
Introduction

DEBORAH WOODYARD-ROBINSON

The preservation of digital resources is about finding ways to maintain our digital heritage, whether it exists in the form of e-journals, database records, Web sites, emails, digital images, audio-visual materials, interactive programs, or any other kind of binary data. Libraries frequently engage with a wide variety of these resources and understand how quickly and easily we lose use of them when computers change or links break.

I have been privileged to work in national libraries for nine years, devoting my time solely to the task of digital preservation. In that time many different solutions have been proposed, and as many opinions have been voiced about whether these solutions will work. While it is often perceived that these opinions contradict or argue with each other, the real answer is in finding the balance of what works in a particular situation. I hope that through this issue of *Library Trends* we may start to see how each solution has its own benefit in a particular context.

All the articles in this issue agree on the challenges of digital preservation: the burgeoning volumes of digital resources to be dealt with; the temporary nature of digital materials, particularly the rapid changes in their formats; and the massive organizational revolution required to work with and preserve digital materials, including the associated costs. However, the context in which digital preservation challenges apply is a highly important factor in making the right choice of a solution. For example, the context of the data can change its meaning and use. The context of the resource can change its relevance and integrity. The context of a collection in an organization can impact its value, and the context of a collecting organization can affect legal rights and obligations for digital preservation.

The articles in this issue illustrate a variety of solutions in a range of contexts, from the implementation of large systems to issue-specific solu-

tions that are still in development. The first three articles feature large-scale digital preservation storage systems being built in three different environments: an academic library, a national library, and a media organization. We can see similar issues arise in each story but with different details that impact the solutions chosen. The interesting contrast here is largely affected by legal rights and responsibilities.

The heterogeneity of digital materials being collected increases complexity in systems and solutions, as shown in the first article by MacKenzie Smith, who describes the development of DSpace at MIT. A significant lesson illustrated by DSpace is that complex systems benefit from an incremental building approach such as spiral development.

Cathy Smith describes the imperatives of the British Broadcasting Corporation (BBC) as a British media body and explores the responsibilities of organizations other than libraries to provide preservation and access to digital information. Developing trust and cooperation with organizations such as the BBC may ease the burden on libraries and archives to be the sole keepers of digital information.

The article by Johan Steenbakkers exemplifies the commitment of national libraries around the world to solving the difficulties of digital preservation. The Koninklijke Bibliotheek (KB), or National Library of the Netherlands, has provided strong leadership and innovation in digital preservation issues. Johan explores the impermanence of digital library collections and the new roles and responsibilities of digital publishing, in particular scientific journal publishers. The KB has been at the forefront of archiving system development and looking for practical preservation strategies by engaging with commercial research partners to try to encourage widely available and supported solutions. Part of this is a strong commitment to sharing and cooperation.

The focus of the issue then moves on to selection and acquisition of digital resources for preservation. Publishing models have changed radically, and the Internet provides an easy conduit for the proliferation of digital information; therefore libraries are faced with divining new and manageable methods of finding, selecting, and acquiring resources for their collections. Libraries often license or link to material online rather than store it locally, so it is no longer acceptable, legal, or even possible to passively keep digital publications in a collection.

Two articles in this issue present examples of a variety of approaches to selecting and capturing resources from the Internet for the purposes of long-term preservation. Margaret Phillips summarizes the differences between selective and comprehensive collection of resources, particularly as these approaches are employed by national libraries, and presents the history of the National Library of Australia's PANDORA Project. This focused Web archiving project has recently reviewed its selection processes,

and some interesting conclusions are discussed. The ability for expansion of the program was considered limited and the way forward is seen as cooperating with international research and developments in automated methods to complement the work already accomplished.

A method developed at the Bibliotheque Nationale de France for comparing the results of different Web archiving approaches is described by Julien Masanès. The discussion focuses on trying to find a balance between the cost of Web archiving practices and the quality of the resulting collections. Besides emphasizing the complexity of Web archiving, this article demonstrates that the choice of method depends on the requirements of the archive. The most promising aspect in Web archiving development is the enhanced ability of automated processes to both improve the quality of materials and reduce the cost of obtaining them, which are currently strong opposing forces.

