

# COLLECTING AND PRESERVING DIGITAL MATERIALS

A HOW-TO GUIDE FOR HISTORICAL SOCIETIES

BY SOPHIE SHILLING

# CONTENTS

- Foreword
- Preface
- 1 Introduction
- 2 Digital material creation
  - Born-digital materials
  - Digitisation
- 3 Project planning
  - Write a plan
  - Create a workflow
  - Policies and procedures
  - Funding
  - Getting everyone on-board
- 4 Select
  - Bitstream preservation
  - File formats
  - Image resolution
  - File naming conventions
- 5 Describe
  - Metadata
- 6 Ingest
  - Software
  - Digital storage
- 7 Access and outreach
  - Copyright
  - Culturally sensitive content
- 8 Community
- 9 Glossary
- Bibliography

# FOREWORD

How the collection and research landscape has changed!! In 2000 the Federation of Australian Historical Societies commissioned Bronwyn Wilson to prepare a training guide for historical societies on the collection of cultural materials. Its purpose was to advise societies on the need to gather and collect contemporary material of diverse types for the benefit of future generations of researchers. The material that she discussed was essentially in hard copy format, but under the heading of 'Electronic Media' Bronwyn included a discussion of video tape, audio tape and the internet.

Fast forward to 2018 and we inhabit a very different world because of the digital revolution. Today a very high proportion of the information generated in our technologically-driven society is created and distributed digitally, from emails to publications to images. Increasingly, collecting organisations are making their data available online, so that the modern researcher can achieve much by simply sitting at home on their computer and accessing information via services such as Trove and the increasing body of government and private material that is becoming available on the web.

This creates both challenges and opportunities for historical societies. To start, societies are increasingly digitising their collections for preservation purposes as well as to make them more readily accessible for both internal organisation and external access. These matters are partly discussed in this new training guide, and direction is given to a wider range of information and guidance.

A major focus of this guide is material that is born-digital, that is, created in digital format. A good deal of such material is ephemeral. For example, historical societies, like so many other groups, increasingly communicate with their members by email messages and distribute their newsletters in the same way. But how much of this is being kept or collected? At the same time, much of the information that future researchers will need is being created digitally, and unless someone consciously collects it, it may disappear.

So historical societies face a dual challenge in collecting born-digital material. One is preservation of their own records and publications, and the other is to collect the sort of material that future historians will be looking for when studying our communities.

The FAHS commissioned Sophie Shilling to examine the question of data that is born digital, and how societies can work within the digital world and can collect material for the future.

Don Garden

President

# PREFACE

This guide aims to assist Australian historical societies in preserving their digital collections. The author hopes to have filled a gap in the current literature surrounding digital preservation, that being a practical, plain English guide covering all aspects of curating and keeping digital collections.

Although digital preservation is a highly technical subject, the advice given in this guide demonstrates what is practical for small Australian historical societies. By following these guidelines, your newly created digital collections will have the best starting point to be preserved and accessible into the long-term. Any future digital preservation efforts will be made much easier.

The guide assumes a reasonable level of computer literacy and offers further advice for those with more knowledge. Several digital preservation solutions have been provided, accounting for differing levels of funding, experience, equipment, and knowledge.

Also included is how to write digital preservation plans and policies, how to foster positive change in a society, and risk management of digital collections. It is recommended that this guide is read through from start to finish before embarking on a digital preservation project.

# 1 INTRODUCTION

The volume of digital materials that we produce is staggering. In historical societies, these digital files could be the results of a **digitisation** project, or they could be **born-digital** (digital works which are not copies of analogue documents). They could be collection items, records of business, or communications. How can this wide variety of digital material be managed?

## Introduction

There are numerous benefits to **curating** a digital archive. Firstly, the collection will expand, either due to digitisation, or better management of valuable **born-digital** documents (of digital origin, for example a document produced in Word). Secondly, keeping documents that relate to the legacy of a historical society will enrich its history for future generations. You may also need to keep born-digital documents to comply with legal retention periods (see Digital Material Creation). If you're making collection items available digitally, improving access is a great way to increase your historical society's reputation and awareness of the collection – after all, the collection is less valuable if no one knows what you have.

This is a step-by-step guide to collecting and keeping a digital collection. The practical steps to preserving digital collections are described in four main steps: Select, Describe, Ingest, and Access, and these are supported by thorough descriptions of different types of digital content, how to plan your digital project, and a comprehensive glossary of terms.

This guide is split into sections based on the Digital Curation Centre's Curation Lifecycle Model,<sup>19</sup> which is a good model for data preservation, and follows the life of digital materials from creation to disposal. At the end of each section is a checklist to complete before moving on to the next section. The practical steps to preserving digital collections are described in four main steps: Select, Describe, Ingest, and Access; and these are supported by thorough descriptions of different types of digital content, project planning, and a comprehensive glossary of terms.

## 2 DIGITAL MATERIAL CREATION

Understanding the origins of digital materials is crucial in choosing how to preserve them. This section describes the differences between **born-digital** and **digitised** content. The wide range of style and **format** of born-digital materials requires best practice data management while still in use, and assessment for **long-term preservation** thereafter.

## BORN-DIGITAL MATERIALS

The way we produce information has changed. Audio and video recordings which were once on film and tape are now published to **stream** online. Manuscripts, which were once handwritten or typed pages and diaries are now Word documents and blogs. Correspondence is now more often via email than post. Instead of making physical photo albums, we can share photos on Instagram, Facebook, and Flickr. We create and share this content and others can modify it by liking, sharing, and commenting. These materials which have a completely digital origin are called **born-digital** materials.

### CURATING BORN-DIGITAL MATERIALS

As a historical society, you may have a document describing the kinds of materials you collect. You may need to update your collections policy if it is not broad enough to include born-digital materials. If you do not have a collections policy, now is the time to write one! Most collecting institutions make their collection policy visible to the public.

To go one step further and explicitly state the need to collect born-digital materials will set your society up to successfully **curate** a born-digital collection. Or, write a statement for your website to attract those who are seeking a suitable institution to donate their digital collections.

**Word processed** file types such as DOC and DOCX save a Microsoft Word document in its original format, but over time the appearance and content can age, change, and be subject to software **obsolescence**. By opening the document in Microsoft Word and choosing *Save As* in **PDF** or **PDF/A** format, you can be sure that the appearance of the document will not change. The choice to change the **format** depends on whether the content or the appearance of that document is more important. Letters, invitations, and newsletters – which are often included in historical collections – are usually produced nowadays on word processing software. These are excellent candidates for accessioning to our digital cultural collections. Treat these born-digital documents as you would their **analogue** equivalent and archive them. Most digital newsletters will be circulated in a PDF format which is suitable for preservation. Keep numbering of your newsletters consistent – you might like to use an acronym of the title or organisation, followed by the issue number or date in YYYYMMDD format (file naming is covered in more detail in the Select section).

**Social media** pages such as Facebook, Twitter, and Instagram belong to a subset of born-digital materials called **dynamic data**. After the account holder has posted content, other users can add likes or comments, which changes that content. Whilst the account holder owns the copyright to any original information they post (text, image, or video), the social media platform may have some rights to publish.<sup>24</sup> A user's profile from most social media platforms can be exported as a personal archive, in which all messages, status updates, and comments are downloaded as **HTML** files, packed with the images into a **zip file** and downloaded from a user's Account page. Similarly, a page admin can export a page's data as a personal archive.



WHEN COLLECTING NEWSLETTERS, YOU MIGHT LIKE TO USE AN ACRONYM OF THE ORGANISATION, FOLLOWED BY THE ISSUE NUMBER AS THE FILE NAME.



## Digital material creation

<input type="checkbox"/>	Name	Type	Compressed size	Password ...	Size	Ratio
	html	File folder				
	messages	File folder				
	photos	File folder				
	videos	File folder				
	index	Chrome HTML Document	6 KB	No	12 KB	55%

A FACEBOOK USER'S DOWNLOADED PERSONAL ARCHIVE. FILES ARE SAVED IN HTML AND CAN BE OPENED IN A BROWSER. WHILE IT WON'T LOOK LIKE FACEBOOK IT WILL HAVE THE INFORMATIONAL CONTENT OF THE USER'S PROFILE.

As well as correspondence, **email** accounts contain promotional materials, newsletters, notifications, invoices, bills, and spam. The National Library of Australia recommends using subfolders to sort emails of permanent value,<sup>42</sup> keeping in mind that business-related documents can be subject to legal retention periods. You might consider making subfolders in your inbox and routinely **back-up** emails of long-term value. To do this in an email account that is used in-browser (e.g. Gmail, as opposed to Microsoft Outlook, which is a software program) you will need to download as an archive, by going to Account Settings. Emails can be exported and saved as **EML** files.

Whilst **websites** are stored by your chosen website host, a local copy should also be retained and archived. Websites written in **XHTML** format are suitable for preservation or can be saved in a WebARChive format called **WARC**. Some web pages are periodically captured and archived, found on The Wayback machine (<https://archive.org/web/>), and Australian web publications are captured on Pandora, run by the national Library of Australia (<http://pandora.nla.gov.au/>).

The original format of any **digital art, photo** or **video** should be maintained, as well as a preservation friendly copy. Photographs from society events, exhibitions, landmark community happenings, and more are priceless to your historical group. Convert to a preservation friendly format (see Select), save, and back up in at least two other locations. Use a storage hierarchy model to save your photographs in series (more on that in Store). You may have a different **series** for each event, and one for any images shared on social media. You can use captions and comments from social media and event flyers to add context and richness to your collection.

