**Australian Government**

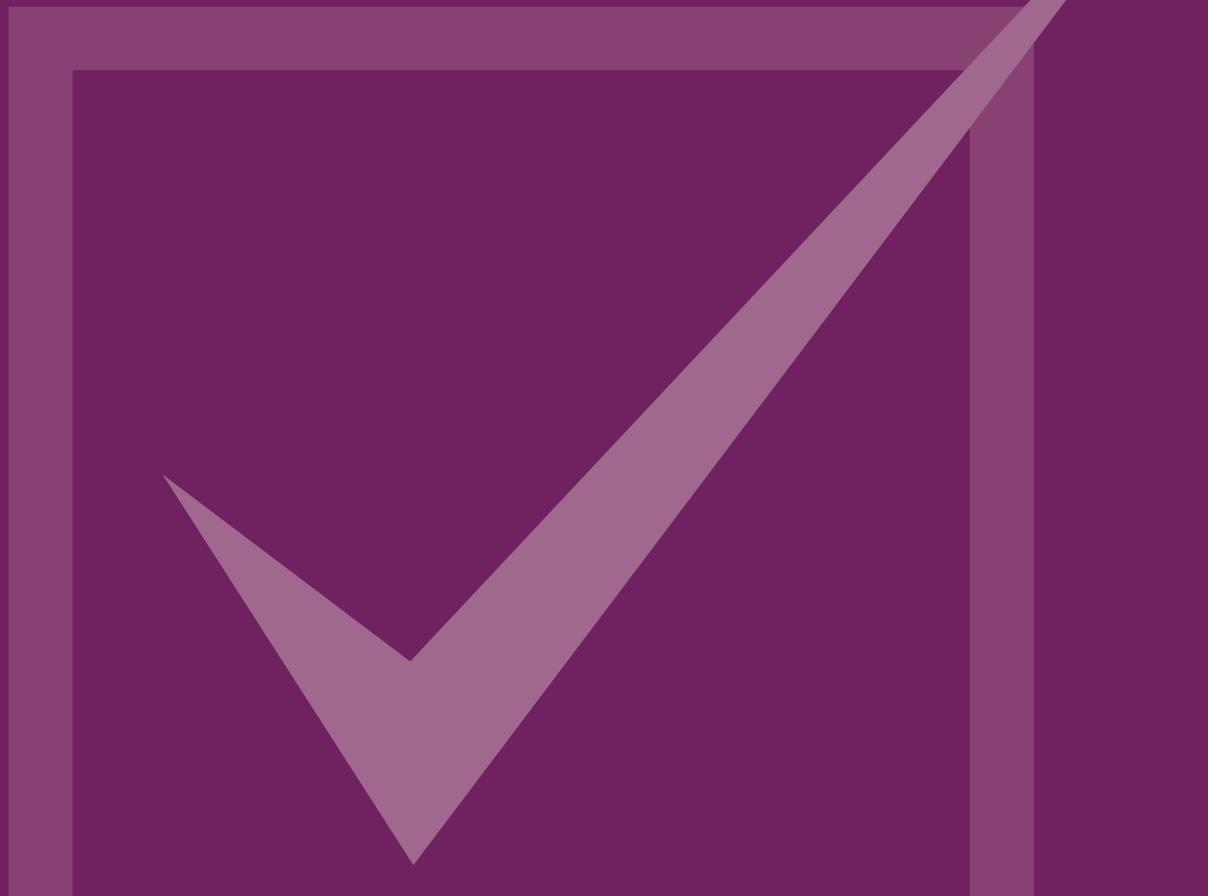**Information Management Office**

18

# Digitisation of Records

## Better Practice Checklist

**Practical guides for effective use of new technologies in Government**

# Introduction

The use of new technologies by Australian Government departments and agencies provides opportunities to deliver better services and information and streamline administrative procedures. Not only have cultural agencies such as libraries, archives and museums used new technologies to digitise their collections in order to provide their clients with access to these collections wherever they live or work, but other agencies also have digitised records and other materials to facilitate streamlined online processes.

A key role of the Australian Government Information Management Office (AGIMO) is to identify and promote 'Better Practice'. This checklist has been created to help agencies optimise their use of new technologies by effectively digitising materials. The items in the checklist are, however, not mandatory. The checklist focuses on the creation of digital versions of material, not the management of 'born digital' items.

This checklist has been created for staff responsible for managing processes to capture, copy, store or provide access to digital material. This checklist focuses on non-technical issues.

It should be noted that the checklist is not intended to be comprehensive. Rather, it highlights key issues for agencies. The checklist is iterative and draws on the expertise and experience of practitioners. The subject matter and issues are reviewed and updated to reflect developments.