From the moment digital resources become part of a collection, they require thoughtful management for their longevity. A major part of the management task is recording vital information needed to use and understand the data stored in an archive. The subset of information that supports processes for long-term maintenance is called *preservation metadata*. The point of balance required for preservation metadata is between creating hand-crafted, resource-intensive, comprehensive, high-quality metadata and affordable, automatically extracted and created, absolute minimum core metadata of unknown accuracy. It has not yet been agreed how much preservation metadata will be enough. How to create even the basic elements is still a matter of research. It is agreed, however, that preservation metadata will be necessary, and three articles in this issue discuss some of the main developments in this field.

Steve Knight at the National Library of New Zealand (NLNZ) explains for us the origins of preservation metadata and the perceived need for its creation. The NLNZ was an early player to propose an implementation for preservation metadata and has moved on to start developing automated tools to assist the population of required fields. This work demonstrates the viability of creating preservation metadata.

Another important aspect of preservation metadata work internationally is the need for consensus among digital archiving communities on what metadata is required in a system for long-term preservation. An extension of this is the creation and use of standards for such metadata. Priscilla Caplan and Rebecca Guenther have headed up an international body of cooperation to develop practical solutions. Their article describes the work of the PREMIS (Preservation Metadata Implementation Strategies) working group and their attempts to define what is required preservation metadata, and it examines the methods of metadata implementation. This valuable work will require further testing and cooperation before realistic consensus will be achieved.

Technical metadata is proving particularly challenging to create and manage, and even the PREMIS group defers this detail to proposed projects such as the Global Digital Format Registry (GDFR), modelled in the following article by Stephen Abrams. A distributed network of cooperating registries is envisaged for the global benefit of long-term digital preservation.

The penultimate article explores the use of networked environments to support the use and preservation of digital resources. Reagan Moore and Richard Marciano tie together all the elements required for digital preservation in a manageable model for a distributed environment. Their article provides a technical and methodical approach for maintaining large amounts of data and explores the benefits of data grids for long-term preservation.

Toward the end of this issue we find the articles becoming more technical in nature. This forces us to consider the complexity of the problems of digital preservation and the need for us, as library business experts, to be able to precisely describe our requirements for our digital repositories to people who have the technical expertise to build them.

The final article pulls us back out of the detail to look at the big picture of how the United States is managing and funding digital preservation initiatives. It explains this through a description of the National Digital Information Infrastructure and Preservation Program (NDIIPP), an impressive and ambitious program being watched closely worldwide. The eight funded projects and four affinity groups discussed are all bound to make significant progress in addressing the complex issues related to digital preservation.

Familiar themes are evident throughout the articles. Many of the projects refer to each other, and all discuss similar concepts. Standards such as the Open Archival Information System (OAIS) reference model¹ are also recognized universally. This is a result of the cooperation and information sharing that will be essential for the future of digital preservation. It will bring us closer to recognizing the scale of digital preservation solutions available and enable each of us to find our own balance.

NOTES

1. The OAIS reference model is a broad and logical model that can guide the logical development of an archival system. See http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html for more information.

Deborah Woodyard-Robinson, Digital Preservation Consultant, Woodyard-Robinson Holdings Ltd., P.O. Box 31308, Ilam, Christchurch, 8001, New Zealand, deb@woodyard-robinson.com. Deborah Woodyard-Robinson is currently working as an international Digital Preservation consultant based in New Zealand. Starting digital preservation work in 1996 as the first dedicated Digital Preservation Officer at the National Library of Australia, she was an author of groundbreaking work on the NLA Preservation Metadata Guidelines and contributed to the PADI (Preserving

Access to Digital Information) Web site and development of the PANDORA Web archiving project. In 2001 Deborah pioneered a digital preservation position at the British Library (BL) in London. As Digital Preservation Coordinator for over three years, she was a BL representative on the UK Digital Preservation Coalition and the Dutch KB/IBM Long Term Preservation Study and participated in international working groups such as PREMIS (Preservation Metadata: Implementation Strategies). Deborah also managed the initiation of the BL Web Archiving Programme, assisted in the creation of the UK Web Archiving Consortium, and developed a risk analysis method for the management of digital materials before moving to New Zealand in 2004.