Archiving Cultural and Community Memories in a Networked Information Society: A Japanese Perspective

Shigeo Sugimoto
Faculty of Library, Information and Media Science
University of Tsukuba, Japan
sugimoto@slis.tsukuba.ac.jp

ABSTRACT
Many digital collections of cultural heritage resources, commonly referred to as digital archives, have been developed in Japan. This paper covers the multiple facets of digital archiving for a comprehensive understanding of Japanese activities and issues learnt since the 1990s, starting from the development of digital archives and related activities. It then discusses some issues for sustainable development of digital archives: social, usability, technological and longevity. As building and maintaining digital archives is expensive, it is important to demonstrate their value to ensure their sustainability. Continuous effort must be made to build really usable digital archives, and to develop appropriate measures to evaluate usability. Such effort must include human resource development and community networking.

INTRODUCTION
Many projects to develop digital collections of cultural heritage have been carried out worldwide since the explosive growth of the Internet and the Web in the mid-1990s. The author has discussed digital archives and metadata in the context of community memory and cultural resources in previous articles (Sugimoto, 2014; Sugimoto, Mitsuharu, Tetsuya, & Tsunagu, 2015). This paper looks back at specifically Japanese projects involving cultural heritage and community memories and discusses some issues from the author’s perspective—with the intention to provide an overview of Japanese digital library activities more comprehensively.

Large digital library projects were carried out in the 1990s. For example, American Memory and Making of America by large research libraries, and the Digital Library Initiatives by the U.S. National Science Foundation and other funding bodies are the most famous projects. American Memory and Making of America developed large collections of historical resources. The Digital Library Initiatives have funded many research projects of different types—information and communication technology (ICT) centric projects, collection-centric projects, and user-centric projects. In Japan, several major digitization projects were started by the National Diet Library (NDL) and major memory institutions. IPA/JIPDEC conducted a comprehensive digital library project in collaboration with the National Diet Library (Mukaiyama, 2000). The National Diet Library started the Kindai Digital Library, a huge digital collection of books published in Japan after the revolution in the 19th century. The Japan Center for Asian Historical Records at the National Archives of Japan started building a digital collection of Japanese governmental records from the
revolution in the 19th century to the end of World War II. The museum community has also had major works for building digital collections. Cultural Heritage Online is a collaborative project by the nation-wide museum community under the Agency for Cultural Affairs of the Japanese government.

In Japan, digital archive is commonly used as a term meaning a large collection of digital resources, primarily in the cultural domain. Various types of digital collections have been built as a digital archive—digitized books, digital images of museum resources, 3D images of heritage sites, etc. Libraries, museums, archives and related institutions, referred to as memory institutions, are the key players in building digital archives.

There are also digital archives for non-cultural resources. Building digital archives of disaster records is a crucial activity as it not only keeps records of disasters, but also keeps community memories for the future. Many digital archives were created after the Great East Japan Earthquake and Tsunami on March 11, 2011. Many digital photographs and videos were taken at the disaster by people who witnessed the disaster, and many photographs and other data were created after the disaster to record the disaster itself and recovery from it. There are many non-digital materials which should be archived in addition to born digital resources. Thus, these archives collect both born digital materials as well as digitized materials. The National Diet Library is providing a portal for the Great East Japan Earthquake and Tsunami archives, which is named Hinagiku.

Digital archives refer to a collection of digital objects of cultural content which may be born digital or converted into digital, for example photographs, videos, audio records and documents. On the other hand, many databases of cultural resources that have been or are being built provide only metadata of the cultural resources with no primary digital objects. These are not strictly digital archives from the viewpoint of primary digital objects, but this paper also mentions them as the border between digital archives and databases is not crisp, and metadata about cultural resources are a key component of digital archives. This paper first looks back at these projects and related activities in Japan, and then discusses issues from the author’s perspective.

