A Framework for Good Digitisation in New Zealand

Version 2.0, June 2009

1. Introduction

This document outlines good practice principles and criteria for selecting and prioritising objects for digitisation from publicly held collections. Its aim is to provide the conceptual framework for the development of tools, guidelines and business cases for digitisation.

2. Purpose of framework

The framework for good digitisation has been produced primarily to enable the development of collaborative mechanisms for comparing and prioritising different types of collections and materials for digitisation. The focus of use is for selection policy development for collecting institutions (such as museums, archives and libraries) and other public bodies that hold significant content over the long term, as well as for funding agencies and budget-holders for digitisation activity.

3. A focus for collecting institutions and public bodies

As access to affordable digital technologies becomes widespread, digitisation activities are occurring in many different areas, one of the fastest growing being the commercial publishing and media industries. Digitisation of back issues of newspapers is becoming a new revenue stream for a number of newspaper companies, while digital transfers and restoration of movie stock and television programmes opens up new commercial opportunities for the entertainment industry. Digitisation can also create commercial opportunities for controlling new distribution of out of circulation material, as the Google Books programme has shown. However, in contrast to this private sector activity, success for the publicly funded and non-commercial sector is much more unevenly experienced.

In perhaps one of the first surveys of its kind in New Zealand, a 2002 research report of non-commercial digitisation in New Zealand found that activities were occurring in two distinct strands. The first strand was largely local and central government driven and generally focused on digitisation that supported business processes and electronic document management strategies[1]. This was a trend later reinforced by the 2003 E-government Strategy, which emphasised electronic delivery of public information and services[2]. The second strand was largely concentrated in libraries and museums, driven by the demand from users to increase access to culture and heritage resources over time. The research warned that unless addressed, this split could result in a large body of historically and culturally valuable documents that were never properly archived or preserved.

---

[1] *Creating a Digital Babylon? Results of a Survey of Digitisation Activities in New Zealand,* D.G. Dorner, B. Chawner, S. Searle, Victoria University of Wellington, 2002
[2] See http://www.e.govt.nz/archive/about-egovt/strategy/strategy-june-2003

In the last few years, as standards have emerged and become more widely adopted, the prospect of a significant loss of historically important records has waned somewhat. However, as recently as 2006 concerns were still being expressed about the lack of standards and non-proprietary tools in place for the government's Community Partnership Fund digitisation projects.[3] A number of these projects involved digitisation of unique community-held information, but with no guarantee that the results will be accessible in the long-term or that analogue originals will continue to be cared for. This suggests that unless there is widespread understanding of good practice, digital technology risks speeding up the loss of original heritage materials of the future, where digitising is erroneously viewed as a modern form of preservation.

In 1992 UNESCO launched the *Memory of the World* programme in response to a growing awareness of the poor state generally of preservation and access to documentary heritage in various parts of the world.[4] Historically, despite the best efforts of many museums, libraries and archives, many of the world's original heritage materials have been lost. War, social change, careless destruction and lack of resources characterise many of the past and present threats to documentary heritage. In the digital age, those threats now extend to format obsolescence, data corruption, and loss of contextual information in content that takes digital form. The core message of the *Memory of the World* is that information, and its systematic retrieval is the basis of the memory of civilisations. Safeguarding that information and keeping it accessible is not a trivial task.[5]

Drawing from this initiative, safeguarding access to New Zealand's own documentary heritage was the subject of a recent policy review led by the National Library. In 2005, a public consultation with a wide range of heritage institutions was undertaken, with a report prepared in 2006 on the findings. Adapting the *Memory of the World* definition, the report defined documentary heritage as including:

"…published and unpublished material; hand-written, type written and printed material on paper; material created and held in electronic format; photographs and works of art on paper and other media; sound recordings on magnetic tapes and compact discs; moving pictures on film and videotape; and oral history recordings and transcripts."[6]

This definition covers a wide variety of the materials that are frequently proposed as candidates for digitisation, whether for purposes of access or preservation.

