Routledge
Taylor & Francis Group

Digital archives and metadata as critical infrastructure to keep community memory safe for the future – lessons from Japanese activities

Shigeo Sugimoto*

Shigeo Sugimoto received BE, ME and PhD degrees from the Department of Information Science, Faculty of Engineering, Kyoto University in 1977, 1979, and 1985, respectively. He joined the University of Library and Information Science (ULIS), Tsukuba, Japan in 1983. After the merger of ULIS with the University of Tsukuba, he has been teaching at the Graduate School of Library, Information and Media Studies, University of Tsukuba. He is a professor at the Faculty of Library, Information and Media Science, University of Tsukuba and currently a director of the Research Centre for Knowledge Communities, University of Tsukuba.

This paper discusses some key issues for digital archives and metadata in a networked information environment to keep our community memory for the future. The paper is based primarily on the experiences and lessons learnt by the author from his research activities on metadata and digital archives. The author participated in a study group on digital archives hosted by the Ministry of Internal Affairs and Communications of the Japanese Government from February 2011 to March 2012. The group discussed the promotion of digital archives in Japan, particularly at memory institutions. The Great East Japan Earthquake of 11 March 2011 caused serious damage in the north-eastern part of Japan, especially the Pacific coastal regions. This disaster significantly affected the discussions of the group. The basic lesson that the author learned is that digital archives built on a robust information environment are essential for keeping our community memory safe for the future. Not only do the primary digital resources need to be properly maintained and preserved for the future but also secondary resources, metadata and meta-metadata. We need to use Linked Open Data technologies to enhance the usability of such digital resources in the archives.

Keywords: digital archives; longevity of digital resources and metadata; meta-metadata; metadata; metadata interoperability

Introduction

A digital archive is a collection of digital resources maintained for long-term use. Many digital archives have been developed and are being used in our networked information society. Digital archives are built in many different places – not only in cultural fields but also in the science, technology, medical and social fields. Memory institutions – museums, libraries and archives – are key players when building and maintaining digital archives and providing them to users in many different communities today and in to the future. Digital archives are recognised as an important part of the knowledge and information infrastructure in our networked information society. They provide users with inexpensive access to valuable knowledge resources, which is crucial for a democratic society.

*Email: sugimoto@slis.tsukuba.ac.jp

This paper aims to discuss a few key issues with respect to digital archives and metadata. It presents some basic issues and models of digital archives and metadata learnt from the author's digital archive research activities. The paper is primarily aimed at identifying and sharing issues with digital archives as a critical infrastructure for the safekeeping of the memory of our communities and societies.

The rest of this paper is organised as follows: section 2 describes the basic concepts of digital archives and discusses some related activities, section 3 presents the author's committee activities in Japan, section 4 discusses metadata issues for digital archives, in particular from the viewpoint of interoperability, section 5 introduces the digital archives of the Great East Japan Earthquake of March 2011, section 6 presents a few studies on metadata and digital archives at the author's laboratory and section 7 concludes this paper.

Digital archives – basic concepts and related works

Digital archive, in this paper, means a collection of digital resources selected, collected, organised and maintained for long-term use. This is a comprehensive definition of digital archive designed to include different types of digital collections, that is, from very high-technology oriented to off-the-shelf technology oriented, from cultural heritage to science resources and so on. Thus, the meaning of the term is not limited to a digitally archived collection of governmental records. Another term which has a meaning close to digital archive is digital curation. This paper uses these two terms with the same meaning. Another point which should be mentioned is that this paper does not distinguish between digital resources that are born-digital or digitised from non-digital.

The trend towards Open Data and Linked Open Data (LOD) is a crucial issue for the development of digital archives. Once a dataset is open, it is linked and used in combination with other resources by third parties. Digital archives should be open and linkable to each other to enhance their usability and to add value to the archived contents. Metadata is always key for the enhancement.