**Active databases** should be captured periodically. Back-up databases in their original format or XML or CSV.

## DIGITAL RECORDS MANAGEMENT

Digital technology makes information creation very simple and as a result, we're creating and changing information at an unprecedented rate. A historical society might receive dozens of emails in a week – some of these will be important documents of long-term value, while others will be important now but can be discarded once actioned, and some will be of no use to the society at any time. Managing this large volume of digital documents is highly important. These documents must be managed when still in active use while planning and preparing for **ingest** to long-term storage. If managed well, locating these digital documents can be easier than locating physical ones. Instead of emailing a document, or printing copies and distributing by hand, simply create a shared folder on your network or in cloud storage, such as Google Drive. Not only does this reduce paper waste, there is also one master copy which anyone can access and change, if allowed.<sup>35</sup> If the document needs to remain unaltered, simply save as a read-only.

## Digital material creation

Historical societies can improve their digital records management processes by including digital business documents in their digital preservation plan (more on that in Project Planning). Your digital documents must be accessible (write down passwords and store securely), backed up, and the documents must be true and clear originals.

Business documents are subject to legal retention periods. Please note that the following list is a guideline – you should refer to the Australian Government's *Business*<sup>6</sup> website for legal requirements of recordkeeping.

Financial records, such as receipts, rosters, statements, asset registers, and taxation documents

Employee records, such as contracts, financial details, performance history

Policies, such as occupational health and safety, manuals, and procedures

As a rule, these documents should be retained for seven years. The Australian Taxation Office requires records to be retained for five years<sup>10</sup>, and the Australian Securities & Investments Commission<sup>8</sup> and the Fair Work Ombudsman<sup>25</sup> require a seven-year retention period.

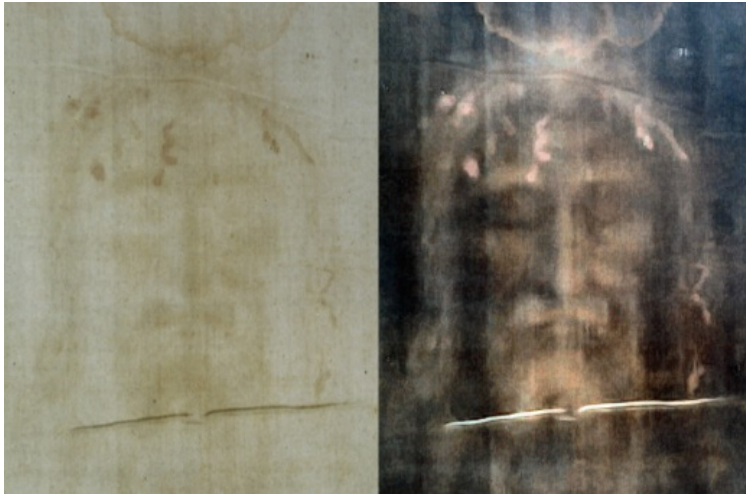
## DIGITISATION

There are many excellent how-to **digitisation** guides<sup>12, 27, 44, 47</sup> but this is not a how-to digitise guide – rather, it outlines what happens after digitisation, how can you preserve that content, and techniques that can enhance digital documents, uncovering crucial information from **analogue** originals.

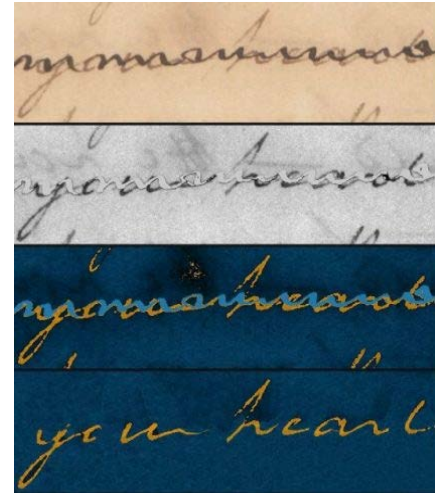
You may choose to digitise your analogue objects to create copies that will be used instead of the originals. This is beneficial for many reasons:

- Browsing can be much easier with digital images on a computer than physical objects, as the only equipment needed to view is a working computer – unlike viewing some analogue formats such as microforms or glass slides.
- Browsing digital copies also reduces access to the physical objects which keeps them better preserved, and the original order of the collection can be 'locked' in place – that is, items are less likely to be misfiled or misplaced.
- Most research needs can be fulfilled by a digital representation of an analogue object in the right system.
- High-resolution, lossless formats can fully capture an image or document in such detail that viewing the original will not provide any extra information.
- In some cases, digitisation can even enhance the originals.<sup>14</sup>

High resolution images can provide more detail than we can see, specialised photographic techniques can pull out layers of detail, and coloured light can be used to view details not visible due to age or damage. For example, a common image of the shroud of Turin is, in fact, a digitally altered photograph. More recently, the Library of Congress used different coloured light to uncover Alexander Hamilton's letters to his wife. The different spectrums of light could isolate the ink he used from the ink used by his son to cross out sections of writing.<sup>34</sup>



DIGITALLY ALTERING THIS PHOTOGRAPH OF THE SHROUD OF TURIN MAKES FACIAL FEATURES VISIBLE. CC BY-SA DIANELOS GEORGODIS.



LIGHT SPECTRUM ENHANCEMENT BY LIBRARY OF CONGRESS.

## DIGITISING YOUR COLLECTION

There are many useful guides to digitising your collection but what should you do after digitisation? You may know that you must keep at least three copies of each **file**, and you may know the best **file format** and resolution, but what about **digital preservation** of your digitised collection? And should it be treated any differently to your born-digital collection?

There may be some barriers to overcome before you can digitise your collection. You may think that you lack the knowledge, or you're overwhelmed by the scale and can't decide where to start. You may not have the equipment that you need. Start by using GLAM Peak's *Digital Access to Collections* modules which break the process into four manageable stages. Use the template below to begin planning your digitisation project.<sup>26</sup>

Digital material creation

<b>PLAN</b> Use this as the base of your Project Plan (see Project Planning)	What will be digitised?	
	How long will it take?	
	Who will do it?	
	How much will it cost? Digital Library Federation's <i>Digitisation Cost Calculator</i> calculates cost of labour and hours. <sup>20</sup>	
<b>PREPARE</b> Preparing documents for digitisation is an excellent opportunity for conservation assessment. Note any damage due to water, bugs, etc.	How will digitised copies be made? Museums Australia (Victoria)'s YouTube channel has many helpful guides.	
	Can you run a working bee to prepare items for digitisation?	
	What are the format, condition, and physical characteristics of the originals?	

Digital material creation

<b>DIGITISE</b>	<p>Capture the whole object. Keep digitised film negatives <i>and</i> positives; the silence at the beginning and end of tapes; the front, back, and spine of books.</p>	<p>What resolution is required for each type of object? Remember that colour is a carrier of information content.</p>	
		<p>Which items require extra care?</p>	
<b>SHARE</b>	<p>Use and share your digitised copies. they will not survive simply as “artefacts” of digital conversion (Conway, 2000).</p>	<p>Who will see your digitised items?</p>	
		<p>How will you present your digitised items?</p>	

## CHECKLIST

You now:

- know what born-digital materials are;
- know how to manage your digital records;
- have begun planning to digitally copy your collection for preservation.

# 3 PROJECT PLANNING

A major hurdle in carrying out a big project is working out where to start. Blindly jumping in with no plan is never a good idea. Equally, your project will be postponed if there is no clear path. By the end of this section, you'll have assessed what you have, what you need, written your digital preservation plan and/or policy, and created a workflow for processing your digital collection.

## WRITE A PLAN

Historical societies often find that new initiatives and projects happen in fits and starts. An energetic and motivated member or two will have a wonderful idea, start to make changes, then lose interest, run into hurdles, or leave. Avoid this in your **digital preservation** efforts by writing a plan. It must be *your* plan. It must be for *your* historical group. It must be based on *your* needs and *your* resources, which will be different to the needs and resources of another historical society down the road. It also must strive for what is practical, not what is perfect. Include the following sections in your Digital Preservation Plan. <sup>16, 27, 48</sup>

### DIGITAL PRESERVATION PLAN

#### PURPOSE

Clearly and explicitly describe why digital preservation is necessary for your organisation.

#### INSTITUTIONAL SETTING

Clearly explain how this initiative fits in the context of the collection and your business practices.

#### SCOPE

This plan can be for your whole digital collection, or you may wish to start with a smaller collection, such as digitised images from a recent exhibition, or your society's emails and other **born-digital** documents.

#### ROLES AND RESPONSIBILITIES

Who is responsible for accessioning? Who is responsible for storage and **back-ups**? Who is responsible for **access**?

#### COLLECTION DESCRIPTION

Information about the objects in the collection included in the scope of this plan. This can include collection items, as well as digital records created in the carrying out of business. List every type of document you have (for example photographs, essays, invoices, membership forms, completed membership forms, EFTPOS receipts, etc.) and for each type identify the creator, location, retention period, and format.

#### RESOURCES YOU HAVE

Write down a list of everything you have. Computers and their operating systems and if they connect to the internet, scanners, cameras, physical storage, cloud storage accounts, etc.

#### REQUIREMENTS FOR PRESERVATION

This includes resources like **hardware** and **software**. Now that you've identified what you have, you can clearly see if you can make do, or if you need anything else. How will you attain this equipment? Think outside the box – perhaps you could ask another nearby historical society if you could borrow their scanner? Or your local library, local government, schools, museums, etc. Preservation requirements also include software and access rights (see Ingest and Access and Outreach).

#### EXPECTED COSTS

Such as outsourced **digitisation**, purchase of equipment, software licenses, cloud storage fees. Also factor in data recovery, in case of data loss. This is an opportunity to investigate grant funding.

#### TIMEFRAME

This measurable goal (or goals) will motivate you and others to complete the project.

