-10-1 Nagata-cho, Chiyoda-ku,
Tokyo, Japan 100-8924
(+81) 335063328
t-kimeza@ndl.go.jp

Shuji Kamitsuna
Kansai-Kan,
National Diet Library
8-1-3 Seikadai, Seika-cho,
Soraku-gun, Kyoto, Japan 619-0287
(+81) 774981484
kamituna@ndl.go.jp

ABSTRACT

The National Diet Library (NDL) has been providing access to digitized library materials via the Internet since 2002. The NDL has been digitizing books and magazines continuously since then, and collecting digitized materials from other institutions. In addition to digitized materials, NDL began to collect online publications (electronic books and electronic magazines) that were not protected by Digital Rights Management (DRM) from the Internet in 2013. The NDL Digital Collections was launched in 2011 to collect, preserve, and distribute these materials. This paper provides an overview of the NDL Digital Collections and discusses current achievements as well as the challenges faced in effecting long-term preservation while meeting the functional requirements of the OAIS reference model.

Keywords

Digital preservation, Long-term accessibility, OAIS reference model

1. THE NDL DIGITAL COLLECTIONS

The NDL digitizes and provides access via the Internet for books, magazines, and other library materials. Prior to 2009, digitization of materials had been limited to materials confirmed to be in the public domain as a result of copyright investigation or for which permission had been obtained either from the copyright owners or the Agency of Cultural Affairs. The 2009 revision of the Copyright Law enabled the NDL to digitize books and magazines for preservation even if they were not yet in the public domain and to provide access to them on the premises of the NDL at the Tokyo Main Library, Kansai-kan, or International Library of Children's Literature. That same year, a major budget allocation for digitization enabled the NDL to digitize more than 2 million library materials. (Table 1)

Table 1. Number of digitized materials

Kind of Materials	Number
Periodicals	1,240,000
Books	900,000
Online Publications	300,000
Doctoral Dissertations	140,000
Rare books and Old Materials	90,000
Historical Recordings Collection	50,000
Others	70,000
Total	2,790,000

The original system, however, lacked both scalability of storage and the functionality to stream digitized recordings. In particular, there were several systems in place, which were difficult to operate and maintain individually in terms of cost and manpower.

The NDL Digital Collections (Figure 1) was developed and integrated with existing systems in 2011 to collect and preserve a wide variety of materials. Functionality for providing people with visual disabilities access to Digital Accessible Information SYstem (DAISY) materials was added in 2012, for enabling publishers to upload online publications with metadata to the NDL Digital Collections was added in 2013, and for collecting doctoral dissertations in electronic format as well as for transmitting digitized material to public libraries was added to the system in 2014.



Figure 1. The NDL Digital Collections

The NDL strives to conform to all legal requirements in digitizing books and magazines as well as collecting and preserving online publications from the Internet. It also provides a variety of services to diverse users, who include Diet members and their staff, employees of government agencies, and patrons from the general public, both in Japan and overseas. Digitized materials in the NDL Digital Collections have bibliographic data, which enables even patrons who lack any specialized knowledge to search and access digitized materials through the system's browsing function.

2. COMPARISON WITH OAIS REFERENCE MODEL

We have described how the NDL Digital Collections was designed to collect, preserve, and distribute digitized materials in variety of diverse formats. Next, we will discuss how it compares with the OAIS reference model (ISO 14712: 2012).

The OAIS reference model defines six functional entities: Ingest, Archival Storage, Data Management, Administration, Access and Preservation Planning. Digital information for preservation is handled as an information package in the OAIS reference model.

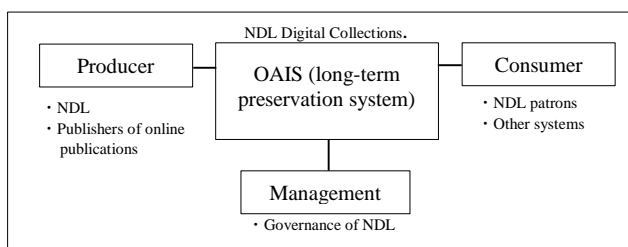


Figure 2. OAIS reference model and NDL Digital

The individual Archival Information Packages (AIP) that comprise the NDL Digital Collections are not preserved in a single comprehensive archival file. Instead, image files of digitized materials are placed in archival storage per each book or magazine. On the other hand, preservation metadata and package information are stored in a Relational DataBase Management System (RDBMS) that is separate from the archival storage. This configuration was adopted in consideration of accessibility to preservation metadata and package information as well as future flexibility to change information package file formats.

The NDL Digital Collections ingests numerous digitized materials by means of a collective registration method. This method directly collects digitized materials from external hard disks. There are also functions for collecting electronic books and electronic magazines via web crawlers or by uploading via a website.

The number, contents, and metadata of ingested materials are checked by staff using an administrative interface. The file format and required metadata items are checked by the system automatically. If necessary, the format of ingested image files is modified to a more suitable format for access and preservation. For example, an image file in TIFF format are converted to JPEG2000.

Preservation metadata is compliant with PREMIS ver. 2.2. Message digest, the date and hour of registration, the staff ID and the file format are recorded as fixed information. Persistent identifiers in info URI are given as reference information. DOIs are given to a part of digitized materials, such as doctoral dissertations in electronic format or rare books and old materials.

In addition to information obtained when materials are digitized, other bibliographic data is added from an integrated search service, NDL Search, for use as metadata when searching. An extended NDL Dublin Core Metadata Description (DC-NDL) is used as the format for metadata information.

We adopted the most suitable devices available for archival storage. In addition to the cost and capacity of devices for archival systems, other factors taken into consideration were creditability, read/write capability, and scalability. Since 2011, we have adopted a distributed file system and constructed a petabyte-class storage system.

Data management of information such as metadata, access restriction, and digital rights is performed by RDBMS.

Administration duties performed by NDL staff include negotiating with publishers and making policies. Monitoring tools are used to administer systems and to collect logs and statistics automatically.

Access management is performed by the NDL Digital Collections server, which converts digital material archived in JPEG2000 format to JPEG and transmits it per each patron request. Audio and video recordings are streamed to the patron from the server. DAISY materials are provided to visually impaired patrons via streaming and download.

3. NEXT STEP

We face a number of future challenges in providing functionality that is not yet implemented within the NDL Digital Collections. In particular, we have yet to implement Preservation Planning to provide environments for accessing to obsolescent file formats. We also need to negotiate with publishers regarding the preservation of online publications with DRM.

Ensuring long-term access to materials that are difficult to migrate involves preserving environments, including hardware and OS, and applications as well as instructions on how to use them. In the future, emulation will be an essential part of the NDL Digital Collections.

4. REFERENCES

- [1] National Diet Library, National Diet Library Digital Collections, http://dl.ndl.go.jp/?__lang=en
- [2] T. Kimezawa, "The Status of compliance with OAIS reference model in the "National Diet Library Digital Collections" Transformation from "Digital Library from the Meiji Era," *Journal of Information Processing and Management*, vol. 58, pp.683-693, 2015.
- [3] PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata. 2012, Version 2.2, 265p, <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>