Memory institutions and governments have been playing important roles in building digital archives. In North America, American Memory at the US Library of Congress has more than 20 years' history of the development of large digital collections of cultural resources, and the Digital Public Library of America provides a broad range of resources. Europeana is a huge international collaborative project in Europe and its strength is not only in international collaboration but also in the use of LOD technologies for value addition by third parties.[1] In Asia, there are many digital archive projects as well. The Taiwan e-Learning and Digital Archives Program (TELDAP) is a good example of the use of digital archives in education.

Disasters, which may be caused by nature or by humans and may be large or regional, affect our lives and societies. On one hand, disasters are one of the major risks for valuable cultural resources of communities. On the other hand, memories and records about the disasters are also an important asset for the communities. Digital archives have a crucial role in keeping memories of disasters for the future.[2] In Japan, there are several archives that record the Great East Japan Earthquake which are described later in this paper. Community memory resources may, however, be destroyed by disasters which happen slowly, for example climate change and social changes of aging communities.

The Digital Curation Centre in the UK fosters long-term use of digital resources and their collections. Digital preservation and the longevity of digital resources are obviously important for digital archives. More importantly, longevity adds value to the

resources and archives.[3] Longevity of digital resources has been discussed since the 1990s. The risks of digital preservation have been discussed mainly from technological and/or management viewpoints. Longevity of metadata is a crucial aspect for digital archives as well as longevity of primary resources in digital formats, because we lose important handles to access and render the resources if their metadata is lost. Longevity of meta-metadata, which is metadata about metadata, is key to keeping metadata understandable and usable over time. Therefore, registries to keep information about metadata and archives are crucial. For example, the UK National Archives built a registry service named PRONOM to keep information about file formats.[4]

Provenance of resources is recognised as critical information for digital preservation as provenance is one of four main categories of Preservation Description Information of the Open Archival Information System.[5] The World Wide Web consortium has proposed a model to describe provenance which is not specifically designed for digital archiving but gives us a generalised model for provenance description.[6] Metadata provenance is also crucial to track changes in metadata schemas as well as mappings among metadata vocabularies. Metadata provenance description is meta-metadata because it is a description of metadata. Eckert and colleagues reported a model for meta-metadata based on the Resource Description Framework (RDF).[7] In general, preservation of metadata as a document would not be as complicated as preservation of computer games because metadata is described as textual data. However, the requirement to keep track of semantic changes of metadata and meta-metadata in a machine-interpretable form presents a challenge.

Open Data and Linked Open Data are very important movements for memory institutions and related communities. Governments are making an effort to make governmental resources open and usable for the public in order to add more value to their data. LOD is a crucial movement to link data available on the Web by semantic links.[8] Europeana, for example, is a large-scale application of LOD. Federated search has been a common function among digital archives. However, those digital archives participating in the federated search have to agree on a protocol and metadata schema for search in advance, that is, semi-open data. LOD provides us with a less tightly connected model for resource access across different archives. Thus, it is crucial to link digital archives by semantic links based on LOD across communities and over time.

Digital archives of memory institutions in Japan

Overview of digital archives

In Japan, the National Diet Library (NDL), the Agency for Cultural Affairs (ACA) under the Ministry of Education, Culture, Sports, Science and Technology, and the National Archives of Japan (NAJ) have been developing large digital archives – for example, the Digital Library of Books published in the Meiji era and early twentieth century by NDL.[9] NDL has a wealth of experience in developing digital archives and federated search services,[10] Cultural Heritage Online (CHO) by the ACA and National Institute of Informatics (NII),[11] and the digital collection of government records covering the period from after the revolution in the nineteenth century to World War II by the Japan Center for Asian Historical Records at NAJ.[12] NII is a national hub for scholarly information that provides CiNii as a portal to scholarly information resources.[13]

Some university and regional public memory institutions have been working on digitisation of their holdings and regional heritage. National museums have also developed digital collections of their holdings. For university libraries, digitisation of cultural heritage is not a central service but they have been keenly working on electronic journal

services and institutional repositories. Some national museums have produced digital collections and collaborated with ACA for the development of CHO. On the other hand, digitisation activities of the regional memory institutions, ie, Museums, Libraries and Archives (MLA), are still limited.