DIGITAL ARCHIVE DEVELOPMENTS AND ACTIVITIES—AN OVERVIEW OF JAPANESE ACTIVITIES

Overview From an Institutional Viewpoint

Many digital libraries and archive projects were launched in the 1990s. Some projects have already ended while some are ongoing. In addition to memory institutions, there are other stakeholders in the development of digital archives, such as government sectors, universities, industries and non-profit organizations. This section provides an overview of the digital archive developments and related activities in Japan from the viewpoint of institutions responsible for building and managing them:

1. The National Diet Library (NDL), a legal deposit library in Japan, has several large scale digital collections of books, manuscripts, and records of the parliament:
   - The Kindai Digital Library at the National Diet Library provides about 350,000 books open to the public via the Internet. NDL is planning to integrate the Kindai Digital

1 Digital Library from the Meiji Era, http://kindai.ndl.go.jp/?lang=en
Library with other digital cultural heritage collections\(^2\).

- NDL has been providing records of the Japanese parliament since 1890. The collection includes records of both the imperial parliament before 1947 and the current parliament after 1947\(^3\).
- As Web pages are important resources for information on the community and society today, NDL is archiving Web resources, mainly those published by government and public sectors\(^4\).
- NDL is a nation-wide aggregator of digital information for libraries and library users (e.g., the portal to the Great East Japan Earthquake and Tsunami archives, Hinagiku\(^5\)).
- NDL provides their authority resources as a Linked Open Data (LOD) dataset (e.g., name authorities and NDL subject headings\(^6\)).

2. *The National Archives of Japan* (NAJ) has digital archives of government records and documents. Their digital archives include both digitized and born digital materials. Digital preservation of government records is a key responsibility of the NAJ\(^7\). The Japan Center for Asian Historical Records (JACAR) at the NAJ has a large digital collection of government records and documents spanning 1867 to 1945 which are collected mainly from the NAJ, Archives of Ministry of Foreign Affairs, and Archives of the Former Japanese Army and Navy\(^8\). The development of the JACAR database started in the late 1990s, and it has about two million catalog records and 30 million page images. This database is a very important resource for history studies of Japan and East Asia from the 19th to 20th century.

3. *The national museums and the Agency for Cultural Affairs* are the main players for building digital collections of cultural heritage held in museums. Cultural Heritage Online collects digitized resources not only from national museums but also regional museums\(^9\), and is a collaborative project to build a portal for cultural resources in Japan. National museums collaborate to build e-Museum, a collection of national treasures and important cultural prosperity of Japan\(^10\).

4. *University libraries* have worked on digitization of cultural heritage supported by the Ministry of Education, Culture, Sports, Science and Technology since the late 1990s. For example, the Kyoto University Library provides a digital collection of cultural heritage, though the main focus of university libraries has moved to institutional repositories. Japanese Institutional Repositories Online (JAIRO) is the collaborative activity of many

---


\(^5\) NDL Great East Japan Earthquake Archive, http://kn.ndl.go.jp/

\(^6\) Web NDL Authorities, http://id.ndl.go.jp/auth/ndla

\(^7\) National Archives of Japan Digital Archive, https://www.digital.archives.go.jp/index_e.html

\(^8\) Japan Center for Asian Historical Records, http://www.jacar.go.jp/english/index.html


university libraries\textsuperscript{11}.

5. \textit{Research institutions, museums and archives in academia}: Research institutions in the humanities, social sciences and area studies have large collections of digital resources. The National Institutes for the Humanities, which is composed of six major research institutes and museums, is developing a system to promote sharing of research resources accumulated by their research institutes\textsuperscript{12}. The Historiographical Institute at the University of Tokyo has a rich set of data of Japanese historical and cultural resources\textsuperscript{13}. The Center for Social Research and Data Archives at the University of Tokyo has a large set of data archives for social science research\textsuperscript{14}. The Center for Integrated Area Studies at Kyoto University provides many databases created for area studies in collaboration with research institutes in Kyoto University, that are rich in resources on area studies, particularly in Asia and Africa\textsuperscript{15}. Digital humanities is a new scholarly domain closely connected to digital archives in both user and developer communities. The Art Research Center at the Ritsumeikan University provides many databases of Japanese cultural resources\textsuperscript{16}. The Digital Humanities project for Japanese culture is one of the main activities at the Art Research Center.

