Guidance for Digital Preservation Workflows

Authors: Kevin Bolton, Jan Whalen and Rachel Bolton (Kevinjbolton Ltd)

Any enquiries regarding this publication should be sent to: asd@nationalarchives.gov.uk.

Introduction

The guidance was commissioned by the Archive Sector Development department (ASD) of The National Archives (TNA). It aims to support archives in the United Kingdom to move into active digital preservation work by providing those who work with archives:

- Practical examples of workflows for managing born digital content, that you can change and use in your own organisation.
- Actions for how to process and preserve born digital content, including using free software.

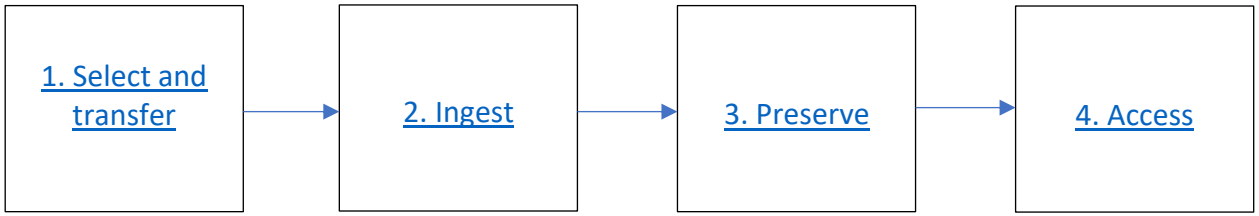
In this guidance, a workflow is a number of connected steps that need to be followed from start to finish in order to complete a process.

You do not need a significant level of digital preservation knowledge in order to follow the guidance. Certain terminology is explained in the glossary. In the guidance we refer to "digital content" or "content" – this is what we hope to preserve. Digital preservation literature often calls this "digital objects".

The guidance will show you which steps are "Essential" and you may prefer to follow these steps only. It is better to do something, rather than nothing! We are not promoting a 'one size fits all' approach and expect archives to use and adapt the guidance depending on the needs of their organisation.

Each step includes links to software, online training and further guidance such as links to documentation, blogs and videos. You can also find templates in Appendix B that which can be presented to your IT departments to make a case for installation of some of the key pieces of software.

The guidance is arranged in four sections covering the following workflows:



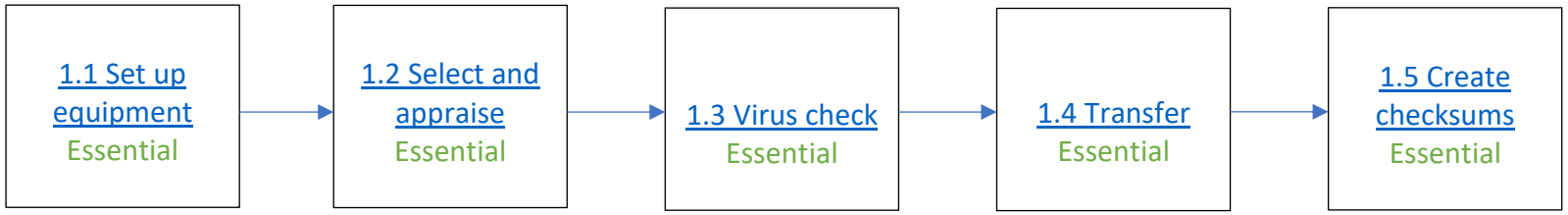
The following table summarises the software you may require to carry out these workflows.