The disaster caused by the great earthquake and tsunami on 11 March 2011 was recorded by many people using digital devices. MIC, NDL, Tohoku University,[14] Harvard Reischauer Center[15] and other partners have been collaborating to collect digital resources to build a digital archive of the disaster. NDL, as the hub of the partners, opened Hinagiku, the digital archive of the disaster, in March 2013.[16] Hinagiku works as a portal to the participating digital archives. The experience at Kobe University Library in building their disaster archive for the earthquake that happened in January 1995 helped the recent archiving activity.

Archiving is an ongoing task. Digital archives for a major disaster play an important role from many viewpoints – information resources for governments to build their risk management plans; scholarly resources for researchers in many fields such as civil engineering, social risk management, risk management technologies, history and sociology; educational resources at many levels; and community memory resources for people in regional communities. An important challenge for a disaster archive is to enhance the usability of archived resources for the general public across communities, for example reference services using the disaster archives, and help for people and communities to record their personal or local memories of the disaster.

MIC Study Group on Digital Archives and NDL Roundtable on Digital Information Resources

The Ministry of Internal Affairs and Communications (MIC) hosted a study group from February 2011 to March 2012 to discuss issues in the promotion of digital archives of MLAs.[17] The author participated in the group as its chairperson. The NDL hosted a Roundtable on Digital Information Resources to discuss issues in the promotion of the collection and use of digital information resources, including the promotion of collaboration among MLAs. The roundtable started in 2010 and ended in 2012.

Both groups shared concerns about the development of digital archives in Japan, particularly developments at regional MLAs. MLA collaboration was a main topic of the NDL roundtable. NAJ have been developing their digital archives as a core function. On the other hand, digitisation at national museums and their metadata exchange with other institutions are less visible. There is a common understanding that MLA collaboration is crucial to promote use of heritage resources. Connecting resources stored at different institutions is vital for MLA collaboration, which means that metadata sharing is key to connecting the resources of MLAs. However, metadata sharing among MLAs is still limited because there is no well-recognised standard and limited development of shareable metadata at museums. Another issue shared among the participants was the accessibility of museum resource information via the Internet.

Discussions at the MIC Study Group

The group first agreed that MLAs, especially those founded by regional governments, are facing common barriers such as a lack of human resources and insufficient IT skills and experience. Longevity of digital resources is also a common concern for memory institutions when investing their precious funds.

The great earthquake happened on 11 March 2011, soon after the first meeting of the MIC Study Group. The disaster significantly affected the discussions of this study group. Many MLAs facing the Pacific coast were seriously damaged by the tsunami. Many rescue activities were carried out 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, rescue activity soon after the disaster was not possible and continuity of the regional community was threatened.

The study group agreed that promotion of digital archives is a key strategy for MLAs aiming to keep their community memory safe from disasters and changes to social environments, in spite of arguments of lack of longevity of digital resources compared with physical resources. What the author learnt from this discussion is that physical things are easily lost in disasters and digital resources properly stored and backed up are more robust against disasters. Another important point agreed by the study group is that it is crucial to create and maintain metadata of resources to make them usable also in the networked environment. This means that it is important to maintain information about metadata, that is, meta-metadata and metadata schema. Thus, we agreed to understand the problems in a layered view – collection building, maintenance and provision of access to digital resources, data storage and preservation to share the digital resources, metadata storage and sharing to share information about the digital resources, and meta-metadata storage and sharing to share information about metadata.

Digital preservation is obviously critical for the longevity of digital archives. It is necessary to reduce both the risk of loss of contents and costs of preservation in both financial and environmental respects. The most fundamental issue for a memory institution is to establish its preservation policies and a maintenance system for preserved resources. Sharing an archiving infrastructure looks reasonable but the different requirements of memory institutions can make it difficult.