#### RISKS

Expected risks and mitigation strategies.

## POLICIES AND PROCEDURES

Good policies result in good decisions. This is a crucial element of succession planning and will assist in the continuity of the project with new users. A policy is often a public-facing document, and as such may increase your chances of funding if you are applying for grants. Use the following policy template<sup>11, 45</sup>. You should also find or write procedure documents for any software and hardware you use, to ensure that everyone involved knows the processes, and to make it easier to train new users of your system.

### DIGITAL PRESERVATION POLICY

#### PURPOSE

Clearly define why your organisation requires a digital preservation initiative. Specify whether **digitisation** is for preservation and/or **access**.

#### CONTEXT

Describe the organisation's background, and how this initiative will relate to and complement organisational objectives.

#### SCOPE

Does this initiative relate to the entire collection or a small number of items? Be specific. What are the priority items? What factors define a document for long-term storage?

#### CHALLENGES

Identify challenges and risks, such as software and hardware obsolescence, growth of the collection, compliance with copyright and existing policies.

#### PRINCIPLES

Key digital preservation concepts, reference models, values, and philosophy.

#### ROLES AND RESPONSIBILITIES

Who is going to be responsible for what parts of digital preservation? Which positions?

#### STANDARDS

for example OAIS, ISAD(G), internal documents such as technology standards.

#### COMMUNICATION

Regular meetings, education, outreach.

#### REVIEW CYCLE

Be specific. Technology policies need to be reviewed regularly.

#### ACCESS AND USE

Broadly, will there be access restrictions?



## RELATED POLICIES

Collections Management Policy, Archives and Records Access Policy, etc.

## GLOSSARY

Necessary due to the technical nature of digital preservation.

## FUNDING

Grant funding applications usually open at the start of the calendar year. GLAM Peak's *Digital Access to Collections*<sup>27</sup> guide provides excellent advice for a successful grant application:

Convey your urgency.

Partner with another historical society or group of historical societies, or even another organisation. For example, your society may wish to share hardware such as a specialty scanner.

Start with a small project, for example, one digital exhibition and build from there.

View successful grant applications or find someone who can review your application before submitting.

## GETTING EVERYONE ON-BOARD

Any change is likely to attract resistance. Whether this is due to fear of the unknown, disruption to routines, lack of confidence, loss of control, poor timing, or lack of purpose, you can mitigate the risk of resistance to change by:

communicating a sense of urgency;  
establishing a team to ensure continuity;  
communicating a clear vision through your plan;  
creating short-term goals, and  
starting small and building on success.<sup>46</sup>

## CHECKLIST

You now:

- have written a digital preservation plan;
- have written and published a digital preservation policy;
- have applied for grant funding;
- are equipped to deal with resistance.

## 4 SELECT

This section focuses on preparing the individual **files** for preservation. By the end of this section you will know which **file formats** and image **resolution** are suitable for your preservation needs and will have created a file naming system for your collection. Then, your **digital objects** will be ready to describe.

## BITSTREAM PRESERVATION

Digital files are made up of millions of **bits** (zeroes and ones). Every time a file is opened and saved, or converted to a different file type, the bits change. Think of it like handling a piece of paper over and over again – the pages will stain gradually from the oils on your hands, you might accidentally crease, fold, crumple, or tear the paper. By saving the original sequence of bits in a digital file, it will be possible to go back to this file at any point in the future and be able to access and open the file. This is called **bitstream preservation**.

Bitstream preservation can be applied to any type of digital object, however it is not always necessary. Put simply, it comes down to whether the intellectual content or the integrity of the file is more important. For example, the bitstream of a digital artwork should be preserved to maintain the integrity of the artist's work and the software to view that file should be described in the **metadata** of that item to view it as the artist intended. However, a business document might not require the same treatment, as the content of the text is more important than the file format or appearance of the document.

Bitstream preservation is also important if a historical society receives a donation of a personal digital archive. This is because the way the person has arranged their digital records and the format of those records is just as important as if they were analogue documents. Manual actions to preserve the bitstream of a digital collection include:

1. Save materials on a computer that isn't connected to the internet
2. Copy files from the **carrier** as a disk image to retain metadata
3. Save the disk directory information (file names, sizes, formats) with the files
4. Place write blockers or restricted access on the folders (right click and choose "Give access to")
5. Write down every action taken in a ReadMe file<sup>23</sup>

This may not be a realistic option for organisations with limited funds or skills. At the very least, the following actions should be taken to preserve the content.

- Save in the original format, plus make an **open format** copy
- Don't make any changes to the item that will inhibit future use
- Document everything you do in the metadata

## FILE FORMATS

**File formats** dictate the software that may be used to view content. For example, a VCR tape is a format of video, which cannot be played using a reel-to-reel player. The same goes for digital formats: a DOCX (Word Document) cannot be opened using Microsoft Excel.

While the seemingly endless array of file formats may seem overwhelming, there are only a limited number that are preservation friendly. Any format that is **open source** will carry less risk of technological **obsolescence**. Open source means that the file type is not reliant on one software package to view it. File format obsolescence occurs when the proprietary software that creates a type of file falls out of common use. To mitigate this risk, migrate old formats when necessary (*Save As* another format).

The most important factor for file formats is consistency. The **ingest**, management, and migration of your collection if necessary will be easier if the collection is uniform. When considering migration, think about whether the format or the content is more important. If content is more important, migration to a new format will be beneficial, however if it is an artwork or an archival document, keep the original and make a copy in an updated format. If you're accepting donations of digital objects, specify the preferred file format for your collection.

Another aspect of file format to consider is the amount of disk space that will be needed because some file types are larger than others (for example TIFFs are a large image file, while JPEGs are compressed and take up less disk space). As a rule, the preservation copy of a digital file (particularly images) should be of the highest resolution your system can sustain. This ensures that you have a good quality copy of the item, should anything happen to the original. The following table provides a list of file types suitable for preservation.<sup>38-40</sup> This is not a comprehensive list but contains common file formats that are obtainable for small historical societies.

---

*A **master copy** is the original digital object. You might hear the master copy referred to as a **preservation copy**. An **access copy** is derived from the master copy, and is a smaller file.*

---

Text and documents			
Comma-separated values	.csv	Preservation friendly	Text file in a spreadsheet format
Microsoft Word document	.doc, .docx	Somewhat preservation friendly	.doc is at risk of becoming obsolete. Convert both .doc and .docx to PDF/A if possible.
PDF/A	.pdf	Preservation friendly	For best results, text files should be saved in PDF/A
Images			
JPEG	.jp2	Not preservation friendly	JPEGs are compressed. Use as access copy. Convert to TIFF if possible.
Tagged Image File Format	.tif, .tiff	Preservation friendly	The most commonly accepted format for images. Save TIFFs with no <b>compression</b> .
Audio			
MP3	.mp3	Somewhat preservation friendly	MP3 is a compressed audio file type. Convert to .wav if possible.
FLAC	.flac	Somewhat preservation friendly	.flac files use lossless compression. This is a good alternative to .wav if you're concerned about file size.
Wave	.wav	Preservation friendly	This format is so widely used that it has become the most preservation friendly. Files can be large due to the uncompressed format.
Video			
Motion JPEG 2000	.mj2, .mjp2	Somewhat preservation friendly	This <b>open standard</b> file format is under consideration as a digital archive format.

Select

<b>MPEG-2</b>	.mp2	Not preservation friendly	MPEG-2 is the precursor to Motion JPEG 2000. Convert if possible.
<b>MPEG-4</b>	.mp4	Somewhat preservation friendly	This file type is suitable for <b>digital preservation</b> .

Non-lossless or **Lossy** file types compress part of the file to save storage space. This is possible by removing random bits of the file, which maintains the overall content of the original. Repeated lossy compression will cause a file to deteriorate. Alternatively, **lossless compression** can decrease the file size whilst maintaining the integrity of the original. TIFFs are a type of image file that supports lossless data, while JPEGs are a lossy file format. JPEGs are very useful for an **access copy** or for display on a website, but TIFFs are much better suited for preservation.

## IMAGE RESOLUTION

As a rule, digitised collection items should have two copies in the highest possible resolution (the **master copy**). In addition to this, smaller objects require a higher resolution to be able to view detail without pixilation.<sup>27</sup> Don't be concerned with pixel dimensions: the resolution is what is important, as it will remain the same regardless of the dimensions of the image.

Material type	File use	Resolution
<b>Printed text (as an image)</b>	Master	300ppi
	Access	300ppi
	Thumbnail	72ppi
<b>Photograph</b>	Master	300ppi if larger than A4 600ppi if A4-A6 1200ppi if smaller than A6
	Access	300ppi if larger than A4 600ppi if A4-A6 1200ppi if smaller than A6
	Thumbnail	72ppi
<b>Map</b>	Master	Maximum allowable
	Access	600ppi
	Thumbnail	72ppi
<b>Newspaper</b>	Master	300ppi
	Access	300ppi
	Thumbnail	72ppi
<b>Object/artwork</b>	Master	Maximum allowable
	Access	600ppi
	Thumbnail	72ppi

## FILE NAMING CONVENTIONS

**File** naming will differ between organisations, however, there are some file naming conventions that should be followed. In addition to the following guidelines, ensure file names are consistent across

Select

the collection. An easy way to link the digitised copy to its **analogue** original is to use the **item number** somewhere in the file name.

Keep file names short, but meaningful.

Avoid repetition in file paths.

If numbers are used, ensure they are the same length, i.e. add zeros at the start of the number if necessary.

Don't use common words at the start of the file name, e.g. 'letter', 'journal'.

Avoid special characters in file names.

For peoples' names, use the full surname and initials. It doesn't matter whether the initials come before or after the surname, as long as it is consistent throughout the collection.