6. \textit{Regional libraries}: Prefectural public libraries play an important role in collecting digital resources of regional cultural heritage. The Tokyo Metropolitan Library provides Tokyo Archive, a digital collection of historical and cultural resources of Tokyo\textsuperscript{17}. The Akita Prefectural Library has a digital collection of the regional cultural heritage, which includes non-tangible heritage such as folktales spoken in a local accent\textsuperscript{18}. The Akita Prefectural Library has been working as a regional hub for memory institutions in Akita prefecture. The Okayama Prefectural Library has been running the Digital Okayama Encyclopedia as a set of services provided on the Internet, including digital collections of cultural heritage as well as resources created by regional communities\textsuperscript{19}. There are many projects developing digital resources, though in general, regional activities are not large because of limited financial and human resources.

7. \textit{Industries and non-profit organizations (NPOs)}: In the industrial sector, Toppan Printing Co. Ltd. (Toppan) and Dai Nippon Printing Co. Ltd. have created high quality digital cultural resources using advanced information technologies, such as virtual reality, and collaborating with stakeholders. Participation of the commercial sector in digital archives is important not only for the development of new technology, but also for creating new

\textsuperscript{12} Resource Sharing, National Institutes for the Humanities, http://www-e.nihu.jp/e/sougou/kyoyuka/index.html
\textsuperscript{13} Historiographical Institute Database, University of Tokyo, http://wwwap.hi.u-tokyo.ac.jp/ships/db-e.html
\textsuperscript{14} Center for Social Research and Data Archives, University of Tokyo, http://csrda.iss.u-tokyo.ac.jp/en/
\textsuperscript{15} Center for Integrated Area Studies, Kyoto University, http://www.cias.kyoto-u.ac.jp/
\textsuperscript{16} Art Research Center, Ritsumeikan University, http://www.arc.ritsumei.ac.jp/ (in Japanese)
\textsuperscript{17} Tokyo Archive, http://archive.library.metro.tokyo.jp/da/top (in Japanese)
business models which are crucial for covering digitization and maintenance costs, a vital issue for the sustainability of digital archives. There are various NPO-based activities: creating digital archives of regional content to keep the memories of the community for the future, promoting high-quality digital archives of cultural content, and supporting collaboration for building digital archives at small and medium memory institutions.

8. **Governmental sectors:** Since the 1990s, the Japanese government has been promoting ICT for the building of better information infrastructures. In the government sector, some ministries have been working to promote the creation and use of digital archives in the cultural domain from several different viewpoints such as:

- Digital archives as a crucial knowledge infrastructure to support development of regional economy, education and communities;
- Digital archives as a crucial resource to promote international recognition of Japanese culture, and create a market of Japanese cultural content for publishing and tourism.

In these activities, memory institutions and other stakeholders are involved as main players in building digital archives and providing digital resources for the public for education, research, entertainment, and so forth. In 2011, the Ministry of Internal Affairs and Communication hosted a study group for promoting digital archives in the networked information society. It mainly discussed issues to promote digital archives of cultural resources at the national and regional institutions. Discussion on promoting development and enhancing the usability of digital archives is continuing as a part of the intellectual asset strategy planning under the Cabinet Office.

**Non-Conventional Cultural Heritage Archives**

This section covers digital archive projects and related activities which are not in the conventional cultural heritage domain—pop-culture, videos and films, and disaster records.

**Pop-Culture Archives—Manga, Games, Animations, etc.**

Pop-culture resources such as manga, video games and animations are crucial resources of Japanese culture because of their popularity and market size. These databases are collections of bibliographic data with no primary resources included, due to intellectual property issues and the proprietary nature of the content. Among the pop-culture resources, manga has had the longest history of publishing. However, manga books and magazines were not collected by libraries as part of their mainstream collection because they were not recognized as regular library material. The Kyoto International Manga Museum provides a rich set of bibliographic data of manga primarily created for their holdings. The Agency for Cultural Affairs has developed a Media Arts Database which covers manga, video games, animations and media arts. In the pop-culture domain, there are many resources produced by amateur creators which are exchanged and traded at markets, some of which are very large events, but they are poorly collected and recorded. There are also many fan-created information resources about primary pop-culture resources, an example being Web pages describing manga and games. These fan-created information resources are useful for resource discovery in the pop-culture domain.