The final report of the study group recommends the promotion of the following four principles. These are primarily agreed for the development of digital archives at small- and medium-scale MLAs.

(1) enhance open access to resources at memory institutions – moving from paper-based management to digital resource management;
(2) develop human resource development infrastructure;
(3) develop system and service infrastructure;
(4) develop metadata information-sharing infrastructure.

Library information services are fairly well standardised and provided via the Internet. However, the management of resource information at museums is, in general, not as well standardised as library resource information because of the diversity of collections. Providing information about museum holdings via the Internet enormously improves the accessibility to those resources. Human resource development is a necessity for regional MLAs developing digital collections and improving their services via the Internet. The report recommends building a network of people and organisations to help MLAs develop their human resources. The report also recommends building a cloud environment for MLAs. For museums, the report encourages the use of Cultural Heritage Online as their shared infrastructure to make their heritage resources visible on the Internet. The group also agree that the Linked Open Data movement is crucial to

improving accessibility and usability of digital archives. As an example of good practice, the group discussed the LODAC museum project hosted at NII, which is collecting information resources from museum sites using LOD technology.[18] The report recommends using URIs to identify every resource and to adopt the International Standard Identifier of Libraries and Related Organisations (ISIL, ISO 15511:2009) as the base identification scheme. The group agreed that sharing metadata information is crucial to enhancing metadata interoperability. A fundamental service for metadata information sharing is metadata schema registries that suit the Web, particularly LOD environments. Another key issue in sharing metadata is the rights statements about the metadata, a kind of meta-metadata.

Metadata issues for digital archives

Interoperability – Dublin Core-based view

Metadata interoperability is a key issue for digital archives in two respects – interoperability across domains and over time. The Dublin Core Application Profile (DCAP) is an important framework for metadata interoperability.[19] DCAP splits structural features and semantic components of metadata to define interoperable metadata across domains. In order to enhance semantic interoperability among the metadata schemas, DCAP suggests metadata schema designers use existing terms rather than define new terms because metadata vocabularies are the principal semantic basis of the metadata. Formal schemes to share metadata schemas are necessary for better metadata interoperability. Both the RDF and the Web Ontology Language (OWL) are important standards for metadata schema.

Meta-metadata – a generalised view for metadata interoperability

Metadata is defined as '(structured) data about data'. There is a data about a metadata, that is, meta-metadata, meta-meta-metadata and so forth. The 1-to-1 (one-to-one) principle of Dublin Core, which says one metadata for one instance, is a simple and crucial model for understanding the importance of meta-metadata.

Let us take a catalogue record of a book as an example. The record is eventually updated and the processing history is recorded. The update information and processing history is metadata of a catalogue record. This metadata is crucial for catalogue management but is not metadata of any book. As a catalogue record is metadata of a book, metadata about the catalogue record is meta-metadata of the book. Learning Object Metadata (LOM), for example, has a category named meta-metadata, which includes attributes for management of metadata. For another example, provenance description of metadata of a preserved resource is meta-metadata of the preserved resource. Thus, provenance information of metadata of a preserved resource is important as well as provenance information of the preserved resource.

Longevity of metadata schemas

Longevity of metadata schemas involves both semantics and infrastructure. A basic semantic problem is the change of meanings of metadata terms over time, namely the semantic drift of words and phrases. Another is the longevity of identification schemes for identifying metadata terms, vocabularies and schemas.

An obvious solution from the semantic perspective is to preserve metadata vocabulary definitions as a document. A metadata schema registry described in the next section is an efficient tool to maintain metadata vocabularies over time and across communities. It is necessary for the long-term maintenance of metadata to include appropriate references to the preserved metadata vocabularies from metadata schemas, that is, metadata for metadata schemas. Standardised metadata schemas and their vocabularies are maintained by maintenance agencies of metadata standards. Unfortunately, not all maintenance agencies live forever, so there is a risk that community-based metadata vocabulary documents will be lost over time. Keeping metadata schema information in a registry trusted and shared among the communities would be a simple solution.