Use dashes (-) or underscores (\_) to denote spaces to avoid ambiguity in machine-read functions.

For example, one way to name a file might use the ISAD(G) standard, such as AU\_WWHS\_MS\_0027.005. Here is how the file name is broken down:

International Standard ISAD(G)	What is being described	How it is represented in the file name
<b>Country code</b>	Australia	AU
<b>Repository code</b>	Woop-Woop Historical Society	WWHS
<b>Collection</b>	Manuscript	MS
<b>Series</b>	Series 27: The Dumbledore Papers	0027
<b>Item</b>	Item 5	005

Or, you might prefer to use whole words in your file names, like MS\_Dumbledore\_005. This is acceptable, as long as the file names are consistent across the collection. Write down your file naming key so that your collection is named consistently.

## CHECKLIST

You now:

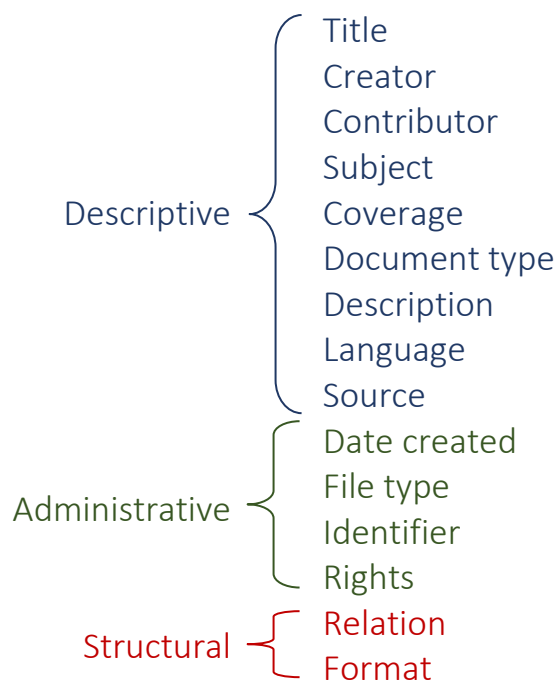
- understand when to preserve the bitstream;
- know which file formats are suitable for preservation;
- know suitable image resolutions for your collection items;
- have created your file naming conventions.

# 5 DESCRIBE

**Metadata** is crucial in keeping digital collections. By the end of this section you will know how to describe (add metadata to) your objects and collections in a way that will make them readable for generations to come.

## METADATA

The best way to make a file discoverable is to record its metadata. The word metadata means ‘data about data’ and contains much of the same information as a library catalogue entry. There are three categories of metadata which describe the minimum metadata required for preservation:



Create a guide to generating metadata for your society. This ensures consistency and will make preservation, management, and your exit strategy much easier. You can simply store metadata in a spreadsheet if you don't have software that can save metadata, or just use your existing catalogue! Here you can see that a lot of metadata is the same as what is in a catalogue record; all that is missing is the technical metadata about the digital file.

<b>Title</b>	Jasper Jones : a novel / Craig Silvey.	Title="Jasper Jones"
<b>Author</b>	Silvey, Craig, 1982-, (author.)	Creator="Silvey, Craig"
<b>Published</b>	Crows Nest, N.S.W. : Allen & Unwin, 2010.	Publisher="Allen & Unwin"
<b>Copyright</b>	©2009.	Rights="Copyright 2009"
<b>Content Types</b>	text	Type="Text"
<b>Carrier Types</b>	volume	Format="pdf"
<b>Physical Description</b>	397 pages ; 20 cm.	Subject="Racism"
<b>Series</b>	Premiers' Reading Challenge Years 9 and 10	Subject="Bookclub collection"
<b>Subjects</b>	Racism -- Fiction. Bookclub collection. Australian fiction -- 21st century. Friendship -- Juvenile fiction. Secrecy -- Juvenile fiction. Australian fiction. City and town life. Secrecy. Friendship. Secrecy -- Fiction. Friendship -- Fiction.	Subject="Australian fiction --21 <sup>st</sup> century"

COMPARISON OF TROVE CATALOGUE ENTRY (LEFT) AND DUBLIN CORE METADATA (RIGHT).



## Describe

The elements of a metadata set will depend on the **metadata standard**, of which there are many. In cultural history collections, the most common metadata standard is **Dublin Core**, and this is what Trove uses – so if you plan for your collection to be **harvested** by Trove, try to use Dublin Core (see Access and Outreach for more on this). Dublin Core allows for a mix of media, allowing for the description of books, photographs, manuscripts, and **audio-visual** materials.

Usually, archival software is only compatible with a handful of metadata standards, so you won't have many to choose from. The ones that you're most likely to see in your software are Dublin Core,<sup>22</sup> **ISAD(G)**,<sup>29</sup> and **LIDO**,<sup>30</sup> which all have their own guidelines.

## CHECKLIST

You now:

- understand how to create metadata
- have chosen which metadata standard you will use

## 6 INGEST

**Ingest** is the stage in which your **curated** digital items are moved into your digital storage. In this section you will choose the type of software you need, the amount and type of digital storage that is best for you, and how to arrange your collections in a **storage hierarchy**, to ensure that your digital collections are stable and **accessible** into the long-term.

## SOFTWARE

It is not possible to recommend a ‘one size fits all’ for digital collection software as it will depend on your needs. Before choosing software, identify what you need it to be capable of. Perhaps one software package will do everything you need, or perhaps you will need two or three different applications. While digital collections programs vary in their capabilities, they can be defined in broad categories, as follows.

### THE BARE BONES REPOSITORY

It is possible to preserve your digital materials without specialty software. This option is more time consuming but is the cheapest. For this option, your written policies and procedures (see Project Planning) will have to be thorough.

Follow these steps:

1. **Ingest** (could be as simple as a copy and paste)
2. Add metadata (simply store this in a spreadsheet, like a database)
3. Store (could be a server, or an external hard drive. Don’t forget to **backup** offsite!)
4. Preserve (save metadata in a spreadsheet, and monitor your collection by making it accessible)
5. Access (online or offline, or onsite only)

### COLLECTION MANAGEMENT SOFTWARE

At the most basic level, it is possible to manage digital files using data management software. Please note that this is not **repository** software – just a metadata management tool. Some of these applications allow a small amount of storage for low-resolution photographs. If you choose this software, you will need to put access permissions on your master file storage folders and back up in at least two other locations regularly. Applications such as Microsoft Access, eHive, or similar are perfect for collection management.

### OPEN SOURCE DIGITAL PRESERVATION AND ACCESS SOFTWARE

While **open source** software is free, it will incur costs for any IT support required for installation, meaning this option could have high initial costs, but very low running costs. These programs usually have very good online support communities, like Artefactual’s Google Group which has regular posts from the creators and users. Programs like AtoM and Omeka are constantly being developed and new releases and patches are frequent.

### PROPRIETARY DIGITAL PRESERVATION SOFTWARE

Proprietary software is a low-risk option, and usually quite user-friendly. Proprietary software does tend to be a costly option, as licenses need to be purchased and updated.

Once you have chosen your software, you will need to learn how to use it. If you’re already computer savvy, you may be able to get by with the software documentation and online tutorials (you can usually find them on YouTube). If you’ve chosen an **open source** application, online forums are excellent for troubleshooting. Proprietary software companies provide support, usually at a fee unless it is included in the software license. There are classes and workshops for some programs – Victorian Collections, for example, runs free workshops regularly to teach their program to new users.

Software **obsolescence** can be avoided by refreshing your software to the most recent updates. Routinely audit your software to ensure not only that it is up-to-date, but also to ensure that you are

getting the most out of it. Stay abreast of new software innovations and move to new software if it will be in your collection's best interests.

## DIGITAL STORAGE

The lifetime of storage mediums usually exceeds the operable time of the medium, due to software obsolescence. Digital storage was once commonly on floppy discs, then compact discs, but has now moved on to flash drives, hard drives, or SSD hard drives, and cloud, and these will no doubt be superseded by newer technology. Avoid running into problems by having routine check-ups on your digital collections and migrate onto new hardware as necessary. While many **carriers** will last decades, the equipment that is required to view the content might only be in common use for a limited number of years<sup>2</sup>.

Based on the size of your collection and the quality of the images you would like to archive, calculate how much storage you will need. File size depends on image resolution and image dimensions, which are discussed in more detail in the next section. For calculating storage requirements, assume that a preservation copy will be approximately 60 megabytes.

$$\text{Number of images} \times 60\text{MB} = \text{required disk space}$$

For example, a collection of 5,000 images multiplied by 60MB = 300 GB, or 0.3 TB. Regardless of which combination of digital storage you choose, don't forget to allow for growth in your collection.

Hardware faults are best avoided by ensuring your **backups** are regular (frequency depends on how often content is added or changed) and diverse. Backup your digital archive on at least three different storage formats in different locations. Make at least two of these of the highest quality (**master copies** or **preservation copies**). You might have master copies stored on a computer and in cloud storage, and back-up using a hard drive. Choose at least three of the following storage options:

Desktop computer/laptop

Server

External hard drives or other media

Remote storage

Home computer

Tablets/smartphones

Cloud (For cloud you will need a reliable internet connection, which is particularly relevant for regional historical societies. Also note that some cloud storage services change ownership of your data. Always read the user agreement.)

