

Digital collection strategies at the National Diet Library

Toshiyuki YAMADA

Director General at the Kansai-kan

National Diet Library, Japan



Copyright © 2013 by **Toshiyuki Yamada**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License

<http://creativecommons.org/licenses/by/3.0/>

Abstract:

The National Diet Library (NDL) has so far formulated “National Diet Library Electronic Library Concept”⁽¹⁾ and “National Diet Library Digital Library Medium Term Plan for 2004”⁽²⁾, and constructed digital collections by such means as digitization of its holdings and web archiving.

The NDL started to provide the digital contents converted from its holdings with the “Digital Exhibition”⁽³⁾ in 1998. Copyright-cleared books published in and after the Meiji era (1868-1912) have been provided on the “Digital Library from the Meiji Era”⁽⁴⁾ since 2002. Digitization in the NDL has been promoted, mainly owing to two factors: first, the Copyright Act was revised in 2009 to enable the NDL to digitize even holdings within copyright protection period for the purpose of preservation and to provide them in digital form within the NDL facilities; and second, a large budget was allocated for digitization in the supplementary budget of FY2009. Now, more than two million items of digitized contents⁽⁵⁾ are available.

In November 2002, the NDL started archiving Internet information resources based on permission from each site owner. In July 2009, the National Diet Library Act was amended to enable the NDL to collect and archive without permission the Internet information resources provided to the general public by public institutions. So far, the NDL has collected⁽⁶⁾ and provided more than 50,000 websites in the Web Archiving Project (WARP).

The NDL has also been working on the construction of the NDL Great East Japan Earthquake Archive, aiming to actively collect and preserve records on the Great East Japan Earthquake of March 11, 2011; the disaster-afflicted areas as they were before the earthquake, and their recovery process; and earthquakes, tsunamis and accidents of atomic power plants in the past.

On June 15, 2012, the Act for Partial Revision of the National Diet Library Act was enacted⁽⁷⁾, enabling the NDL to collect online materials (digital books and journals provided on the Internet etc.) In accordance with the amendment, the NDL will start collecting online materials provided free of charge and not protected by DRM (Digital Rights Management) from July 2013 as the first step.

To ensure solid operation of the above-mentioned projects, the NDL drew up “Mission and Objectives 2012-2016”⁽⁸⁾ and “Strategic Objectives” in FY2012. I would like to introduce several items related to strategies for digital collections from those macro strategic documents.

(1) “National Diet Library Electronic Library Concept”:

<http://www.ndl.go.jp/en/aboutus/ndlelc-eng.pdf> (English) (accessed 2013-02-04)

(2) “National Diet Library Digital Library Medium Term Plan for 2004”:

http://www.ndl.go.jp/en/aboutus/elib_plan2004.html (English) (accessed 2013-02-04)

(3) Online Gallery:

<http://www.ndl.go.jp/en/gallery/index.html> (English) (accessed 2013-02-04)

(4) Digital Library from the Meiji Era:

http://kindai.ndl.go.jp/?__lang=en (English) (accessed 2013-07-19)

(5) See the following page as for digitization of the NDL holdings:

<http://www.ndl.go.jp/jp/aboutus/digitization.html> (Japanese) (accessed 2013-02-04)

(6) Web Archiving Project:

http://warp.da.ndl.go.jp/info/WARP_en.html (English) (accessed 2013-02-04)

(7) See the following pages as for the partial revision of the National Diet Library Act in June 2012:

*http://www.ndl.go.jp/jp/news/fy2012/_icsFiles/afieldfile/2012/06/15/pr120615.pdf
(Japanese) (accessed 2013-02-04)*

http://www.ndl.go.jp/en/publication/ndl_newsletter/186/866.html (English; National Diet Library Newsletter, No. 186, December 2012) (accessed 2013-02-04)

(8) Mission and Objectives 2012-2016:

<http://www.ndl.go.jp/en/aboutus/mission2012.html> (English) (accessed 2013-02-04)

1. Introduction

Digital collection strategies at the National Diet Library (from now on, just NDL) have evolved through the following four stages.

First, the starting point of the digitization of the NDL’s collection dates back to June

1998, when the NDL had completed digitization of rare books and old materials and started to provide them through the “Digital Exhibition” in the NDL website. Around the same time, in May 1998, the NDL formulated the “National Diet Library Electronic Library Concept (which I will call the “Concept”)” which represents the desired form of the electronic library to be created by the NDL. The Concept identified three categories of digital resources to be provided by the NDL: offline publications such as CD-ROMs and DVDs, digitized printed materials and Internet information resources.