URIs are the base identification scheme of resources on the Web. Each URI is an identifier of a corresponding metadata term or a resource. All of these links should be maintained over time but it is not easy to preserve links consistently because domain names are not persistent.

Metadata schema registry – a crucial service for metadata schema sharing across communities and over time

A metadata schema registry (or simply metadata registry) is a service to store metadata schemas and their components and to provide them on the Internet.[20] The primary instances stored in a registry are metadata vocabularies, that is, terms which represent metadata attributes, value classes and so on. On the Internet, both humans and machines use metadata, which means that every metadata term should be properly labelled for humans and, at the same time, should be given a unique identifier for machines. Therefore, proper localisation of metadata labels, and management and maintenance of URIs of metadata terms are the fundamental requirements for metadata interoperability across communities and over time.

The Open Metadata Registry[21] provides a rich set of metadata terms. Schema.org[22] and Linked Open Vocabulary[23] also provide rich sets of metadata vocabularies. The author has been involved in the development of two registries – DCMI Metadata Registry[24] and MetaBridge.[25] Both provide metadata schemas in RDF. MetaBridge stores not only metadata vocabularies but also description set profiles.

Digital archives and disaster – some lessons learnt

Great East Japan Earthquake digital archive

Memory institutions are responsible for keeping our memory for the future. As mentioned in a previous section, there are large digital archiving efforts led by the Japanese government to record the disaster and keep community memories into the future. Hinagiku at NDL, launched for public access in March 2013, is based on the collaboration of many public and private sectors – universities, libraries, regional and national broadcasting stations, newspaper companies, IT companies, local communities and so forth. NDL, Tohoku University and other partners have been collecting the resources and are continuing their efforts. An important issue for these archiving efforts is to get broader interest and higher recognition among the general public.

Hinagiku works as a search portal to access many databases – 29 participating databases including 16 databases dedicated to the East Japan Earthquake. These participating databases are organised by different types of organisations – university libraries,

regional public libraries, newspaper and broadcasting companies, regional public sectors and Non-Profit Organisations, and so forth.

In-depth discussions on some crucial issues are required to further develop the services such as pictures which contain private content and shocking scenes, secondary use of archived resources, reference services provided by regional memory institutions for regional people using archived resources, and so forth. A crucial issue for Hinagiku is the promotion of its use by those people living in the area and seriously affected by the disaster to keep their memories for the future and to help them recover from the disaster. Regional needs and nationwide needs are sometimes hard to satisfy at the same time. For example, terms and phrases used in regional communities are useful for regional people keeping their memories in the archive but are not always easy to understand for other users. The issue left for future research and development is to extend the functionality of the digital archives from preservation-and-access to linking people via the archived resources in order to help them keep their community memories.

The power of cloud and crowd

Regional government offices, museums, libraries and cultural heritage sites facing the Pacific coast were seriously damaged in the disaster. Physical objects were lost or damaged. Loss of metadata, for example catalogue records and provenance information, caused great difficulties for recovery of objects found after the disaster. An obvious conclusion from this experience is that physical objects are easily lost and digital copies of important resources should be stored in a robust environment to avoid the loss of memory.

Disasters on that scale do not happen frequently. However, smaller disasters, which do happen more frequently, also cause serious damage to the memories and records of regional communities. Environmental and social changes may cause serious damage, which may happen slowly but can be serious. There are arguments about the longevity of digital media and the costs of digitisation but the author considers that digital archiving is a solution for keeping community memory resources safe for the future regardless of the resource types – born-digital or non-digital, tangible or intangible. An important lesson is that institutions and communities that cannot afford digital archives should share a trusted repository to keep their memories safe.