---

Video and Film Archives

The National Film Center, founded in 1952\textsuperscript{21}, at the National Museum of Modern Art is an archive for films. NHK (Japan Broadcasting Corporation) has an archive of their video and audio content\textsuperscript{22} in digital form, including 800,000 programs and 6,000,000 news items as of March 2015. Their content is used in their video-on-demand service, with many video resources made accessible to the public for free. The National Film Center also provides a collection of digital resources created from their non-film holdings, which falls under cultural heritage covering the history and culture of Japanese movies.

Disaster Archives

The disaster caused by the great earthquake and tsunami on March 11, 2011, was recorded by many people using digital devices such as mobile phones and digital cameras. The public sector, research sector and non-profit organizations have created many records after the disaster in the forms of photographs, videos, sound and voice recordings, and documents. Mass media platforms such as broadcasting stations and newspapers produced massive amounts of material as well. The Ministry of Internal Affairs and Communication, the National Diet Library, International Research Institute of Disaster Science at Tohoku University\textsuperscript{23}, Harvard University’s Reischauer Institute of Japanese Studies\textsuperscript{24}, and other partner organizations collaborated to build digital archives of the disaster. The National Diet Library, the hub of the collaboration, opened the portal for the digital archives named Hinagiku in March 2013. The primary goal of disaster archives is to record what happened in and after disasters, but they can still be a part of archives of regional historical records.

LESSONS LEARNT IN DIGITAL ARCHIVE DEVELOPMENT

The primary role of digital archives is “preservation and access,” which means that on one hand a digital archive preserves cultural heritage in a digital format and on the other provides access to digitized cultural heritage via the Internet. This is generally agreed on, but in reality there are many arguments against this as a general goal. The issues and questions below were raised in the 1990s and early 2000s with regard to the development of digital archives:

- **Cost recovery issues**: How can we recover costs to create and maintain digital resources?
- **Legal issues**: How can we manage legal issues such as intellectual property, privacy and accessibility?
- **Longevity issues**: How can we preserve digital objects properly? How can we inherit digital archives when their maintenance organization ceases to exist?
- **Usability issues**: Are digital archives well used? How can we enhance usability of digital archives?
- **User issues**: Who uses the digital archives—scholars and researchers, educators, students, or the general public? For what purposes do the users use them?

None of these questions is easy to answer. The author considers these factors, on top of

\textsuperscript{21} National Film Center, http://www.momat.go.jp/english/fc/

\textsuperscript{22} NHK Archives, http://www.nhk.or.jp/archives/en/


\textsuperscript{24} Digital Archive of Japan’s 2011 Disasters, http://www.jdarchive.org/en/home
limited financial and human resources, to be the reasons why small and medium memory institutions are reluctant to build digital archives. The following discuss issues for the development of digital archives which the author has learnt in the course of his research activities:

1. **Digital archives** are crucial to keeping cultural heritage and community memories safe for the future. The author has participated in several discussion groups on digital archives hosted by the public sector. One study group was hosted by the Ministry of Internal Affairs and Communication to discuss promotion of digital archives at memory institutions as a knowledge infrastructure. The study group started with the underlying agreement that memory institutions, especially those founded by regional governments, are facing common barriers such as the lack of human and financial resources, and that longevity of digital resources is also a common concern for memory institutions when investing their funds. As the group was launched in February 2011, the earthquake and tsunami disaster of March 11, 2011, significantly affected the discussions of this study group. Many memory institutions in the region were seriously damaged. Many rescue activities were carried out after the disaster but it was not possible to perfectly recover the damaged resources. The situation in the area affected by the nuclear plant accident was much worse. Because people had to move out of the affected area, any rescue activity immediately after the disaster was not possible, and the continuity of the regional community was threatened. The study group agreed that promotion of digital archives is a key strategy to keeping community memories and cultural heritage safe from such disasters and changes of social environments. The author learnt from this discussion that physical things are easily lost in disasters and digital copies play a big part in keeping community memories and cultural heritage safe for the future.