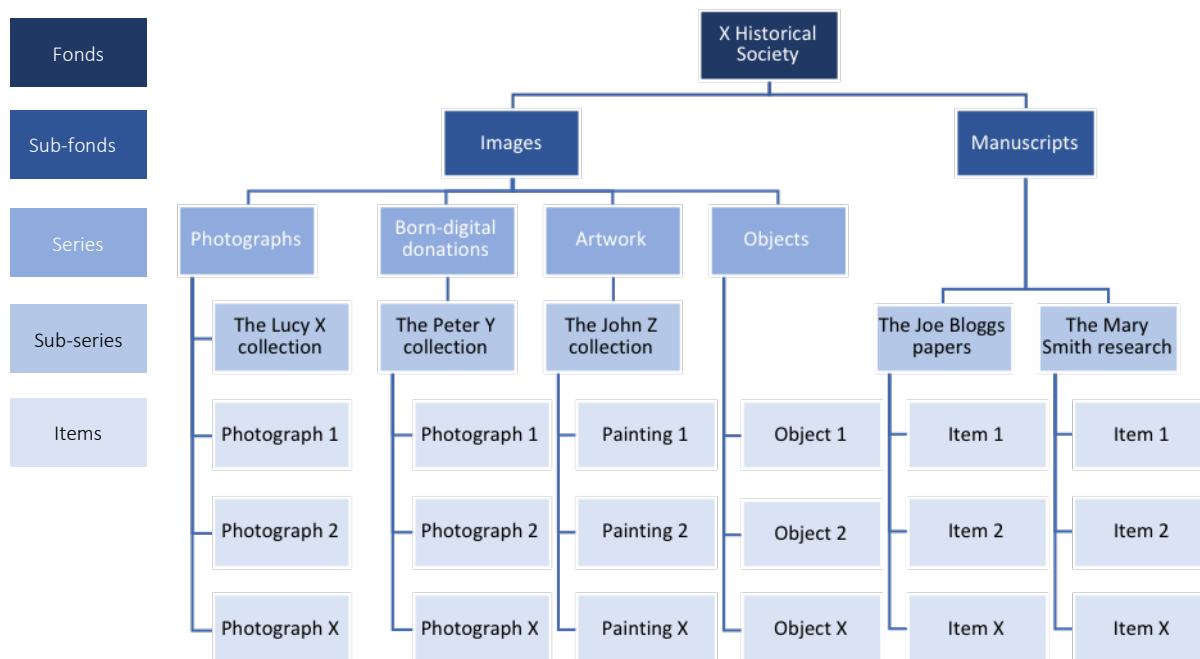
## STORAGE HIERARCHY

Before you begin saving your digital collection, consider the structure of the folders. Do you need to separate your own collection from donated digital copies? Do you need to separate your photographs from your artworks?

Folder structures, where possible, should mirror the way the physical items are stored, without being tied to a physical location. For example, if all the images are separated by **analogue** format (i.e. glass slides stored together, photographic negatives stored together) then the digitised versions should be stored the same way. However, if a sub-collection is stored in "Shelf A", don't store the digital

## Ingest

versions the same way! What happens if your physical storage is modified and the collection is moved? It may be helpful to create a diagram of your storage hierarchy.



EXAMPLE STORAGE HIERARCHY OF A HISTORICAL SOCIETY COLLECTION. THE LEVELS OF THE HIERARCHY ARE DESCRIBED IN THE GLOSSARY.

## CREATE A WORKFLOW

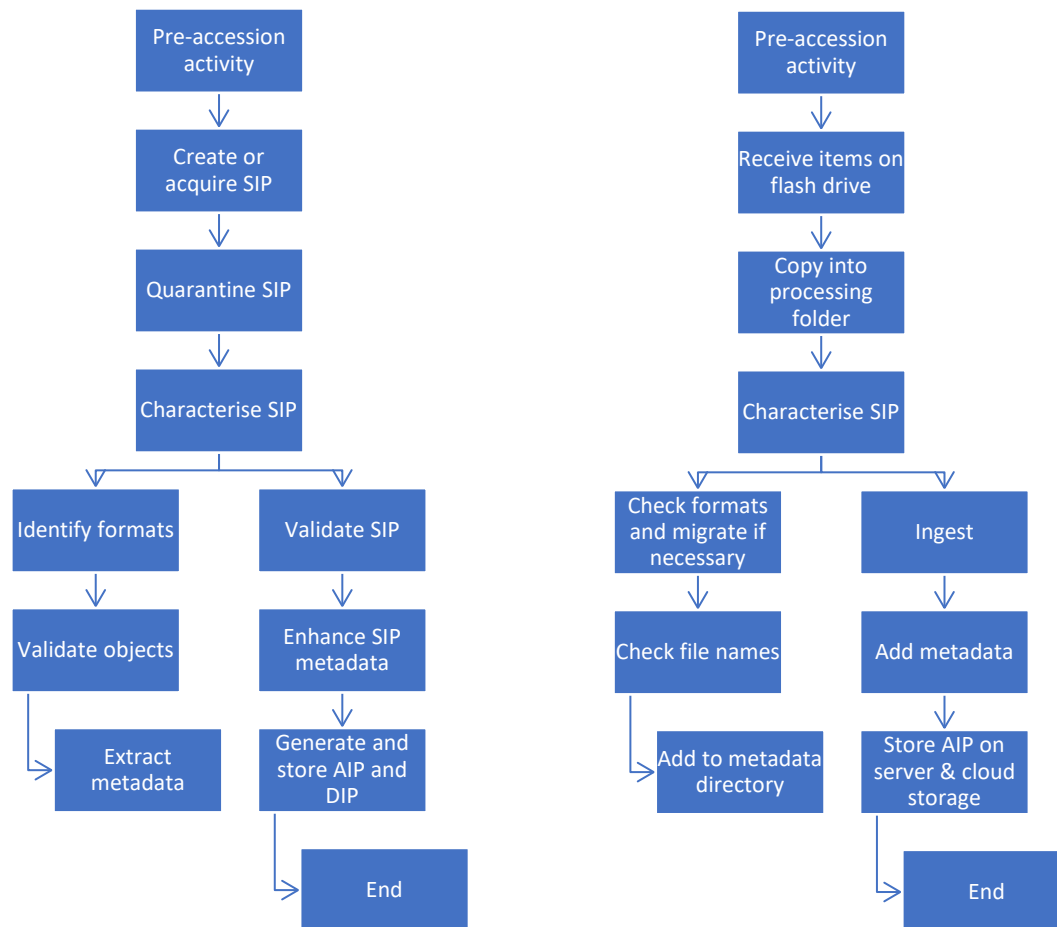
Whether your digital preservation program is simple or requires many steps, it is recommended that a workflow is established. This ensures that every step is carried out every time. Below, you can see the elements that are necessary for a digital preservation workflow, and an example of a real workflow from the Royal Historical Society of Victoria.

Let's break this workflow down in plain English. When preparing digital data, three "packages" will be created. These packages are theoretical – what is in these packages does not need to be saved in a single folder. Thinking of them as packages just ensures that we're checking that all the necessary data exists. The three packages are Submission Information Package (**SIP**), Dissemination Information Package (**DIP**) and Archival Information Package (**AIP**). They have the same information but do different things. To illustrate this, we're going to use the example of flat-pack furniture.

In the showroom, there is a cabinet. This is the AIP: the master copy, which stays in the showroom. Anyone who wants to buy that cabinet must take an exact replica as a flat-pack, and instructions. This is the SIP: the data, and how that data fits together (i.e. its metadata). That flat-pack can be taken home and put together. This becomes the DIP: the copy.

## Ingest

At RHSV, a digitisation company supplies a flash drive with the digital copies. These files are copied into a processing folder on the server, checked for item numbers, titles, and file format. They are transferred to the repository by the software Archivematica, which adds the files' metadata to the directory. It automatically stores the AIP, and a DIP is uploaded to the access software AtoM.



ELEMENTS OF A DIGITAL PRESERVATION WORKFLOW.<sup>11</sup>

EXAMPLE WORKFLOW FROM ROYAL HISTORICAL SOCIETY OF VICTORIA.

## CHECKLIST

You now:

- have chosen your software
- have three locations to store your collection
- have created a storage hierarchy
- have written a workflow to ensure consistency
- have established a routine check schedule to mitigate the risk of software obsolescence

# 7 ACCESS AND OUTREACH

A collection's value increases with awareness of it. This section will assist you in making your collection visible and fully or partially **accessible** if you choose.

## Access and outreach

Outreach doesn't have to mean complete public **access** to your collections. It is simply about increasing awareness. There is less value in your collection if no one knows about it – promoting your collection will increase interest in your society. You may be able to provide **thumbnails** or **watermarked** images of your collection to limit reuse. This will encourage more purchase requests for your images!

Preserving digital content is intrinsically linked with enabling access to it.<sup>37</sup> Think of it like a manuscript collection, kept in drawers. If no one accesses it for ten years, what are you likely to find when you finally open those drawers? Faded paper? Moth-eaten documents? **Digital objects** (like **analogue** objects) age – the best way to ensure that they are in good condition is to view them.

One of the best ways to increase awareness of your collection is to have it harvested by **Trove**. Trove is a search aggregator, which means that a user searching in Trove for any topic will pull results from all of Trove's sources. The only metadata **elements** required by Trove are title, **unique identifier**, and format. For more information on having Trove harvest your catalogue, see the Trove website.<sup>43</sup>

If you plan to sell digital copies of your collection items, you might display low-resolution or watermarked images on your society's website, Flickr, or eHive. Then, your high-quality images can be supplied upon request, for a fee.

Try to find ways to integrate your digital collection with the rest of your business practices and write down your outreach plans and policies. Writing down your policies formalises them, increases the chances of their survival, and is a great way to prove your eligibility for grant funding.