Loss of metadata causes serious problems. Once metadata of an object is lost, it is very expensive to recover the metadata even if the archived object is kept safe or rescued. It is particularly hard to recover contextual information such as provenance, classification and value descriptions of the archived object. The intellectual power of humans is essential for recovering or adding contextual information. Crowdsourcing technology and practices would be one way to solve this problem. Morishima and colleagues presented a platform for crowdsourcing that applies human intelligence to semantic micro-tasks on a logical framework for artificial intelligence which is designed to cope with a large volume of data.[26] They apply the platform to track the running paths of tornadoes and to correct semantic errors in the bibliographic data of NDL. In the former example, it is easy for regional people to identify the location where a photo of a tornado is taken but very hard to automatically find the location. On the other hand, it is easy for computers to merge many data created by people into a database.

Linking digital resources across communities and over time

Linking born-digital resources on the Web

Electronic books are important new cultural resources. In traditional print publishing, electronic catalogue records of books are created separately from the books. However, in the electronic publishing environment, both book contents and catalogue records are realised as a digital instance and distributed on the Internet and Web. On the other hand, video and audio resources are also in a digital format and broadcasting stations broadcast their programs in a digital format. Thus, the borders separating books, video and audio are diminishing in the digital and networked environment. In the conventional print-centric publishing and analogue broadcasting environment, metadata is created separately from the primary resources, for example books and TV programs and their catalogue data. However, in the digital environment, metadata could be encoded and embedded within the primary resources. In the digital environment, the border between archived and non-archived resources is unclear because both resources are stored in a database and served over the Internet.

Linked Open Data has a large potential for the development of digital archives in the open community. Metadata embedded in a resource works as a link connecting resources. Those metadata semantically link resources. On one hand, those semantic links will add value to the digital archives; on the other hand, maintaining the semantics of links as well as maintaining links is a new research issue.

A cloud-based model for digital preservation

The preservation of digital objects is a crucial function for digital archives. Preserved objects may have longer lives than the archival systems where they were born. Cloud environments have potential for MLAs as robust environments to keep and share digital resources. The hierarchical layered model based on the layered model of cloud computing is worth discussing because it clarifies the layers of services – from domain-neutral services in the base layers to domain-specific services in the application layers.

The variety of digital objects is always increasing because the variety of devices and digitally published resources is always increasing, for example smart phones, tablet PCs, electronic books and games. In addition, metadata which describe intellectual, logical and physical features of every digital object have to be preserved with the object. The variety of metadata schemas is always increasing. Thus, memory institutions have a demand to store a large variety of digital objects in accordance with their own requirements, which is likely to increase the complexity of any digital preservation. The layered model shown in Figure 1 contains an additional layer that bridges the application-specific requirements and an application-neutral preservation system based on the Open Archival Information System (OAIS).

Cloud computing is a natural place for small memory institutions to implement their digital archives. A cloud computing environment is defined as a layered architecture, for example as Infrastructure as a Service (IaaS) and Software as a Service (SaaS). Hiding detailed information defined in lower layers is a primary function of the layered architecture. On the other hand, because 'packaging' is a central idea of OAIS we need to re-define 'packaging' suitable for a layered architecture of cloud computing. Askhoj and colleagues showed a layered model for preservation in the cloud environment where a layer for preservation is defined between an application service layer and a platform layer.[27] It maps entities in the OAIS reference model into the layered model.
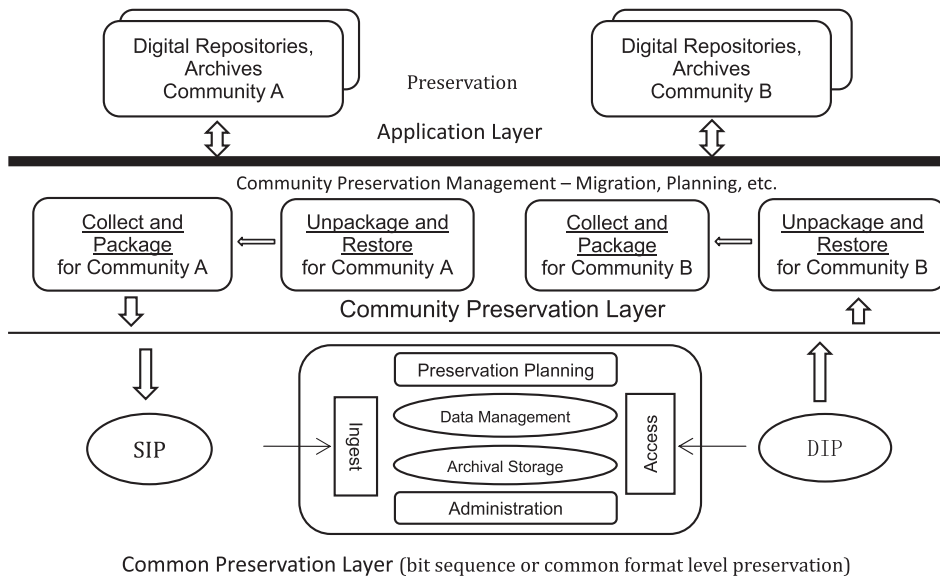
Figure 1.   OAIS and layered model.