2. **Human networks** are important for collaborative development. The Ministry of Internal Affairs and Communication study group recommended forming a human network to support development of digital archives at small and medium size memory institutions which lack human resources for developing and maintaining digital archives. The author was involved in an activity named Digital Archive Network (DAN) that shared knowledge and information about digital archives of cultural heritage and community memories. He learnt from DAN activities that regional and cross-regional human networks are useful in both building digital archives and enhancing their usage and usability.

3. **Metadata sharing on the Internet and Web** is obviously important for building digital archives. Europeana, the Digital Public Library of America, and Hinagiku, all of which provide a rich set of digital resources, aggregate metadata provided from participating institutions and organizations. Metadata interoperability across archives is crucial because of the need to aggregate heterogeneous metadata from diverse participating institutions—libraries, archives and museums—in diverse communities. Semantic linking across archives is a challenging but necessary task to connect digital archives, and Linked Data (or Linked Open Data) technologies are commonly used in these large digital archive portals. The Ministry of Internal Affairs and Communication study group recommended using the International Standard Identifier for Libraries and Related Organization (ISIL) as a base identifier to specify the domains of the resources, and to define their Uniform Resource Identifier (URI) identification scheme. Longevity of metadata and metadata interoperability over time, that is, temporal interoperability of metadata, are crucial issues
for digital archives. In the Internet environment, metadata has to be both human and machine interpretable. Machine interpretability of metadata over time is crucial but not well-studied. In particular, semantic linking in the Linked Data environment depends heavily on the longevity of the URI, and the stability of the meaning of the metadata terms.

**DISCUSSION**

This section discusses the lessons learnt by the author from his research activities as well as from participation in committees and groups organized by public sector organizations. The discussion is organized into four aspects—social, usability, technological and longevity.

**Social Aspects—Policy and Legal Issues**

Digital archives can be categorized into several different policy categories, including open, in-house or dark archive, public or proprietary content, and crawler-accessible or human access only archive. Legal issues are closely connected to policy issues in the design and management of digital archives.

As there are so many policy and legal issues for digital archives, it is impossible to find a panacea for all the issues. This section introduces only a few issues that are often discussed—intellectual property, privacy and ethical issues. Intellectual property is always an important issue when developing digital archives. Rights management for orphan works is a central issue for digital archive development in any cultural domain (e.g., books, pictures, music), so we need a reasonable solution to make the approval process to use orphan works more efficient.

For example, disaster archives contain many photographs of unidentified individuals. The images of individuals on a photograph are shaded because the cost to identify the individuals and get permission from them is high. On the other hand, the cost to shade off individuals may be a barrier to opening digital archives to the public. As another example, there is a concern among scholars of historical map databases that historical maps may include location names which may be improperly used for social discrimination. Even if a historical map is a cultural resource, the information on the map may cause problems for people living and/or born in the area brought about by improper usage. These sorts of problems are common when handling historical resources.

Secondary use of archived resources and metadata provided by digital archives is important to enhance the usability and usage of the digital archives. There is discussion on the secondary use of digital archive resources but, in general, conditions for secondary use of digital archive resources vary case-by-case. Therefore, specifications of the conditions of secondary use are important information for users of digital archives. Condition descriptions of secondary use of metadata about archived resources are also important.

**Usability Aspects**

Digital archives should be highly usable for different types of users, i.e. scholars, domain specialists, educators, students and the general public. Search and browse functions are crucial components in the design of digital archives to enhance usability. Application Programming Interfaces (API) are also crucial to open digital archives for third party developers and user created information services. It is important that these digital archives are highly usable in order to continuously get funds to keep the digital archives alive. However,
there is no easy solution to this issue because of the variety of users, as well as the plethora of archived resources. Most of the cultural heritage archives in Japan, as well as other countries, have been built by national and public institutions. Under weak economic circumstances, wide recognition of the importance of digital archives is a crucial factor in keeping digital archives sustainable.