Regarding the physical format electronic publications, the NDL Act was amended in 2000 to allow the NDL to start acquiring these publications by legal deposit. In the same year, the NDL crystallized the Concept and drew up the “Basic Implementation Plan for Electronic Library Services (Basic Plan).” Digitization of books and the experimental project of archiving websites with the permission of their owners were started.

Second, in 2004, the NDL formulated the “National Diet Library Digital Library Medium Term Plan for 2004,” replacing the Basic Plan. This new plan argued that building the digital archives will be based on the two following pillars; digitization of the NDL’s collections and extensive acquisition of online information resources.

Third, the NDL made a great leap forward in its digitizing project when a large budget for digitization was granted in FY 2009. The NDL Act (amended 2009) enabled the NDL to collect Japanese Internet resources provided by the public sector such as national and local government agencies without permission. The NDL Act (amended 2012) established new scheme to legally collect online materials (digital books and periodicals provided on the Internet, etc). In addition, the NDL started to build the Great East Japan Earthquake Archive to collect and preserve records and information related to the powerful earthquake that struck Japan on March 11, 2011, and to provide it, together with other institutions’ contents, through integrated search.

The fourth stage started in 2012, when the NDL formulated a five-year plan “Mission and Objectives 2012-2016 (Vision 2012).” Now the staff members of the NDL are tackling several issues in accordance with the “Strategic Goals” which set out more practical targets under Vision 2012.

This paper shows the NDL’s past and current strategies for digital collection building by categories.

2. Digitization of the NDL's collection

As one of the two pillars of the NDL's digital collection, digitization of its own collection began with the Rare Books and Old Materials. The outcome was provided in the "Rare Books Image Database" on the Internet in March 2000. Next, the NDL focused on digitizing the books published after the Meiji Restoration in 1868, which have been available in the "Digital Library from the Meiji Era" on the Internet since October 2002.

< Top page of the "Digital Library from the Meiji Era" >

The screenshot shows the top page of the "Digital Library from the Meiji Era" website. At the top, there is a navigation bar with a language dropdown menu set to "English", and links for "FAQ" and "Help". The main header features the title "近代デジタルライブラリー" (Digital Library from the Meiji Era) in large Japanese characters, with the English translation below it, and the logo of the National Diet Library (国立国会図書館). Below the header, a blue search bar prompts users to "Enter title, author or keywords (in Japanese)." with a "Search" button. To the right of the search bar are two buttons: "Theme Search" and "Advanced Search". The page is divided into several sections: "Featured Items" with a sub-header "Items from this database are introduced with commentaries." and a link to "Featured Items (Japanese Only)"; "News" with a date "2013-05-30" and a text snippet about an English interface; and "Other Digitized Contents" which lists various collections: "Rare Books and Old Materials", "Historical Recordings Collection", "Handwritten Newspaper Ishinomaki Hibi Shimbum", "Official Gazettes", "Doctoral Dissertations", "Constitutional Materials", and "Materials on the Allied Occupation of Japan". Each collection is accompanied by a small representative image.





However, not all had been smooth sailing; the digitization had to proceed on a meager budget which only allowed an annual increase of the digitized contents provided to the public of only by some ten thousand. The ratio of digitized contents to all the books held by the NDL reached only 2% by 2008 (around 150,000 items as of the end of FY2008).

A major turning point came in 2009 when the supplementary budget for the purpose of stimulating the economy allocated 12.7 billion yen to the NDL for digitization.

In addition, in 2009, the Copyright Act was amended to make it possible for the NDL, the sole deposit library in Japan, to digitize its holdings regardless of their condition for the purpose of long-term-preservation and to provide the digital images within the NDL's premises, without permission of the rights holders. Before the amendment, the Copyright Act allowed libraries, including the NDL, to digitize only severely deteriorated items.

As a result, by the end of FY2012, the NDL had successfully digitized most of the Japanese books published before 1968, often-used Japanese periodicals published before 2000 and doctoral dissertations accessioned during FY1991-2000. As shown in Table 1 below, the amount of the production made up one-fourths of the NDL's collection including Rare Books and Old Materials, Japanese books, Japanese periodicals and doctoral dissertations.

