NATIONAL DIGITAL HERITAGE ARCHIVE PROGRAMME

Conference Review – ARCHIVING 2011

Cynthia Wu
July 2011

Document Revision History and Control

Version	Date	Author	Version Notes
1.0	15/07/2011	Cynthia Wu	

CLIO Reference No: #

Table of Contents

Conference Review – ARCHIVING 2011

1. Introduction

Purpose

This document provides a review of the ARCHIVING 2011 Conference held in Salt Lake City between May 16 and May 19, 2011.

Reference Documents

Document	Source	Date
JHOVE	http://hul.harvard.edu/jhove/	2003-2009
JHOVE 2	http://www.jhove2.org/	2011
PREMIS	http://www.loc.gov/standards/premis/	2011

2. Conference Review

I was fortunate to attend the Archiving 2011 conference held by the Imaging Society for Science and Technology in Salt Lake City, Utah, from May 16 to May 19th. This conference included a wide range of technical presentations, from in-depth digitizing technologies, development of Digital Preservation systems and best practices, to Digital Curation Internship programmes. Due to the geographic location of the conference, many of the speakers were from USA, with the exception of a few speakers from various institutions across Europe and Asia. Conference participants were also invited to join various technical short courses, of which I attended two relating to JHOVE 2 and PREMIS.

Long Term Digital Preservation

Digital Preservation is 'an ongoing, proactive process of preserving information and its significance over time' (Korenkova & Hägerfors, 2011). As more institutions around the world begin to consider the need for long-term digital preservation and partake in such processes and activities, the digital preservation community are gaining better insights into the complexity and unique issues it faces during the preservation process. During the conference, several speakers reiterated the need to work collaboratively. Jay Verkler from FamilySearch in the United States pointed out that sharing digital preservation knowledge and experiences could help accelerate the evolution of this field. Margarita Korenkova and Ann Hägerfors from the Luleå University of Technology in Sweden further pointed out that the digital preservation community needs a unified information quality assurance framework. This framework can be used as guidance for all institutions considering long-term digital preservation, as well as criterion to ensuring organisations maximise the value of the objects they preserve. This idea is interesting as most of the better-known frameworks only covered the infrastructure and metadata part of the digital preservation process. There has yet to be any formal framework for measuring the quality of digital preservation, though one could argue that this can be influenced by other constraints such as skills and technology available within any institution.

Current State of Digital Preservation

As our own National Digital Heritage Archive moved into version 2.1.2 at the end of last month, and with 2.2 and version 3 of the system in sight, the conference was an excellent reminder of how much the National Library of New Zealand has achieved in the field of Digital Preservation. A number of institutions within the United States had recently revised their systems to improve performance, interoperability, scalability, metadata capture and support of new file formats. Many others are still in the process of designing and building their own digital preservation systems. As the conference speakers pointed out that collaboration is key, the National Library of New Zealand's programme of outreach to the international communities has enabled us to build stronger relationships both within the Digital Preservation realm and other related fields. In turn, this outreach programme ensures we remain relevant, helps us to be innovative and stay at the

forefront of this field.

Digital Curation and Preservation as a Service

Etelä-Savon tietohallinto Oy (ESTH), a company based in eastern Finland, is currently running a noteworthy project called SARKK. SARKK is a non-profit based project aiming to provide Digital Archive Services for Finnish municipalities and municipal federations. These services encompass long-term preservation of stored objects, as well as permanent storage of these objects. The SARKK system provides integration mechanism with its customers to enable submission, searching and retrieval functionalities through the use of SARKK adapters. The SARKK project manages such services via service contracts, but ownership of the objects remaining with the customers. While this project provides comprehensive and cost-effective Digital Curation and Preservation services and expertise to its clients, some digital preservation strategies, such as migration and emulation were deemed to be unsuitable due to increase in complexity, risks, and costs.

The National Digital Heritage Archive, based on ExLibris' Rosetta system, has an infrastructure that can support similar services. The current Government Digital Archives Programme is a start towards this business model, but National Library can expand the scope of this Digital Archive to support smaller institutions where digital preservation expertise and technology resources are limited. The physical system can also be further developed to increase its scalability and flexibility. This sharing of services can encourage collaboration, expand our knowledge on newer file formats and raise the awareness of Digital Preservation across New Zealand.

Digital Curation Internships

The University of Houston's Digital Services Department has created an interesting Digital Curation internship programme since 2009. Under this programme, student interns rotate through four steps of digital collection creation: digitization, metadata creation, project management, and collection promotion. The last three steps of this programme are especially useful. Interns are required to describe digitised objects using national and local metadata standards, as well as being provided with training on the Dublin Core standard. The next step requires interns to examine past projects within the department and submit reports on the project resourcing and any possible challenges. Interns are required to write short articles for new digital collections, which will be published using various Web 2.0 technologies, such as Twitter, Facebook, and Flickr. The main advantage of this programme is its integration with the Master of Library Science (MLS) programs, promoting awareness for Digital Curation and enriching MLS students with practical experience, which may not be covered comprehensively within their theoretical studies.