Comments about this checklist and suggestions for additional checkpoints can be directed to:

Better Practices and IMSC Secretariat
Australian Government Information
Management Office
Phone: (02) 6271 1514
Email: **better.practice@agimo.gov.au**

Copies of this and other checklists are available at **www.agimo.gov.au/checklists**.

## Acknowledgments

This checklist was developed with the assistance of Australian Government agencies. In particular, AGIMO would like to thank Screensound Australia, the National Archives of Australia and the National Library of Australia.

# 18. DIGITISATION OF RECORDS

## Why digitise?

Creating and providing access to digital copies of material give an agency's clients and customers the potential to pursue educational, cultural appreciation and commercial opportunities wherever they live or work. Digitisation also enhances the potential for synergies with other digital collections through shared descriptive information in consolidated or federated online databases.

For audiovisual materials, digitisation minimises the generational loss inherent in older analogue and compressed digital audiovisual formats, and it reduces the costs of migration required to manage rapid changes in formats.

Note that this checklist focuses on the *digitisation* of print and other materials. The checklist does not directly cover issues associated with *imaging*, which involves the creation of an image of the item. While there are numerous differences in the processes, a key difference is that digitisation means that all the information within the item is available in electronic form and can be searched. Imaging provides a copy of the original, but elements within the item cannot be searched unless the copy is further processed by Optical Character Recognition (OCR) software.

# ✓ SUMMARY OF CHECKPOINTS

## BEFORE YOU START

- ☐ Identify the business case for digitisation
- ☐ Consider the digital content value chain
- ☐ Consider general principles for digital preservation
- ☐ Consider appropriate technical standards
- ☐ Consider the equipment to be used
- ☐ Consider resource allocation
- ☐ Consider rights and permissions management

## MANAGING DIGITAL ITEMS

- ☐ Apply metadata to the digital item
- ☐ Consider using a collection management system to manage digitised items
- ☐ Determine a delivery method for the information

## INFRASTRUCTURE ISSUES

- ☐ Estimate collection growth and storage requirements
- ☐ Develop a backup strategy and disaster management plan
- ☐ Future-proof the system

# ☑ CHECKPOINTS

## BEFORE YOU START

### ☐ Identify the business case for digitisation

A business case should be identified to ensure that key issues have been considered and that the decision to proceed is based on sound reasoning. Issues to be considered would include the required services, stakeholders' interests, resourcing, interoperability, risk assessment, privacy and recordkeeping requirements.

Further details on developing business cases for ICT projects can be obtained from the Australian National Audit Office's *Internet Delivery Decisions* Better Practice Guide **www.anao.gov.au**.

### ☐ Consider the digital content value chain

The digital content value chain represents the value that is added to digital content through key stages, from conception to use. Consideration of the value chain will inform the development of any business case.

There are numerous models for the digital content value chain. The following diagram provides a commonly understood model for digitisation programs in Australia.
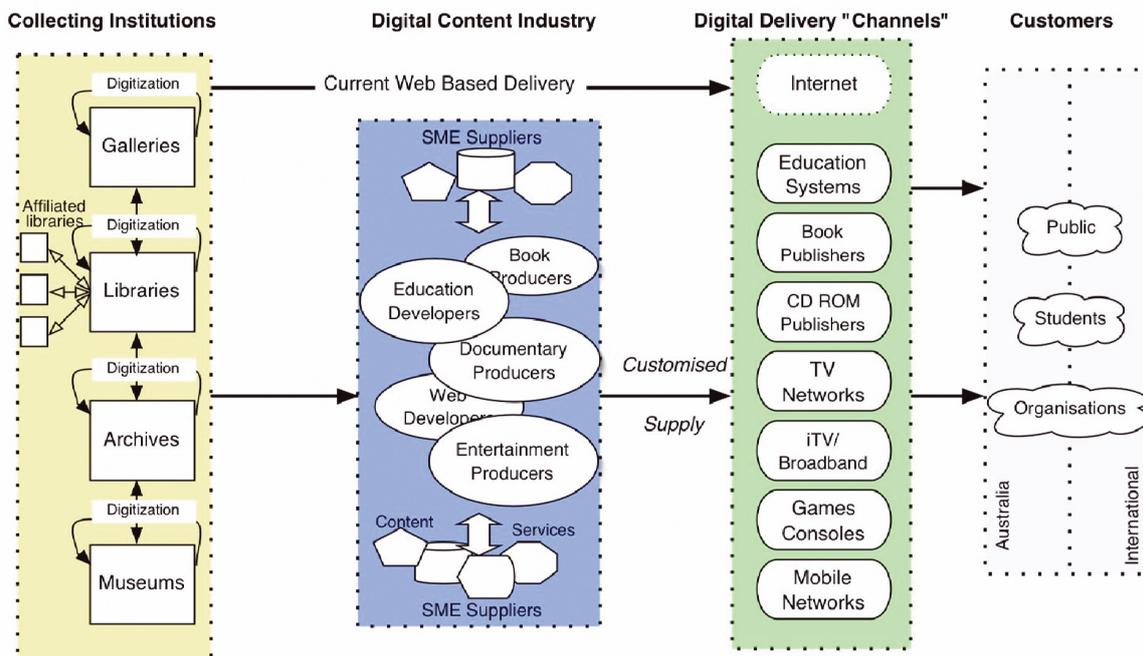