While the policy review did not focus on digitisation and digital documentary heritage, it did identify a number of issues relating to digital technologies and access. In many cases collecting institutions have backlogs of cataloguing which limit access. The growth of the digital environment also increases expectations for those institutions to store, preserve and make accessible born-digital information, which requires balancing with the needs of traditional materials.

---

[3] *New Zealand Digital Content Strategy Submission Summary*, 2007
[4] UNESCO Memory of the World http://portal.unesco.org/ci/en/ev.php-URL_ID=23929&URL_DO=DO_TOPIC&URL_SECTION=201.html
[5] *Preserving our Documentary Heritage* UNESCO 2005
[6] *A Report on New Zealand's Documentary Heritage: Preservation and Access* June 2006, National Library of New Zealand

Digitisation often increases demand for access to the original item, as awareness of what is held in a collection increases. This can place greater pressure on, and risk to, the original than may have existed before digitisation.

Overall, while digital technologies present many opportunities to increase access to documentary heritage, the report concluded they also present many challenges in relation to digital preservation and digitisation, which few institutions outside those managed or governed by the Crown were well placed to deal with.

The review stopped short of offering solutions to the current capacity of the heritage sector. However, it seems evident that any move to publicly fund or coordinate collaborative digitisation efforts in New Zealand needs to take into account the risks for the sector as well as the risks for both traditional and digital documentary heritage.

Protecting documentary heritage is a legitimate starting point for public policy development on digitisation. Given digital technologies and digitisation are also capable of directly assisting the heritage sector to better manage collections for the long-term while improving access for the public, improving the capability of the sector to utilise those technologies is also a logical point of future intervention.

4.   Preservation and access

Protecting and managing New Zealand's documentary heritage means addressing preservation as part of the access equation – without preservation, there is no possibility of access over an extended period of time. Preservation activity is undertaken for long-term access, and almost always involves making surrogate copies of documentary material to 'back up' or to reduce wear in addition to other efforts aimed at conserving originals for as long as possible.[7] Considerations of risk to the original, rarity of the original, access costs involved in viewing, and demand for use are constraining factors for physical collections and frequently lead to the creation of preservation and 'access' copies. These increasingly lend themselves to digitisation.

Coupled with the historically unprecedented volumes of content being produced and stored in computer-based information systems, curators, archivists and librarians struggle to deal with the competing priorities between artefacts, paper, film, electronic and digital collections. Traditional access and preservation rules are turned upside down. Ancient fragile and unique originals can have digital copies made available to anyone with a computer. Modern electronic or digital records stored on magnetic tape or floppy disk can begin to decay through 'bit rot' and become unreadable within a year or two, while paper records stored properly can last 500 years.

It has been argued that as digital technologies evolve, digitally-based access and preservation strategies are converging, leading to an opportunity to address high volumes of decaying content (moving image in particular) through industrial-scale techniques in a way that addresses both access and preservation needs.[8]

---

[7] Conservation is a difficult concept in the digital environment, as migration strategies may preclude any attempt to retain original storage media beyond one or two generations of migration, if that. Use of redundant arrays of hard drives (RAID) is one such example.
[8] *Defining the Digital Archive*, The Film Archive, 2005

Similar arguments have been made for mass digitisation of books, with strategies aiming to have access copies as well as perpetual digital copies being actively pursued by a number of academic libraries across the U.S. and the U.K.[9]. There are some problems with these approaches however. One problem is cost. For example, by some estimates, storing film digitally as opposed to on traditional film stock can be up to twelve times more expensive once migration strategies and storage volumes are taken into account.[10] In the case of digitised books, the digital versions may be maintained as a long term additional cost on top of the physical collection management, while the current quality of the industrial digitisation processes used has been under question.[11]

Understanding access in a digital environment means understanding that being digitised does not of itself equate to making content more accessible. The explosive growth of super-abundant digital content right now is directly altering the utility and value of digitised documents, a relationship that to date has rarely been considered by digitisation programmes. The presence of super-abundant digital content alters the way users search for information, with the sheer scale of digital content available to be found introducing criteria beyond mere availability into the mix. Quality metadata – often costly and time-consuming to create - is one of those criteria.