< Table 1: Current state of digitization of NDL's collection (as of the end of FY2012) >

Type of materials	Number of collection (A)	Digitized materials (B)	Undigitized materials (A-B)	Operation rate (B/A)
 Rare Books and Old Materials	290,000	90,000	200,000	1/4
 Japanese Books	4,360,000	900,000	3,460,000	1/5
 Japanese Periodicals	4,610,000	1,120,000	3,490,000	1/4
 Doctoral Dissertations	390,000	140,000	250,000	1/3
Total	9,650,000	2,250,000	7,400,000	1/4

As of May 2013, the total number of digitized contents shown in Table 1 includes 470,000 items provided on the Internet, of which 360,000 were released to the public after copyright clearance such as checking the year of authors' death to confirm expiration of copyright or discovering copyright owners to obtain necessary permission. Among those available on the Internet are 86,000 items provided based on the compulsory copyright license from the Commissioner for Cultural Affairs (the chief of the Cultural Affairs Agency), the scheme defined in the Copyright Act to enable the use of copyrighted works by depositing money under the ruling of the Commissioner for Cultural Affairs (In Japan, the Cultural Affairs Agency is in charge of copyright issues) provided that the whereabouts of copyright holders and their inheritors are not found.

< Display example of the NDL digitized contents >



The screenshot displays the NDL Digitized Contents website interface. At the top, there is a navigation bar with the title "NDL Digitized Contents", a language dropdown set to "English", and links for "FAQ" and "Help". Below this is a search bar with a "Search" button and a checkbox for "Include 'Onsite only'". There are also buttons for "Collection Search", "Theme Search", and "Advanced Search". A banner below the search bar states: "This site enables a user to search and view a variety of resources, collected and stored by the National Diet Library of Japan."

The main content area is titled "Rare Books and Old Materials". It features a description: "Out of the 300,000 items stored in the Rare Books and Old Materials Room of the NDL, about 92,000 are available as digital images." Below this, it lists categories: "These items are: Japanese items created before the end of the Edo period (1868), Chinese items created before the end of the Qing dynasty, including *Nishiki-e* (colored woodblock prints), *E-zu* (old maps and plans), and nationally-designated important cultural properties. About 20,000 items are available only within the facilities of the NDL, because of relatively recent or unknown publication dates. A List of Items with Analyses and Transcriptions (Japanese Only)".

In the center, there is a thumbnail image of a world map. To the right of the image, the title "地球全国【チキウゼンズ】 司馬江漢岐【シバコウカン】" is displayed, followed by the English title "Chikyu zenzu" and the description: "Copied and etched by Shiba Kokan. S.I.: s.n., c.1794." Below this, a detailed description reads: "A stereographic projection map of the world by Shiba Kokan. Shiba Kokan (1747-1818) was the first person in Japan to use copperplate etching to print a world map. The shape of the map clearly shows that the earth is spherical. He was a painter, but was interested in Dutch studies due to the influence of Hiraga Gennai. He was the first to make copperplate etching".

At the bottom of the page, there is a navigation menu with categories: "Rare Books and Old Materials", "Books", "Periodicals", "Newspapers", "Historical Recordings", "Official Gazettes", "Doctoral Dissertations", "Constitutional Materials", "Materials on the Allied Occupation of Japan", "Prange Collection", and "Internet Resources".

There were also discussions as how to make the most of NDL's digitized contents which consumed enormous amounts of the national budget. In June 2012, the Copyright Act was amended to enable transmitting digital images of out-of-print items to public and university libraries in Japan. By this transmission, users can access and copy them in a public or university library close to their residence. Aiming for its launch in January 2014, the NDL is presently working hard on preparations such as detailing the procedures to accept a claim to exclude a still-in-print publication from the transmitted contents as well as developing the system for transmission.

3. Web archiving

To realize the collection of online information resources as the other pillar of NDL's digital collection, the Web Archiving Project (WARP) was launched to collect websites in November 2002.