## COPYRIGHT

Before you share your images, ensure that your organisation holds the **copyright**. You don't have to apply for copyright – simply stating that you hold copyright of your collection is enough. A statement for fair use could be something as simple as "© X Historical Society. All rights reserved."<sup>4</sup>

If you don't hold copyright, you will need to obtain a license from the copyright holder/s to copy or publish the work. If you gain permission, be sure to get it in writing. The copyright owner holds the right to request a fee for their permission, or to refuse. The copyright owner of a digital **file** should be included in the object's **metadata**. This ensures that copyright is clear for new users in the future.

## SECTION 200AB FLEXIBLE DEALING

This section of Australian copyright law is commonly used by small collecting institutions. It provides a way to give access to content that has ambiguous authorship or is owned by another party. This is a complicated exception to Australian copyright law, so you should consider legal advice for its use.<sup>5</sup>

To qualify under s200AB, use must:

- not be for commercial advantage or profit; AND
- be for the purpose of maintaining or operating educational services or;
- be for giving educational instruction.

The use of copyrighted material must also be for a special case that does not:

- conflict with a normal exploitation of the material; or
- unreasonably prejudice the legitimate interests of the copyright owner.



## HOW LONG DOES COPYRIGHT LAST?

As of 1<sup>st</sup> January 2019:

Type of work	Length of copyright
Literary, dramatic, musical and artistic work with a known author	Life of the author plus 70 years
Literary, dramatic, musical, and artistic work with an author that cannot be identified (including anonymous, pseudonymous and orphan works)	If made public by the copyright owner before 1 January 2019 or within 50 years of creation: 70 years from when first made public If not made public in the above periods: 70 years from creation
Sound recordings and cinematographic films	If made public by the copyright owner before 1 January 2019 or within 50 years of creation: 70 years from when first made public If not made public in the above periods: 70 years from creation
Works, sound recordings, and cinematographic films created or first published by Commonwealth, State or Territory Governments	Creation plus 50 years

For further information:

Australian Libraries Copyright Committee: <http://libcopyright.org.au/><sup>7</sup>

Copyright Agency: <https://www.copyright.com.au/><sup>15</sup>

Australian Copyright Council: <https://www.copyright.org.au/><sup>4</sup>

Copyright: <https://www.communications.gov.au/what-we-do/copyright/><sup>18</sup>

Copyright Act 1968<sup>13</sup>

## LICENSING

If you would like to allow your collection to be open access (i.e. you want to give people the right to share and use your collection), you may attribute a Creative Commons license. More information can be found on the Creative Commons website, which includes a dedicated **GLAM** (Galleries, Libraries, Archives, Museums) section (<https://creativecommons.org.au/learn/glam/>). There are different types of Creative Commons licenses – they are sure to have one that fulfils your specific needs.<sup>17</sup>

## CULTURALLY SENSITIVE CONTENT

AIATSIS (Australian Institute of Aboriginal and Torres Strait Islander Studies) is the guiding body for indigenous studies in Australia. Their website contains a plethora of resources for managing culturally sensitive materials: <https://aiatsis.gov.au/>.<sup>1</sup> The protocols for managing these materials are outlined by ATSILIRN (Aboriginal and Torres Strait Islander Library, Information and Resource Network) on their website: <http://atsilirn.aiatsis.gov.au/protocols.php>.<sup>3</sup>

A statement such as:

*Aboriginal and Torres Strait Islander people should be aware that this content may contain images or names of deceased persons in photographs or printed material.*

Access and outreach

is required to alert viewers to culturally sensitive content.

## CHECKLIST

You now:

- have improved access to your collection or item records;
- have confirmed and stated copyright ownership of your collection;
- have alerted users to culturally sensitive items in your collection.

## 8 COMMUNITY

The cultural heritage and **GLAM** community is a strong network of peak bodies, government organisations, and NGOs. Listed here are some Australian and international organisations who provide advice, assistance, and resources about digital collections.



The ASA state branches run regular workshops for members and non-members and have discussion boards on their website<sup>9</sup>. A number of publications and eLearning courses are available for purchase from their website, here:

<https://www.archivists.org.au/products/shop>.



The Digital Preservation Coalition's Digital Preservation Handbook is an excellent starting point for planning a digital preservation program:

<http://dpconline.org/handbook>.<sup>21</sup>



GLAM • PEAK

A number of resources are available on the GLAM Peak website (<http://www.digitalcollections.org.au/>) which are aimed at small to medium and volunteer-run organisations in Australia. Read about case studies of successful digitisation projects in small organisations<sup>26</sup>. Seminars and workshops are run regularly across Australia and are usually free to attend, but they can book out quickly.



Museums Galleries Australia is the national peak body representing Australian museums, galleries, and collecting institutions<sup>36</sup>. They provide a comprehensive list of links and support organisations to start a digitisation program here:

<https://www.museumsaustralia.org.au/digital>.

Both the national body and state branches provide training on record keeping, including digital collections, at a fee. They also list industry events on their website. See

<https://www.museumsaustralia.org.au/>.

State branch websites can be found here:

ACT:

<https://www.museumsaustralia.org.au/australian-capital-territory/>

NSW: <https://www.museumsaustralia.org.au/new-south-wales>

NT:

<https://www.museumsaustralia.org.au/northern-territory>

Qld:

<https://www.museumsaustralia.org.au/queensland>

SA: <https://www.museumsaustralia.org.au/south-australia>

Vic: <https://mavic.asn.au/>

WA: <http://museumsgalleriesaustraliawa.org.au/>



The NAA provides information about caring for their collection, which is freely available. Find it here: <http://www.naa.gov.au/information-management/managing-information-and-records/preserving/digital-pres/index.aspx>.<sup>40</sup>



The NFSA is Australia's leading audio-visual heritage archive<sup>41</sup>. They offer preservation services for both media and equipment and provide how-to guides here: <https://www.nfsa.gov.au/preservation>.



The Public Record Office of Victoria (PROV) run 'Just Digitise It' workshops and have written a digitisation guide: <https://www.prov.vic.gov.au/community/managing-your-collection/just-digitise-it>.<sup>44</sup>

# 9 GLOSSARY

<b>access</b>	The ability to locate, view, and make use of materials in a collection.
<b>access copy</b>	A copy of an item that is used only for sharing and/or to protect original from damage.
<b>accession</b>	The act of adding a new item to a collection, in a library, archive, or similar.
<b>administrative metadata</b>	Information pertaining to the management of an object.
<b>AIP</b>	Archival Information Package. The digital object and its metadata, bundled together inside a digital repository.
<b>algorithm</b>	A rule, formula, or set of steps that dictate computer processes.
<b>analog(ue)</b>	Non-digital materials or equipment for example paper, cassettes, glass slides, and the equipment used to view them.
<b>appraisal</b>	Determining the value of a collection item to assign retention periods.
<b>archival copy</b>	The high quality full-resolution copy of an item.
<b>Archival Information Package</b>	See <b>AIP</b> .
<b>audio-visual</b>	Items in pictorial and/or audio format, e.g. Photographs, motion pictures.
<b>authenticity</b>	The trustworthiness of an item, including its provenance. An authentic item is free from tampering and corruption.
<b>AV</b>	See <b>audio-visual</b> .
<b>back-up</b>	Secondary copies of an item, stored in case the <b>archival copy</b> is destroyed.
<b>bit</b>	The smallest unit of data that a computer can store. Represented in zeroes and ones.
<b>bit stream preservation</b>	The maintenance of a digital object's bit stream over time, ensuring no corruption or changes are made.
<b>born-digital</b>	Documents that have never been in <b>analogue</b> form; originally captured in electronic formats, for example word documents, emails, HTML, digital photographs etc.
<b>browser</b>	A computer program that accesses the web, e.g. Google Chrome, Internet Explorer, Firefox, Safari.
<b>byte</b>	A unit of digital information, made of six <b>bits</b> .
<b>carrier</b>	The physical medium in which digital content is stored, for example a CD, DVD, hard drive, memory card, thumb drive, etc.
<b>checksum</b>	A method of identifying the integrity of a <b>file</b> . Checksums are created by a computer, using algorithms. Some digital preservation software will create checksums automatically.
<b>complex digital object</b>	A digital object that consists of more than one file type, for example the scanned pages in TIFF format, plus the transcription and metadata of the diary.
<b>compression</b>	An action performed on a digital file to reduce the size of the <b>file</b> for storage or transfer. Compression can be <b>lossy</b> or <b>non-lossy compression</b> .

<b>COPTR</b>	Community Owned digital Preservation Tool Registry. A registry of digital preservation tools, applications, and software.
<b>copy</b>	A duplicate of an original document; digital or <b>analogue</b> .
<b>copyright</b>	The legal right to publish, print, or perform literary or artistic material.
<b>corruption</b>	The alteration of a digital object to the point where it can no longer be read by a computer.
<b>culturally sensitive content</b>	Content which could cause distress or sadness to a cultural group, or content which might not normally be publicly visible.
<b>curate</b>	The selection, presentation, and management of content or objects.
<b>dark archive</b>	An archive that is not publicly accessible.
<b>data capture</b>	The process of collecting data, often automated.
<b>data storage</b>	General term for collecting data in a digital storage medium.
<b>descriptive metadata</b>	Metadata that describes the intellectual content of an object. Used for discovery and identification.
<b>digital native</b>	A person familiar with computers and digital technology from an early age.
<b>digital object</b>	Simple digital objects such as text, image or sound files, or complex digital objects made by combining a number of other digital objects, such as websites.
<b>Digital Object Identifier</b>	See <b>DOI</b> .
<b>dispose</b>	An action at the end of the digital curation lifecycle. <sup>19</sup> The deletion of data that has not been selected for long-term curation in accordance with an organisation's policy.
<b>DOI</b>	Digital Object Identifier. A permanent identifier of an online object used in case the web address changes, the user will be redirected to the new address. Created by the International DOI Foundation and found at <a href="http://www.doi.org">http://www.doi.org</a> .
<b>digital preservation</b>	The management and applied processes necessary for long-term retention of a digital collection.
<b>digital repository</b>	The location in which digital objects are stored and maintained.
<b>digitisation</b>	The capture of <b>analogue</b> material into a digital format through photography or scanning.
<b>DIP</b>	Dissemination Information Package. The digital object and its metadata, bundled together for <b>access</b> .
<b>discovery</b>	The ability to find an item in a collection or archive. Discovery is made easier with good metadata.
<b>disk space</b>	The storage capacity of a digital storage device.
<b>Dissemination Information Package</b>	see <b>DIP</b> .
<b>DOI</b>	See <b>Digital Object Identifier</b> .
<b>Dublin Core</b>	A commonly used metadata <b>element</b> set.
<b>duplication</b>	The creation of a copy, usually unintended. Manage duplicates by keeping a <b>master copy</b> .
<b>dynamic data</b>	Information that changes frequently, therefore attracting challenges to preserve. For example, social media data such as a post on Facebook.
<b>electronic records</b>	Day-to-day business records that are created and maintained electronically
<b>element</b>	One component of a metadata set.