Authenticity, persistent identifiers for referencing purposes, long-tail discovery, the ability to reformat for different devices or platforms, and the ability to re-use, re-mix or share the content legally are features highly valued by digitally literate users used to services from the likes of Google, Flickr and iTunes. Furthermore, the rapid growth of internet business models for sound and video content threatens to make stand-alone text and images 'second-best' as users seek out rich, integrated, multimedia experiences. This makes the availability of digitised audio-visual heritage content of equal if not greater importance than other documentary material if that heritage is going to be at all accessible.

5.   Defining "Good digitisation"

Digitisation is a popular term to convey a wide range of meanings and expectations, and as such lacks any precise definition. It is evident that there are many types of digitisation activities that have little in common with each other,[12] meaning considerable care is needed in using the term digitisation to convey anything beyond a very broad category of activity.

At its most basic, and as the definition for this framework, digitisation involves:

---

[9] *Preservation in the Age of Large-Scale Digitization A White Paper*, O. Rieger, Council on Library and Information Resources, 2008

[10] *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials*, The Academy of Motion Picture Arts and Sciences, 2007

[11] There has been extensive discussion online by librarians, historians and others, for instance, http://blog.historians.org/articles/204/google-books-whats-not-to-like, http://dltj.org/article/google-scholar-and-books/, http://radar.oreilly.com/2006/09/quality-of-book-digitization.html.

[12] for instance, e-book transcription, digital recording of oral history, and migration of video from VHS to DV tape are all widely considered digitisation activity, yet the purposes, audience and technical nature of the activity are all quite different

digital content creation by making a digital copy or digital recording of analogue information, where that information can reside in a document, artefact, sound, performance, geographical feature or natural phenomena.

Digital content creation includes data-entry and transcription, digital imaging, photography, sound and video recording and transfer – in fact any analogue-to-digital transfer. It excludes transcoding or migration of digital information into a different digital format or media (digital-to-digital transfer), software manipulation or programmed machine creation of new digital information (born-digital information), and analogue output of digital information such as printing or audiovisual playback (digital-to-analogue transfer).

The concept of "good" digitisation is drawn directly from the *Framework of Guidance for Building Good Digital Collections 3rd Edition[13]*, which argues that digital collection development has matured beyond proof of concept and collection-building projects to a point where digital objects, metadata and collections, in order to be "good", must be capable of being building blocks for others to reuse, repackage, repurpose and build services upon. In the same vein, this digitisation framework aims to place digitisation as core activity for collecting institutions, and one that fits within a wider context of good digital collection management and sustainable investment in digital technologies.

Inherent to any digitisation is a consideration of the technologies and techniques used for making digital copies. As with earlier technologies such as microfilm, not all techniques are born equal. While copying is extremely easy to do with digital content, accurately transferring an analogue original into digital form often requires specialised and calibrated equipment and software. Good digitisation needs to address the technology requirements while keeping them in balance them with content driven requirements.

6. Assessment principles

The good digitisation framework emerged from the observation that most existing selection policies have been developed from within single fields or organisations. The framework endeavours to draw together accepted digitisation and preservation principles from different collection and record management fields, and present them in a consistent and cross-disciplinary way.

The framework provides five assessment principles that each has a set of underlying criteria to apply to digitisation candidates:

a. **Selection**: not all content can or should be digitised

b. **Purpose**: digitisation should be for an identified purpose

c. **Access**: digitisation will affect access to the original

d. **Technique**: the digitisation technique will determine usability

e. **Value**: the digital copy should have intrinsic value

---

[13] National Information Standards Organization (NISO), December 2007

The intention is that these five principles can be applied across any digitisation candidates that are covered by the framework's digitisation definition. For example, it should be possible to apply these principles to a museum artefact, a letter and a sound recording that all relate to a particular event or theme, and arrive at an assessment of which is the highest priority to digitise.

Although all five principles should be addressed in creating a selection or digitisation policy, not all of the criteria will be relevant in a given situation. For instance, replacement of the original will very likely only be relevant for a digital preservation programme.