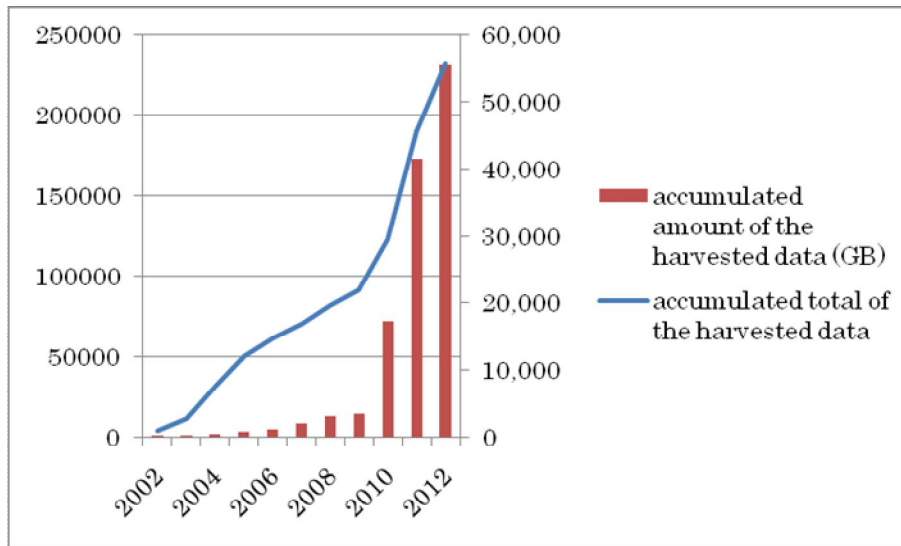
The project was initially piloted experimentally based on permission of copyright holders such as website owners. With regard to the public sector's websites, the NDL worked on law amendment to collect and preserve them without permission; and the NDL Act, amended in 2009 and put into force in April 2010, enables the NDL to do just that.

Websites are harvested with a crawler named "Heritrix" developed by the International

Internet Preservation Consortium (IIPC) of which the NDL is a member. In addition, a part of core works extracted from the website such as online publications, white papers, yearbooks, etc. are assigned metadata and stored. Collected websites are basically available within the NDL's premises, and about 70% of them for which the NDL obtained permission are provided through the Internet as well.

The number of collected websites is 7,466 items and that of harvesting is 55,749 times in total. The data amount was 231TB in the end of March 2013. The transition of data amount collected during FY 2002-2012 is shown in Table 2.

< Table 2: Transition of the accumulated amount of the harvested data >



4. Collection of digital books and periodicals

Though the NDL Act (amended 2009) made possible the collection of websites of the public sector without permission, collection of information resources released on the Internet by private publishers, companies, individuals, private universities and research institutes still requires permission from those owners. Among rapidly-spread digital books and periodicals, the collection, preservation and provision of those issued by the private sector needed another law amendment. The NDL pushed for the legal amendment, and thus the NDL Act saw another amendment in June 2012 (the amendment act comes into force in July 2013) to oblige publishers of digital books and periodicals to deposit their publications with the NDL by applying for the NDL's harvesting with a crawler, uploading digital data manually to the system or sending digital data recorded on a DVD by mail.

Since the NDL was still in negotiation with publishers and rights holders for the collection, preservation and provision of paid digital books and periodicals, the amendment act exempts paid publications and DRM(Digital Rights Management System)-protected publications for some time, and the target for legal deposit is any of the following among free and non-DRM publications published (released) on the Internet and other transmission media, by the private sector: publications assigned with the prescribed codes (ISBN, ISSN, DOI (Digital Object Identifier)) or recorded in the prescribed file formats (PDF, EPUB, DAISY (Digital Accessible Information System)).

In addition, the Academic Degree Regulations (Ordinance of the Ministry of Education, Culture, Sports, Science and Technology) were amended in April 2013 so that a doctoral dissertation must, as a rule, be published through the Internet. The NDL has, so far, relied on donation, not deposit, from universities etc. for collection of doctoral dissertations in paper form, throughout Japan. From FY 2013, it collects and preserves doctoral dissertations as digital contents in the same manner as digital books and periodicals.

5. Great East Japan Earthquake Archive

In order to pass down the records and lessons of the Great East Japan Earthquake of March 11, 2011 to posterity and to serve restoration and reconstruction of the afflicted area as well as future disaster prevention and reduction, the NDL developed a plan to establish the Great East Japan Earthquake Archive right after the event. At the outset, NDL strengthened acquisition and preservation of websites from affected local authorities. The focus on record preservation was also endorsed in the *Reconstruction Vision* and the *Basic Policy for the Post-Great East Japan Earthquake Reconstruction* compiled by the government, and the NDL is now driving for the construction of the fully-open archives by collecting all sorts of records from public authorities, other organizations and individuals in cooperation with government agencies.