<b>EML</b>	A file type for emails.
<b>emulation</b>	The use of programs that emulate old hardware or software.
<b>encryption</b>	A security procedure that translates plain text into a code that cannot be decrypted without the original code.
<b>file</b>	The basic digital unit within a records series.
<b>file format</b>	The type of <b>file</b> , e.g. Images can be .tiff, .jpeg; moving images can be .mov or .mp4.
<b>file format identification</b>	The process of identifying <b>file formats</b> . Most digital preservation software is capable of this and automatically includes it in the file metadata.
<b>finding aid</b>	A document detailing a collection, including but not limited to the history, subject, materials, source, and structure of the collection.
<b>fixity</b>	The extent to which a digital item remains unaltered.
<b>fonds</b>	The whole of the collection, i.e. the chief archive unit.
<b>gigabyte (GB)</b>	A unit of measurement for digital information, equal to 1,000 megabytes.
<b>GLAM</b>	Industry sector comprising galleries, libraries, archives, and museums.
<b>greyscale</b>	Images with a full range of black, white, and greys.
<b>hardware</b>	The physical components of a computer or digital storage device.
<b>harvest</b>	In collections management, harvesting is the collection of descriptive metadata used by a search aggregator, such as Trove.
<b>HTML</b>	Hypertext Markup Language. The coding language used to make websites.
<b>ingest</b>	The intake of digital objects to a repository.
<b>ISAD(G) or ISAD-G</b>	International Standard for Archival Description (G), published by the International Council on Archives.
<b>ISO 15489</b>	An international standard for information documentation and records management.
<b>item</b>	The smallest archival unit e.g. A photograph, a document etc.
<b>item number</b>	A unique number assigned to an item in a collection.
<b>JPEG</b>	Stands for 'Joint Photographic Experts Group'. An image file format that can be compressed.
<b>kilobyte (KB)</b>	Unit of measurement for digital information, equal to 1,000 bytes.
<b>level of description</b>	The position of a <b>file</b> in the hierarchy of a <b>fonds</b> (collection).
<b>LIDO</b>	Lightweight Information Describing Objects. A metadata standard for describing museum objects.
<b>LOCKSS</b>	Lots Of Copies Keeps Stuff Safe. Open source software that creates replicas of digital content.
<b>long-term preservation</b>	In terms of digital preservation, long-term is the time far enough into the future to be concerned about developments in technology that will affect a digital collection.
<b>lossless compression</b>	The action of decreasing file size without losing information.
<b>lossy compression</b>	Random bits of information are removed from a <b>file</b> while maintaining the overall content of the original. Repeated lossy compression will cause a file to deteriorate.
<b>master copy</b>	A secure, high-quality copy of a <b>file</b> , used to make access copies.
<b>megabyte (MB)</b>	Unit of measurement for digital information, equal to 1,000 kilobytes.
<b>metadata</b>	Latin word literally meaning 'information about information'. The description of the content, context, and structure of an item when created and throughout its lifecycle.



<b>metadata extraction</b>	The automated gathering of metadata, completed by digital preservation software.
<b>metadata standard</b>	A set of metadata elements used to describe an object; examples include <b>Dublin Core</b> and <b>ISAD(G)</b> .
<b>METS</b>	Metadata Encoding and Transmission Standard. Metadata presented in XML format. <sup>32</sup>
<b>micro-services</b>	Small processes that contribute to the preservation of digital objects in software.
<b>migration</b>	The process of changing the format of a <b>file</b> , rendering it possible to open on new hardware or software.
<b>MPEG</b>	Moving Picture Experts Group (MPEG, 2018). Responsible for preservation standards of digital moving images.
<b>OAIS</b>	Open Archival Information System. An archival framework that provides the common language we use for digital preservation.
<b>obsolescence</b>	When a type of hardware or software is no longer commonly used or available.
<b>OCR</b>	Optical Character Recognition. Electronic translation of textual images into text.
<b>open access</b>	Freely available content, usually on the web. This content is available to use, replicate, and distribute.
<b>open format</b>	Freely available standards of software that are universally compatible and not dependent on proprietary components.
<b>open standard</b>	Freely available file formats that are universally compatible and not dependent on proprietary components.
<b>Optical Character Recognition</b>	See <b>OCR</b> .
<b>PDF</b>	Portable Document Format. PDF/A is specifically designed for archival uses. PDFs are a fairly universal <b>file format</b> that is suitable for preservation.
<b>persistent identification</b>	A reference to an online document that remains constant into the long-term.
<b>pixel</b>	“Picture element”. Pixels are the dots that make up a digital image. Image resolution is measured in pixels per inch ( <b>ppi</b> ).
<b>pixilation</b>	The distortion of an image due to low resolution. Can occur due to lossy compression or excessive post-capture editing.
<b>ppi</b>	Pixels Per Inch. Signifies image resolution.
<b>PREMIS</b>	Preservation Metadata: Implementation Strategies. <sup>33</sup> One of the digital preservation metadata standards.
<b>provenance</b>	Information pertaining to the origin of a digital object, including any changes that have been made to it over time.
<b>relationship data</b>	Items in a collection are often linked. These links can be represented as relationship data.
<b>repair</b>	The act of fixing or restoring a corrupted <b>file</b> or file element.
<b>repository</b>	See <b>digital repository</b> .
<b>resolution</b>	Picture elements (pixels) that make up an image.

<b>sandbox environment</b>	A test version of software used to experiment with new developments or trial before installation. Many software packages offer a sandbox environment for potential customers.
<b>scanner</b>	A device that captures a digital image of an <b>analogue</b> item such as a photograph.
<b>series</b>	A group of records that are maintained as a whole due to similarity in structure or location.
<b>server</b>	A computer system which stores data and can be accessed by other computers.
<b>SIP</b>	Submission Information Package. The digital object and its metadata, bundled together for <b>ingest</b> .
<b>software</b>	Sometimes also called an application or program. A set of instructions used by a computer to provide services, such as Microsoft Word, Google Chrome, etc.
<b>spam</b>	Unsolicited content via email.
<b>storage device</b>	Anything used to store data. This could be a server, hard drive, USB/memory stick/thumb drive.
<b>storage hierarchy</b>	A visual representation of the structure of a storage system. From top to bottom: Fonds, Subfonds, Series, Subseries, File, Item. Can be used in tandem with a <b>finding aid</b> .
<b>stream</b>	To transmit or receive data over the internet continuously. Commonly used to receive video and audio content. Popular streaming services include Netflix (video) and Spotify (audio).
<b>structural metadata</b>	Metadata which describes where an object fits in relation to the rest of a collection, for example page number, page layout, or related documents.
<b>Submission Information Package</b>	See <b>SIP</b> .
<b>subfonds</b>	The second-highest level in a storage structure.
<b>terabyte (TB)</b>	A unit of measurement for digital information, equal to 1,000 gigabytes.
<b>thumb drive</b>	A small digital storage device. Also called a USB stick or memory stick.
<b>thumbnail</b>	A small, low resolution image used for browsing and sampling.
<b>TIFF</b>	Stands for 'Tag Image File Format'. A lossless image format.
<b>transfer</b>	The act of moving files from one location to another.
<b>Trove</b>	The National Library of Australia's online library database aggregator.
<b>unique identifier</b>	The primary identifier for an item in a collection, e.g. Item number.
<b>URL</b>	Uniform Resource Locator. An address for a location on the internet.
<b>validation</b>	A process that checks for inconsistencies in a new submission into a repository.
<b>WARC</b>	Web ARChive format. The international standard for archived websites <b>file format</b> . <sup>31</sup>
<b>watermark</b>	An overlaid logo, word(s), or image to denote ownership, used over an image or document.
<b>WAV</b>	Wave. A standard <b>file format</b> for audio.
<b>XHTML</b>	An extended version of <b>HTML</b> used to code websites.
<b>XML</b>	eXtensible Markup Language. A commonly used standard for presenting information, including metadata.

## Glossary

### **zip file**

A file with content that is **compressed** for ease of storage or transfer.