While it is envisaged that this framework can greatly assist decision-making, there are no definitive rules or answers for deciding what should be digitised. Each of the principles has to be tempered with good judgement before making a final decision. It is hoped however that the framework will improve consistency and transparency in the exercise of that judgement over time.

7.   The principle of selection

Underlying the principle of selection is the idea that, even with significant resource devoted to it, not all works or creations will be able to be digitised in the foreseeable future. It may also be that not all works should be digitised – for instance, information known to be duplicated, images that are blurry or incomprehensible, or miscellaneous outtakes from recordings. Other material may ultimately be digitised for preservation, but in the meantime are unable to be used for decades due to rights or privacy issues. Where it can be agreed that material can and should be digitised, there remains the question what to prioritise first. The principle of selection means giving priority to material that has the best chance of success in digital form.

Five criteria underlie the selection principle:

- i. **Organised:** A good digitisation candidate will be organised, itemised and described beforehand

- ii. **Scoped:** A good digitisation candidate will be scoped to quantify material for digitisation

- iii. **Scalable:** A good digitisation candidate can be scaled down if not all material can be digitised

- iv. **Permitted:** A good digitisation candidate has been permitted to be copied by the owner

- v. **Cleared:** A good digitisation candidate will not breach legal rights if digitised

8.   The principle of purpose

The purposes that organisations undertake digitisation for vary widely. Clearly understanding what purpose a digitised object may be used for is vital to the

selection and digitisation strategies chosen. Digitisation provides opportunities for increasing availability of content that is out of publication or limited in circulation (e.g. newspaper collections, journals, rare film). Where the format or information can be readily represented digitally without losing the important qualities of the original, a digitised copy can be an excellent surrogate for an original, perhaps protecting an object from added damage. On some occasions digitising is a key strategy for long-term content preservation through migration (e.g. for sound and video tape), while on others digital technology can enable new or different functionality (e.g. virtual collections, 3D manipulation, search across text).

Four criteria underlie the purpose principle:

i. **Availability:** A good digitisation candidate will increase opportunities for access and use by digitising

ii. **Representation:** A good digitisation candidate will be readily represented in digitised form

iii. **Replacement:** A good digitisation candidate will create a complete replacement sufficient to be permanently accessed instead of the original

iv. **Transcends:** A good digitisation candidate will enable uses in digital form that were difficult or impossible to do with the non-digital form

9. The principle of access to the original

Making a digital copy or recording of a non-digital object can be an important step towards increasing access to that object. Some non-digital objects are readily accessible because they are common or easily substituted, in low demand, or are inexpensive to make available (e.g. yesterday's newspaper). Others are limited by privacy or security concerns (e.g. military personnel records), or by donor restrictions on public distribution (e.g. family letters).   In order to gain the most improvement from access, a good digitisation candidate will likely be rare, in demand, at risk, and expensive or time-consuming to make available.

Four criteria underlie the access principle:

i. **Rarity:** A good digitisation candidate is unique or has few viewable or usable copies (regardless of form) that make it not generally easy to access

ii. **Demand:** A good digitisation candidate is in demand for viewing or using

iii. **At risk:** A good digitisation candidate has an unacceptable risk of damage or loss due to current access

iv. **Difficult to access:** A good digitisation candidate is expensive, time-consuming or technically difficult to gain or provide access to

10. The principle of technique

The techniques and technologies used to undertake digitisation activity directly affect the utility and sustainability of the effort involved in digitisation. Where digitisation may damage a rare object (e.g. cracking the spine of a book, breaking a fragile tape spool), or the object itself needs conservation treatment before digitising, careful consideration has to be given to both the preparation involved and the feasibility of the technology chosen for the job. Where the choice of digital format or physical carrier will limit interoperability (e.g. proprietary technologies not in widespread use), or there is limited availability of required hardware or software, usability of the digital object will be lessened. Where there is a strategy to manage the newly digitised content (e.g. as part of a digital collection), create backups and migrate it to new physical carriers over time, the opportunity for return on the effort to digitise will be much greater.