In March 2013, the NDL launched the NDL Great East Japan Earthquake Archive "HINAGIKU," a portal co-developed with the Ministry of Internal Affairs and Communications that enables integrated search and access of sound and videos, pictures, websites, etc. uploaded by various institutions (initially broadcasters, newspaper publishers, research institutes and libraries).

6. Digital Data of Audio and Braille Books for Visually-Impaired, etc.

Due to the Copyright Amendment Act of 2009, libraries and other relevant institutions providing services for the visually impaired, etc. are empowered to produce digital

audio books and transmit them online. In July 2011, the NDL formulated a plan Implementing Services for the Visually-Impaired and set out to launch the transmission of digital data for such persons (such as DAISY data of audio books) produced by the NDL to public libraries and individuals by March 2014.

7. Collecting sound recordings, audio-visual materials, etc.

Under a four-year plan, the NDL has collected about 48,000 digital sound recordings and has made them open to the public in stages since May 2011. The digitization was done by the Historical Records Archive Promotion Conference (HiRAC), composed of associations and organizations related to music and records, from 78-rpm records and original metal discs produced in Japan from 1900 to around 1950. Recordings no longer under copyright are available on the Internet while others can be listened to only within the NDL's premises. Since January 2012, the latter category of copyright-protected recordings are also available, at the NDL's expense for the copyright usage fee, through the NDL's transmission service in domestic public libraries and other appropriate institutions (115 as of May 2013).

8. Further Challenges

The digital collection related challenges to be tackled by the NDL in the next five years are upheld in Vision 2012 and subordinate Strategic Goals.

Regarding the digitization of its collection, the Basic Plan for Digitization was developed in 2013. It argues that digitization should focus on the selection of pre-1980s, i.e. pre-DTP (desktop publishing) software, books. The appropriation of the necessary budget for digitization remains a challenge.

Regarding the archiving of websites, the NDL is a more active kind of national library which, with permission granted, publicly displays the archived pages on the Internet. The challenge of the NDL is to further improve the permission rate and to make more websites available for users.

As for the acquisition and preservation of digital books and periodicals released on the Internet, the NDL's task is to smoothly roll out the acquisition scheme of free ones and doctoral dissertations in 2013 and to push for legislation to include paid digital books and periodicals in the deposit scheme. Such publications are usually DRM protected, so there are technical barriers to acquisition, preservation and provision to be surmounted. Japan has traditionally awarded compensation to the publishers for complying with the

legal deposit, but whether the NDL will retain a similar framework for digital books and periodicals remains to be decided in consultation with the stakeholders.

In regard to the Great East Japan Earthquake Archive, the NDL will continue the comprehensive acquisition and preservation of records including the pre-disaster period, reconstruction process and other past disasters such as earthquakes, tsunamis and atomic power plant accidents. In order to attain nationwide collaboration, the NDL will encourage and support other organizations' similar efforts and integrate search-related information of relevant records such as location information. The scope of the project is not just national but is intended to provide useful information relating to disaster prevention and reconstruction worldwide.

As for the digital data of audio and braille books for the visually impaired, etc., the NDL needs to collect similar data produced by public libraries and transmit them to other libraries and visually-impaired, etc.. Transmitting text data to those individuals in the DAISY format, already legally permitted, remains to be done.

With respect to audiovisual (AV) materials, acquisition through digitization is already on the agenda but some of the historically important AV materials are under the threat of scattering and loss so its implementation needs to be fast-tracked. Also exhaustive acquisition and preservation of broadcast programs as cultural assets are challenges that relevant organizations, including the NDL, need to tackle together.

Aside from the above matters, there are some issues concerning the continued service of digitized materials and deposited born-digital materials: outdated of formats, deterioration of preservation media (discs of DVD and Blu-ray), and the sheer amount and diversity of archived web files. These issues need to be addressed.

Thus far, I have shown the multi-faceted challenges the NDL faces in the matter of digital collection. While acquiring and preserving digital and other cultural resources and passing them onto posterity, the NDL will continue to make efforts to enable everyone, regardless of location and handicap, to use such cultural resources, by resolving all these issues.