Currently the Master of Information Studies (MIS) at the Victoria University of Wellington offers around three courses relating to Digital Preservation, Digital Curation and Digital Libraries. The National Library of New Zealand should consider building similar internship programme to provide MIS students with practical experience in these fields. Such programme can equip students with practical knowledge of ingesting and describing digital objects, the use and importance of metadata (descriptive and technical), current practices

and challenges with Digital Preservation, ensuring accessibility of digital objects through joint effort of digital preservation and delivery mechanism, as well as marketing of these digital collections to the wider public. This programme could also be beneficial for the National Library by strengthening the cohesiveness between various segments within the current National Digital Library along with the business units (such as Alexander Turnbull Library and the Legal Deposit team). MIS students may also assist the National Library to remain innovative through their fresh perspectives.

JHOVE2 and PREMIS

JHOVE and PREMIS are among the most important digital preservation tools and metadata standards, therefore it was an excellent opportunity to learn more about them in short technical courses held as part of the conference.

JHOVE is a joint project of JSTOR and the Harvard University Library to develop an extensible framework for format identification, validation and characterisation. Version 1 of the tool offered 12 format modules. JHOVE2 has been released this April with less supported file formats (ICC color profile, JPEG2000, PDF, SGML, Shapefile, TIFF, UTF-8, WAVE, XML) due to funding issues. The architecture and the API of the tool have been redesigned to identify files based on their internal signature. The characterisation part of the tool now encompasses format identification, validation, and report of significant properties pertinent to the format's preservation. An extra 'Assessment' process has also been added to automatically evaluate an object's acceptability into a repository based on the institution's policies. JHOVE2 is also able to characterise files of different formats within a directory, as well as container files such as arc, zip, or shapefiles.

PREMIS (Preservation Metadata: Implementation Strategies), an international working group developing digital preservation metadata published its first Data Dictionary in 2005. This year, they released version 2.1 which included amendments and clarifications. They are also expecting to release version 3.0 this year, which will include changes to the PREMIS data model, as well as addition and modification to existing semantic units.

Future Trends

Technological advance and exponential growth in electronic information creation and consumption has increased the complexity of Digital Preservation. Organisations such as the FamilySearch in USA have predicted a data growth of over 100 Petabytes (single copy of objects only) within 10 years (White 2011) and 300 Petabytes by 2025 (Creighton, Tilbury, & Evans, 2011). For large organisations, long-term digital preservation activities now require stable, cost-effective and durable storage solutions. As a result, organisations are investigating the use of magnetic tapes to store material that do not require high availability. Tape solutions are being seen as more scalable, less costly, have longer lifecycles and consume

less power. Magnetic tape technology is also evolving to become higher density (more capacity), with better integrity information such as CRC obtained at the time of writing being attached with the object during transfer. However, the use of magnetic tapes for digital preservation is not without challenges, as White (2011) outlined in his institution's analysis in tape technology for large scale digital preservation activities.

Throughout the conference, there was an increasing awareness that further frameworks and standards are required to guide the worldwide library and archiving community in the development of digital preservation activities. Existing models and standards include the OAIS, TRAC, ISO MOIMS-RAC & PREMIS, covering the architecture, audit and certification, and metadata standard for digital repositories. Further guidance for infrastructure, technology and digital preservation best practices are still needed to ensure the digital objects in global digital repositories can remain accessible and functional for generations to come.

Conclusion

All in all this was an excellent and worthwhile conference to keep abreast of the continual innovations within the field, and to meet and exchange ideas with other Digital Preservation practitioners. The technical short courses were also very informative. The conference certainly raised some thought-provoking concepts and ideas, such as providing Digital Curation and Preservation as a service to other institutions, offering digital preservation internships within an organisation, and a unified framework for assessing the quality of digital preservation. The use of tape technologies as a long-term digital preservation storage solution will no doubt be a welcomed solution to many organisations. For further reading, please refer to the full conference proceedings at the Society for Imaging Science and Technology website.

3. References

Caplan P. & Chou, C. (2011). DAITSS grows up: migrating to a second generation preservation system. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

Creighton, T., Tilbury, J, & Evans, M. (2011). Digital object curation at scale. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

Korenkova, M. & Hägerfors, A. (2011). Quality criteria for digital information in long-term digital preservation. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

Ryhänen, K. (2011). SARKK – Comprehensive digital archive services for Finnish municipalities. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

Verkler, J. (2011). Preservation in a digital age. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

Westbrook, R. & Reilly, M. (2011). Growing a mutually-beneficial digital curation internship program that is sustainable and low cost. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.

White, G. (2011). Using tape for large-scale digital preservation. Unpublished paper presented at the Archiving 2011, Salt Lake City, UT.