Discussion and concluding remarks

Digital archives are essential for our communities to keep our memory safe for the future. Metadata is key for the longevity of digital resources and digital archives. Longevity of archived resources is a primary concern of memory institutions. Physical entities are easily lost. The loss is caused not only by natural disasters but also by human activities. There are many arguments about the longevity of digital resources because of rapid technological changes. Longevity is primarily a management and policy issue rather than a technology issue.

A cloud computing environment has great potential for memory institutions building scalable and trusted digital archives because those institutions can use computation and storage power without worrying about human resources for maintaining computer systems. On the other hand, semantic information, which only humans can create, is crucial for memory institutions organising and providing information resources. Therefore, combining human intelligence and machine intelligence is necessary for the further progress of digital archives.

The development of electronic books and publishing industries will affect the organisation of digital collections and their archives. An important point in both the digital publishing and archiving environments is that any instance described in metadata can be a digital instance identified by a URI, so that any instance can be linked to other instances. Metadata and meta-metadata are also first-class instances as well as digital books and images. LOD technologies are obviously crucial for memory institutions to handle networked resources.

There is consensus that digital archiving is important for memory institutions regardless of their size. The lesson learnt from the great disaster of 11 March 2011 is that it is crucial to have securely stored digital copies, if we are to keep cultural heritage and memories of a community safe for the future. We also learnt that a lack or loss of metadata caused further damage to community memory.

Another lesson is the necessity of infrastructure to share metadata in order to improve the interoperability of our digital archives. RDF and LOD augment semantic linkage among the information resources on the Internet. Sharing metadata terms is a good way to improve semantic interoperability of metadata. Metadata schema registries that store and provide metadata terms enable semantic interoperability of metadata terms. Metadata interoperability over time is as important an issue as interoperability across communities. Preservation of both metadata and metadata terms is fundamental to keeping digital contents usable. However, current metadata schema preservation is not perfect.

Communities need information environments customised to their needs. Communities need information environments shared with other communities for information sharing and cost-saving. Thus, both specialisation for each community and generalisation among communities are required for community digital archives. We need not only good technologies but also good management practices and policies to solve these contradictions.