Figure 2.6.5 from Dr Roger Clarke and Peter Higgs. 'Economic value from cultural content: digitisation programs, standards and flow-on effects', p. 11. Creative Industry Cluster Study, 2003. © Commonwealth of Australia.
**www.cultureandrecreation.gov.au/cics/benefits.pdf**

# ☑ CHECKPOINTS

## ☐ Consider general principles for digital preservation

General principles that have been identified by agencies with experience in digitising materials for preservation purposes include:

- Treat the material being copied appropriately during the digitisation process. That is, ensure that the original materials are not damaged by the digitisation process.

- Digitise to a sufficiently high quality so that the material does not need to be redigitised later.

- Use a standardised format that is likely to remain available and supported for a long time and is likely to be easily migrated to successor formats without loss.

- Describe the digital files with suitable metadata to allow efficient and effective management over time.

- Store the files on carriers and file management systems that can be cost-effectively managed and that maintain data integrity.

- Implement proper backup, disaster preparedness and data management arrangements and audit them regularly.

- Consider that the means of providing access to digital files will change over time, necessitating action to keep them accessible.

- Plan to take the steps that will be needed to keep them accessible as authentic/trustworthy copies. The steps and any records generated by this process of ensuring or proving authenticity of the copies will be a record and will require coverage in the agency records disposal authority.

The International Association of Sound and Audiovisual Archives (IASA) has developed a set of principles covering the digital preservation of audiovisual materials. These principles cover the management of errors in digital copies, technological obsolescence, migration to different formats and preservation of originals. Further information is available from IASA **www.iasa-web.org**.

## ☐ Consider appropriate technical standards

Selecting the appropriate technical standards ensures the efficient acquisition and storage of collection items, reduces the risk of format obsolescence, improves audience reach, and allows the sharing and exchange of data across agencies.

While there are a number of technical standards for other formats, the technical standards for digitising audiovisual formats on page 7 are offered as a guide.

## ☐ Consider the equipment to be used

If the digitisation process is to be undertaken in-house, the selection of appropriate equipment will be a significant issue. Business cases, the agency environment and other considerations will determine the equipment to be used.

## ☐ Consider resource allocation

Apart from the resources needed to acquire content in digital form and/or to actually digitise existing items from the collection, agencies should consider the potential resource requirements for a range of supporting tasks. These tasks may include:

- choosing or prioritising the items to be digitised

- obtaining permission and rights from creators and stakeholders to provide access to digital images

- description or cataloguing and metadata management

- retrieval and return of the original material

- handling, preservation treatment and possible rehousing of the original material

- storage and backup of the digitised items

- content quality assurance

- website development

- IT infrastructure development

- ongoing management and maintenance.

## Technical standards for digitising audiovisual formats

| Media | Purpose | Format | Resolution | Other details |
|-------|---------|--------|------------|---------------|
| Audio | Preservation surrogate | BWF | 96 kHz 24bit | - 48 kH 24 bit from low-quality tapes<br><br>- Digital-born items keep their original format |
| | Browsing | MP3 | 44.1 kHz 16bit | |
| Still images | Preservation surrogate | TIFF | 300 dpi | - Colour: 24 bit<br><br>- Greyscale: 8 bit uncompressed |
| | Browsing | JPEG | 72 dpi | - Cropped and/or improved for online viewing if appropriate |
| Video | Preservation surrogate | .mxf<br>.601 or<br>.YUV | Uncompressed or original format | - MXF used for exchange purposes<br><br>- Soundtrack digitised as WAV file |
| | Broadcast or distribution | MPEG-2 | 50 Mb/s I-frame to 5 Mb/s long GOP | |
| | Browsing | MPEG-1 or MPEG-4 | 1.8 kb/s to 500 kb/s (typical) | |

### ☐ Consider rights and permissions management

The *Copyright Act 1968*, including the *Copyright Amendment (Digital Agenda) Act 2000*, may affect items held in agency collections. Agencies may be required to obtain permission prior to copying items or providing access to them. Agencies should consult the Acts to determine how these laws affect their collections and, if necessary, obtain legal advice. The Australian Copyright Council may be able to provide advice **www.copyright.org.au**.