# BIBLIOGRAPHY

## ENDNOTES

- 1 AIATSIS. (2018). Australian Institute of Aboriginal and Torres Strait Islander Studies. Retrieved from <http://www.aiatsis.gov.au/>
- 2 AMIA (2018). "The Association of Moving Image Archivists." Retrieved 26/02/2018, 2018, from <https://amianet.org/>
- 3 ATSILIRN. (2012). ATSILIRN Protocols for Libraries, Archives and Information Services. Retrieved from <http://atsilirn.aiatsis.gov.au/protocols.php/>
- 4 Australian Copyright Council. (2018). Australian Copyright Council. Retrieved from <https://www.copyright.org.au/>
- 5 Australian Copyright Council. (2017). Information Sheet G115v04 Section 200AB: The 'Special Case' or 'Flexible Dealing' Exception.
- 6 Australian Government (2018). *Business*, "Keep the right records". Retrieved from <https://www.business.gov.au/new-to-business-essentials/series-one/keep-the-right-records/>
- 7 Australian Libraries Copyright Committee. (2018). Copyright for Australian Libraries and Archives. Retrieved from <http://libcopyright.org.au/>
- 8 Australian Securities & Investments Commission (2018). What books and records should my company keep?. Retrieved from <http://asic.gov.au/for-business/running-a-company/company-officeholder-duties/what-books-and-records-should-my-company-keep/>
- 9 Australian Society of Archivists. (2017). Retrieved from <https://www.archivists.org.au/>
- 10 Australian Taxation Office (2018). Managing your small business records. Retrieved from <https://www.ato.gov.au/Business/Managing-your-small-business-records/>
- 11 Brown, A. (2013). *Practical Digital Preservation: a how-to guide for organizations of any size*. London: Facet Publishing.
- 12 Collections Trust. (2018). Digitisation: a simple guide for museums. Retrieved from <https://collectionstrust.org.uk/resource/digitisation-a-simple-guide/>
- 13 Commonwealth of Australia. Copyright Act 1968, (1968).
- 14 Conway, P. (2000). Overview: rationale for digitization and preservation. In M. Sitts (Ed.), *Handbook for digital projects: a management tool for preservation and access*. Andover, Massachusetts: Northeast Document Conservation Center.
- 15 Copyright Agency. (2018). Retrieved from <https://www.copyright.com.au/>
- 16 Corrado, E. M. (2014). *Digital preservation for libraries, archives, and museums*: Rowman & Littlefield.
- 17 Creative Commons. (2018). Creative Commons. Retrieved from <https://creativecommons.org.au/>

## Bibliography

- 18 Department of Communications and the Arts. (2018). Copyright. Retrieved from <https://www.communications.gov.au/what-we-do/copyright/>
- 19 Digital Curation Centre (2004-2017). "DCC Curation Lifecycle Model." 2017, from <http://www.dcc.ac.uk/resources/curation-lifecycle-model/>
- 20 Digital Library Federation. (2018). Digitization Cost Calculator. Retrieved from <http://dashboard.diglib.org/>
- 21 Digital Preservation Coalition. (2018). Retrieved from <https://www.dpconline.org/>
- 22 Dublin Core Metadata Initiative. (2018). Dublin Core Metadata Initiative. Retrieved from <http://dublincore.org/>
- 23 Erway, R. (2012). *You've got to walk before you can run: first steps for managing born-digital content received on physical media*. Retrieved from <https://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>
- 24 Facebook. (2018). Help Centre. Retrieved from <https://www.facebook.com/help/>
- 25 Fair Work Ombudsman (2018). Record-keeping. Retrieved from <https://www.fairwork.gov.au/pay/pay-slips-and-record-keeping/record-keeping/>
- 26 GLAM Peak. (2017). Digital Access to Collections. Retrieved from <http://www.digitalcollections.org.au/>
- 27 GLAM Peak. (2018). *Digital Access to Collections Workshop Manual v4*.
- 28 Harvey, R., & Mahard, M. R. (2014). *The Preservation Management Handbook : a 21st-Century Guide for Libraries, Archives, and Museums*: Rowman & Littlefield Publishers.
- 29 International Council on Archives. (1999). ISAD(G): General International Standard Archival Description, Second Edition. In. Stockholm.
- 30 International Council on Archives. (2010). LIDO - Lightweight Information Describing Objects Version 1.0. In: International Council on Archives.
- 31 International Organization for Standardization. (2017). Information and documentation -- WARC file format. In (Vol. ISO 28500:2017): International Organization for Standardization.
- 32 Library of Congress (2017). "METS." Retrieved 26/02/2016, 2018, from <http://www.loc.gov/standards/mets/>
- 33 Library of Congress. (2017). PREMIS. Retrieved from <http://www.loc.gov/standards/premis/>
- 34 Library of Congress. (2018). Technology at the Library: Long-Hidden Text Is Uncovered in Alexander Hamilton Letter. Retrieved from <https://blogs.loc.gov/loc/2018/01/technology-at-the-library-long-hidden-text-is-uncovered-in-alexander-hamilton-letter/>
- 35 Lowry, D., & Durling, S. (2017). Managing born-digital documents. In N. Y. S. Archives (Ed.).
- 36 Museums Galleries Australia (2017). 2017, from <https://www.museumsaustralia.org.au/>
- 37 National and State Libraries Australasia. (2018). Digital preservation. Retrieved from <https://www.nsla.org.au/resources/digital-preservation-principles>
- 38 National Archives of Australia. 2017. Retrieved from <http://www.naa.gov.au/>

## Bibliography

- 39 National Archives of Australia ([2018]). "Preservation File Formats." Retrieved 8/08/2018, 2018, from <http://www.naa.gov.au/information-management/managing-information-and-records/preserving/long-term-file-formats.aspx>
- 40 National Archives of Australia. (2017). Preserving Digital Records. Retrieved from <http://www.naa.gov.au/information-management/managing-information-and-records/preserving/index.aspx>
- 41 National Film and Sound Archive of Australia. 2017. Retrieved from <https://www.nfsa.gov.au/>
- 42 National Library of Australia. (2018). various entries. Retrieved from <https://www.nla.gov.au/blogs/preservation/2016/08/08/managing-your-digital-records>
- 43 National Library of Australia. (2018). For smaller collections. *Trove*. Retrieved from <http://help.nla.gov.au/trove/becoming-partner/for-smaller-collections>
- 44 PROV. (2018). Just digitise it. Retrieved from <https://www.prov.vic.gov.au/index.php/community/managing-your-collection/just-digitise-it>
- 45 Rafferty, E. (2017). Better together: a holistic approach to creating a digital preservation policy in an art museum. *Journal of the Art Libraries Society of North America*, 36(1).
- 46 Schermerhorn, J. R., & al, e. (2011). *Management foundations and applications* (1st Asia Pacific ed. ed.). Australia: John Wiley & Sons.
- 47 State Library of Queensland. (2018). Digitisation Toolkit. Retrieved from <http://www.slq.qld.gov.au/resources/qld-history/connect/connected-collection-of-queensland-memory/digitisation-toolkit>
- 48 University of Tasmania. (2018). Records Management Unit Forms and Templates. Retrieved from <http://www.utas.edu.au/it/records/forms>

## ADDITIONAL RESOURCES

- Barateiro, J. (2010). "Designing digital preservation solutions: a risk management-based approach." *The International Journal of Digital Curation* 5 (1).
- Community Owned digital Preservation Tool Registry (2016). "COPTR." Retrieved 26/02, 2018, from [http://coptr.digipres.org/Main\\_Page](http://coptr.digipres.org/Main_Page)
- Cornell University Library (2003-2007). "Digital Preservation Management." Retrieved 26/02/2018, 2018, from [http://www.dpworkshop.org/dpm-eng/eng\\_index.html](http://www.dpworkshop.org/dpm-eng/eng_index.html)
- Council on Library Resources (2018). "CLIR." Retrieved 26/02/2018, 2018, from <https://www.clir.org/>
- Digital Curation Centre (2018). "DRAMBORA." Retrieved 26/02/2018, 2018, from <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora>
- Digital Curation Centre (2018). "Glossary." 2018, from <http://www.dcc.ac.uk/digital-curation/glossary>
- Dooley, J. (2015). *The archival advantage: integrating archival expertise into management of born-digital library materials*, OCLC Research.
- Harvey, D. R. (2012). *Preserving Digital Materials*. Boston, De Gruyter Saur.
- International DOI Foundation (2017). "DOI." Retrieved 26/02/2018, 2018.

## Bibliography

- Kandiuk, M., et al. (2013). Digital Image Collections and Services. Washington DC, Association of Research Libraries.
- Library of Congress (2017). "Sustainability of digital formats: planning for Library of Congress collections." Retrieved 26/02/2018, 2018, from <https://www.loc.gov/preservation/digital/formats/index.shtml>
- Matusiak, K. K., et al. (2017). "Finding access and digital preservation solutions for a digitized oral history project: a case study." Digital Library Perspectives **33**(2): 88-89.
- McMillan, G., et al. (2011). Digital Preservation October 2011. Washington, DC, Association of Research Libraries.
- MPEG. (2018). The Moving Picture Experts Group. Retrieved from <https://mpeg.chiariglione.org/>
- NSW State Archives & Records (2018). "Benefits, costs and risks of business process digitisation programs." Retrieved 7/03/2018, 2018, from <https://www.records.nsw.gov.au/recordkeeping/benefits-costs-and-risks-business-process-digitisation-programs>
- Ochoa, M. N., et al. (2014). Digital Collections Assessment and Outreach August 2014. Washington DC, Association of Research Libraries.
- Park, E. G. and S. Oh (2012). "Examining attributes of open standard file formats for long-term preservation and open access." Information Technology and Libraries **31**(4).
- POWRR, D. (2012). "Preserving digital Objects With Restricted Resources." Retrieved 26/02/2018, 2018, from <http://digitalpowrr.niu.edu/>
- Society of American Archivists (1997-2017). "A Glossary of Archival and Records Terminology." 2017, from <https://www2.archivists.org/glossary>
- Traczyk, T., et al. (2017). Digital Preservation: Putting it to Work, Springer International Publishing AG.
- University of Michigan Library (2018). "Digital Preservation." 2018, from <https://www.lib.umich.edu/preservation-and-conservation/digital-preservation>