Acknowledgements

Notes

1. Europeana homepage, available at <*http://europeana.eu/*>, accessed February 2014.
2. Crisis, Tragedy, and Recovery Network, available at <*http://www.ctrnet.net/*>, accessed February 2014.
3. Digital Curation Centre, available at <*http://www.dcc.ac.uk/*>, accessed February 2014.
4. The National Archives, 'The Technical Registry PRONOM', available at <*http://www.nationalarchives.gov.uk/PRONOM/*>, accessed February 2014.
5. Consultative Committee for Space Data Systems (CCSDS), 'Reference Model for an Open Archival Information System Blue Book', June 2012, available at <*http://public.ccsds.org/publications/archive/650x0m2.pdf*>, accessed February 2014.
6. W3C Working Group Note, 'PROV Model Primer', 30 April 2013, available at <*http://www.w3.org/TR/prov-primer/*>, accessed February 2014.
7. K Eckert, M Pfeffer and H Stuckenschmidt, 'A Unified Approach for Representing Metametadata', in *Proceedings of International Conference on Dublin Core and Metadata Applications,* 2009, pp. 21–9, available at <*http://dcpapers.dublincore.org/pubs/article/view/973*>, accessed February 2014.
8. World Wide Web Consortium, 'Data – W3C', available at <*http://www.w3.org/standards/semanticweb/data*>, accessed February 2014.
9. National Diet Library, Digital Library From Meiji Era, available at <*http://kindai.ndl.go.jp/?__lang=en*>, accessed February 2014.
10. National Diet Library, 'List of Online Services', available at <*http://www.ndl.go.jp/en/service/online_service.html*>, accessed February 2014.
11. Cultural Heritage Online, available at <*http://bunka.nii.ac.jp/Index.do*>, accessed February 2014.
12. Japan Center for Asian Historical Records, available at <*http://www.jacar.go.jp/english/index.html*>, accessed February 2014.
13. National Institute of Informatics, 'CiNii', available at <*http://ci.nii.ac.jp/en*>, accessed February 2014.
14. Tohoku University, 'Michinoku Shinrokuden Archive', available at <*http://shinrokuden.irides.tohoku.ac.jp/*>, (in Japanese), accessed February 2014.
15. Harvard University, 'Digital Archive of Japan's 2011 Disasters', available at <*http://jdarchive.org/en/home*>, accessed February 2014.

16. National Diet Library, 'NDL Great East Japan Earthquake Archive', 2013, available at <*http://kn.ndl.go.jp/node?language=en*>, accessed February 2014.
17. Ministry of Internal Affairs and Communications, 'Expansion of Social Knowledge Infrastructure Through Digital Archive of Knowledge: Public Announcement of Proposal and Guidelines', 30 March 2013, available at <*http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Releases/Telecommunications/120330_06.html*>, accessed February 2014.
18. LODAC, 'Linked Open Data for Academia', 2013, available at <*http://lod.ac/*>, accessed February 2014.
19. Dublin Core Metadata Initiative, 'The Singapore Framework of Dublin Core Application Profiles', 14 January 2008, available at <*http://dublincore.org/documents/singapore-framework/*>, accessed February 2014.
20. 'ISO 11179: Part 1, Framework', 18 November 2013, available at <*http://metadata-standards.org/11179/*>, accessed February 2014.
21. 'Open Metadata Registry', available at <*http://metadataregistry.org/*>, accessed February 2014.
22. Schema.org, available at <*http://schema.org/*>, accessed February 2014.
23. Pierre-Yves Vandenbussche, 'Linked Open Vocabularies (LOV)', available at <*http://lov.okfn.org/dataset/lov/*>, accessed February 2014.
24. DCMI Registry, 'The Dublin Core Metadata Registry', 2013, available at <*http://dcmi.kc.tsukuba.ac.jp/dcregistry/*>, accessed February 2014.
25. MetaBridge, 'About MetaBridge', available at <*https://www.metabridge.jp/infolib/metabridge/menu/?lang=en*>, accessed February 2014.
26. A Morishima, N Shinagawa, T Mitsuishi, H Aoki and S Fukusumi, 'CyLog/Crowd4U: A Declarative Platform for Complex Data-Centric Crowdsourcing', *Proceedings of VLDB Endowment*, vol. 5, no. 12, 2012, pp. 1918–21.
27. J Askhoj, S Sugimoto and M Nagamori, 'A Metadata Framework for Cloud-Based Digital Archives Using METS with PREMIS', in *Proceedings of ICADL 2011 Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation, Lecture Notes in Computer Science*, vol. 7008, 2011, pp. 118–27. J Askhoj, S Sugimoto and M Nagamori, 'Preserving Metadata in a Cloud', *Records Management Journal, Emerald*, vol. 21, no. 3, 2011, pp. 175–87.