Social, cultural and religious sensitivities may also need to be considered. For example, the copying of material relating to Indigenous groups may be unacceptable to some groups.

Some agencies may find it necessary to implement an electronic rights and permissions management system. Depending on the size and breadth of the collection, these systems can be expensive to implement and maintain.

# ☑ CHECKPOINTS

## MANAGING DIGITAL ITEMS

### ☐ Apply metadata to the digital item

Metadata is broadly defined as 'structured data about data'. Metadata about digital items may describe content to enable users to discover, locate and retrieve required items, but it can also (most significantly to audiovisual archiving) describe the format and technical nature of the work.

There are many standards governing metadata, depending on context. Dublin Core (DC), Australian Government Locator Service (AGLS) and Metadata Object Description Schema (MODS) are three metadata standards that can be used to describe resources. The attributes of the collection and the digital format may require modifications or additions to these schema.

Further information about the Australian Standard for Preservation Metadata is available from the National Library of Australia **www.nla.gov.au/preserve**.

Further information about general metadata issues is available in *Better Practice Checklist 6, Use of Metadata for Web Resources.*

### ☐ Consider using a collection management system to manage digitised items

A collection management system can facilitate the storage, location, retrieval control and access to collection items. Systems can range from a simple management system using a spreadsheet or listing to a fully integrated and automated database application.

Systems used can record physical and intellectual descriptive information, provide a unique number or code for each item, and track the movement of physical items. Other desirable functions can include acquisition, client and rights management.

Data that can be recorded in a collection management system might include title, copyright owner, file format, file size, primary source information, copy history and other factors. The collection management system is itself a record and will need coverage in an agency's recordkeeping system. In addition, if original records are disposed of, access copies need to be managed appropriately.

### ☐ Determine a delivery method for the information

Budget, infrastructure, collection and business drivers will determine how agencies may provide access to the digitised materials. Agencies may choose to provide access to the collection online via their own website or a collaborative repository, or via an FTP server or email, or offline by copying the item to a physical format, such as CD or DVD.

Agencies should also determine whether they will serve clients best by allowing access to all parts of the collection, or by presenting selections. Many agencies find it necessary to provide both forms of access.

## INFRASTRUCTURE ISSUES

### ☐ Estimate collection growth and storage requirements

Storage requirements can grow considerably due to digitisation programs. Collection growth and storage requirements will depend on many factors, including acquisition strategy, copying programs and technical specifications.

### ☐ Develop a backup strategy and disaster management plan

The creation of backup copies of digital items is essential to ensure the long-term preservation of the collection in case of systems failure or other disasters. Agencies should determine the level of risk they can afford and implement a policy to meet these requirements.

Agencies may find it appropriate to hold two backup copies for digital items. Each copy can be stored in a different location, in correct storage conditions. Suppliers should be able to advise on the best storage conditions for particular digital media. Regular backing-up of the collection management system will also assist in disaster recovery.

### ☐ Future-proof the system

Survival of digital content in the long term depends on having stable, sustainable file formats, and systems that will manage content through regular cycles of technological change. Key strategies to consider in future-proofing systems include:

- using formats that prevent loss of data quality (for example, TIFF rather than JPEG for the preservation surrogate copy)

- checking and validating data

- redundancy (that is, having more than one copy of data in more than one location)

- refreshing when individual media begin to deteriorate, before data is lost

- reformatting by cloning data onto newer formats and media

- migrating entire storage systems or pools of data as technology changes

- translating or transcoding content from formats that can't be supported

- emulating older technology with newer hardware and software.

# Other resources

Metadata Object Description Schema:
**www.loc.gov/standards/mods**

Dublin Core metadata schema:
**www.dublincore.org**

Open Archives Initiative:
**www.openarchives.org**

Association of Moving Image Archivists (AMIA):
**www.amianet.org**

International Association of Sound and
Audiovisual Archives: **www.iasa-web.org**

Society of Motion Picture and Television Engineers:
**www.smpte.com.au/eventcat.asp?cid=8**

Preserving Access to Digital Information (PADI):
**www.nla.gov.au/padi**

Guidelines for the preservation of digital heritage
(Unesco, March 2003): **www.unesdoc.unesco.
org/images/0013/001300/130071e.pdf**

Preservation Metadata for Digital Collections:
**www.nla.gov.au/preserve/pmeta.html**

DIGIPRAC-L Discussion List:
**www.nla.gov.au/nla/listserv/digiprac-l.html**

# Better Practice Checklists

**practical guides for effective use of new technologies in government**

**www.agimo.gov.au**