Four criteria underlie the technique principle:

i. **Prepared:** A good digitisation candidate can be readily prepared for copying

ii. **Feasible:** A good digitisation candidate can be digitised with available and appropriate technology and without causing undue damage

iii. **Usable:** A good digitisation candidate can provide a clear metadata description and copyright statement that supports usability and persistence

iv. **Manageable:** A good digitisation candidate can be digitised to a format and physical carrier that will be managed, interoperable and migrated over time

11. The principle of value

The value of a non-digital object will not always be successfully translated once it is digitised. For research and reference purposes, the physical carrier or context can be as important as the information that can be digitised (e.g. marginalia in a book, texture or smell in an artefact, a collection of letters). Its use as a resource and potential economic or social contribution may need to be considered. An original object may have an intrinsic meaning or emotional connection that cannot be replicated digitally (e.g. a bible that belonged to an ancestor). Digitisation opens up new possibilities use of the digitised copy that may not have been intended by the original owners or creators. Ensuring provenance and continued ownership throughout the lifetime of the copy can help protect the integrity of both the original and the copy.

Four criteria underlie the values principle:

i. **Evidence:** A good digitisation candidate will have integrity as a reference source in digital form for research and citation, and remain connected with the context of the original

ii. **Resource:** A good digitisation candidate will have value as a resource, able to be used for new works or opportunities

iii. **Meaning:** A good digitisation candidate will convey the meaning or significance of the original object in digital form

iv. **Integrity:** A good digitisation candidate will have a clear provenance and maintain its integrity in digital form throughout its lifetime

12.   Application of the framework

An initial application of this framework is to provide the foundation for a basic set of selection and analysis aids for identifying potential candidates for digitisation where no other selection criteria exist. These are expected to be particularly useful for those on limited budgets starting out on a digitisation programme for the first time, or for building a new digital collection that draws on a number of non-digital sources.

Through the Digital New Zealand project development process, a digitisation scorecard has been developed and tested, with the first public release of the scorecard being made in June 2009. More information about the scorecard and its usage can be found at http://makeit.digitalnz.org.

It is hoped others will build on these principles and the scorecard tool to improve future knowledge and practice relating to digitisation selection.

13.   Further Reading

Audiovisual research collections and their preservation - by D. Schuller, European Commission on Preservation and Access, 2008

BCR's CDP Digital Imaging Best Practices Version 2.0 - Bibliographical Center for Research, June 2008

Beyond the Paper Trail: National Archives in the Digital Age - by G. Davison, *Dialogue* Vol. 26 No. 2, 2007

The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials - The Academy of Motion Picture Arts and Sciences, 2007

Digitisation Activities: Best Practice Guidelines - The University of Auckland Library, 2007

Digitisation Selection Work - National Digital Forum, 2008

Digital Continuity Action Plan- Archives New Zealand, September 2008

A Framework of Guidance for Building Good Digital Collections 3rd Edition - National Information Standards Organization, December 2007

Handbook for Digital Projects: A Management Tool for Preservation and Access - Northeast Document Conservation Center, 2000

History, Digitized (and Abridged) - by K. Hefner, *New York Times*, March 2007

NZETC Digitisation Selection Policy - New Zealand Electronic Text Centre, 2007

One Word: Digital - by K. Coyle, preprint, published in *The Journal of Academic Librarianship*, Vol. 32, No. 2, 2006

Selecting Research Collections for Digitization - by D. Hazen et. al., Council on Library and Information Resources, August 1998

Task Force to establish Selection Criteria of Analogue and Digital Audio Contents for Transfer to Data Formats for Preservation Purposes - International Association of Sound and Audiovisual Archives, August 2003

Preservation in the Age of Large-Scale Digitization A White Paper - by O. Rieger, Council on Library and Information Resources, February 2008

The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy - International Association of Sound and Audiovisual Archives, 2005

Saving Tape - The New Zealand Film Archive

Sustainability of Digital Formats: Sustainability Factors - Library of Congress, March 2007