**Technological Aspects**

Advanced information technologies are crucial to enrich digital archives and to enhance their usability. For example, advanced 3D sensing and virtual reality technologies are crucial to create high-quality cultural heritage digital archives. Character recognition of old Japanese documents handwritten using a brush, and motion-capture of Japanese dancers are important not only for digitization purposes but also for analysis.

Metadata technology is key for digital archives. Linked Open Data (LOD) is an important metadata issue in the current Web-based information environment. The National Diet Library has been using LOD technologies in their services such as Hinagiku and National Diet Library authorities. Hinagiku defines a metadata element set named NDL-KN. It collects metadata from participating digital archives in RDF form using OAI-PMH. Thus, LOD technologies are well recognized as key technologies to make digital archives open on the Internet. Though the number of LOD datasets is growing, they are still limited. Furthermore, ontology resources (e.g., the National Diet Library authorities and DBpedia) which are essential to semantically connect resources are limited.

Maintenance of metadata sometimes requires semantic error correction. Morishima et al. (2014) has applied a crowdsourcing technology to find and correct errors in the bibliographic database of the National Diet Library. Li, Nagamori and Sugimoto (2015) proposed a provenance description model to keep track of metadata schema change in the Web environment where metadata schemas are shared as an RDF-based metadata instance. It is known that sharing vocabularies is key to enhancing data sharing on the Internet. The Infrastructure for Multi-layer Interoperability (IMI) project is a governmental project to build an infrastructure for vocabulary sharing. A project called LODAC applied LOD technologies to connect museum resources. The Ministry of Internal Affairs and Communication study group had recommended building a system to share metadata schema information. MetaBridge, which is a metadata schema registry based on RDF, is developed following this recommendation.

As maintaining large digital storage is too expensive for small and medium-sized memory institutions, cloud storage seems a reasonable solution. Bit-stream level trusted archive storage may be found, but application-level archive storage which handles various types of file formats is still not highly developed.

**Longevity Aspects**

The longevity of archived resources is a fundamental issue for digital archives because the life-time of digital technologies is rather short when compared to the expected longevity of information resources. The National Archives of Japan have defined their metadata schema

---


for digital preservation. Stable and widely-used file formats such as PDF, JPEG2000 and TIFF are often used in digital archives for static resources. Preservation of dynamic resources such as offline/online games and computer graphic images is handled by each memory institution. In the case of virtual reality data for cultural heritage, the original data taken by sensors and edited for use is maintained for migration to a new virtual reality technology in the future.

Longevity is affected not only by technological changes but also social and socio-technological environment changes. Maintenance organization of cultural heritage is needed not only to keep the heritage objects safe, but also to maintain metadata of the objects properly for the future. As technologies used to create and maintain digital objects and their metadata change over time, it is crucial to keep provenance information about the technologies in order to track the maintenance history.

CONCLUSION

In the early stage of digital library development in Japan, digital collections of cultural heritage were a central topic but there were questions about their value and roles (i.e., who are the users and how useful are the digital collections for the users). “Preservation and access” was a general answer to the questions, but the impact was not obvious, and costs and legal issues were a barrier for small and medium institutions. Europeana has significantly influenced discussions on digital archives in Japan, including its approach of non-centralized development of digital collections, aggregation models, metadata aggregation, and adoption of LOD technologies. Discussions in the government and public sector are from the perspective of development of the economy and education.

There are many unresolved issues relating to the future of digital archives. The author considers the most important point to be the need for continuous effort to build really usable digital archives which should include human resource development and community networking. We need appropriate measures to evaluate the usability of digital archives in order to make digital archives sustainable. Broader recognition of the value of digital archives is also essential.

As the social and technological environments are changing so fast, some problems found in the early stages of the development no longer exist, for example strict limitations of file size and bandwidth of the network. On the other hand, new systems and services have appeared in the past two decades—Google, Facebook, Twitter, YouTube, and so on. The lifetime of digital resources, particularly those of cultural heritage, is much longer than the lifetime of new information technologies and services. Social environments and social systems change slowly but steadily. On the other hand, developing a digital archive is a time consuming process and not as publicly visible as the new technologies and services. Continuous effort is needed to keep digital archives growing for the future.

REFERENCES