Beginner	Intermediate	Advanced
• **Anti-virus software.** • **Copying software** such as Teracopy, Data Accessioner or Robocopy (a Windows command) for transferring or moving content. • Software such as DROID, to **identify and list your content.** • Software such as DROID or AVP Fixity to **create** checksum • Software such as CSV Validator with an integrity schema or AVP Fixity to carry out **integrity checks**.	Same as beginner + : • **Disk imaging software** such as FTK Imager Lite or BitCurator. • **Encryption software** such as VeraCrypt or Bitlocker. • Software such as Quick View Plus and VLC for **viewing or playing** content. • **De-duplication software** such as CSV Validator with the deduplication schema and TreeSize Free.	Same as intermediate + : • **Packaging software** such as Bagger or Exactly**.** • **Validation software** such as JHOVE, Jpylyzer, veraPDF and MediaConch. • Software to help with **analysis** such as Freud and HxD Hex Editor. • Software such as Bulk Extractor that can help **identify sensitive information.** • Redaction software for carrying out **redaction** of sensitive information. • Software for **migrating and converting** file formats such as FFmpeg, ImageMagick, Ghostscript, MIXED, LibreOffice and Apache Open Office. • ePADD for working with **email archives**.

1. Select and transfer

This workflow describes the process of selecting the content and obtaining it from the depositor.

Summary



1.1 Set up equipment — Essential

- Set up a dedicated PC to connect to any storage media holding the digital content. Ideally, only connect it your organisation's systems / internet to perform essential updates to systems.
- You will need equipment to read various types of media you will work with. For example:
  - Readers for DVDs / CDs and floppy disks (particularly 3.5", $5^1/_4$" and 3"). You may wish to look at Kryoflux if you want to read floppy disks.
  - Zip drives, tape drives and a caddy for internal hard drives.
  - See Appendix C for photographs.
- You can use write blockers to prevent changes to the content – especially when using hard drives or floppy disks. The type needed will depend on the media you want to read.
- You may receive content internally or by email / the internet and will require a PC with access to your organisation's systems / internet. Alternatively, use an external hard drive for transfer.
- You may also wish to look at using encryption software (see below) on the hard drive of the PC and external hard drives, especially if you work with sensitive content.

Encryption software

- VeraCrypt
- Bitlocker

Further guidance

- Building a digital curation workstation with Bitcurator (blog by Porter Olsen)
- What's Your Set-up?: Curation on a Shoestring (blog by Rachel MacGregor)
- Digital Preservation Lab: Equipment (blog by University of Michigan Library)
- Outfitting a Born-Digital Archives Program (blog by Ben Goldman)
- Forensic Workstation pt3 (blog by Hull University Archives on write blockers)
- Archivists Guide to Kryoflux

1.2 Select and appraise — Essential

- Your organisation's collection policy should determine selection - the Digital Preservation Coalition Handbook gives a good overview of the key things to consider.
- Ensure that any information from the depositor about Intellectual Property Rights and access restrictions is captured at this stage.
- You could ask the depositor to create a list of the content that is being transferred (they could use the software below to do this).
- At this stage you may wish to carry out appraisal of the content (although this can be done at a later stage - see step 2.5). The Paradigm project offers a good summary of the issues around digital appraisal.
- Create an accession number which will be later used in step 2 (Ingest).
- Create a folder on the PC (e.g. using the accession number for the folder title). You may wish to create subfolders – one for the content (e.g. called "content") and one for any documentation about the content (e.g. called "metadata").

Software

- Karen's Printer or DROID (a depositor could use these tools to make a list of their content)
- Windows Command Prompt (a depositor could also use this to create directory listings)

Further guidance

- Sample transfer lists (Paradigm project)
- Paradigm Project – Appraisal and Disposal
- DPC Handbook: Acquisition and Appraisal
- University of Hull Idiot Guide No. 4: Karen's Directory Printer
- Practical Digital Preservation: In-House Solutions to Digital Preservation for Small Institutions (includes a section on creating folders/directories)

1.3 Virus check — Essential

- Place the media into the appropriate reader or port (remember to use a write blocker).
- If possible, scan the content for viruses using anti-virus software on the media before transferring them (see step 1.5).
- Remove any infected content and decide on action e.g. repair or contact the depositor for clean copies.
- You could leave (quarantine) the content on your PC for 30 days and then re-scan them for viruses before proceeding to step 2 (Ingest). Alternatively if you virus check with two different types of anti-virus software to reduce the risk of missing any viruses.
- Following the quarantine period you may wish to check the checksums created at step 1.4.
- Keep a record of what virus checks you have undertaken (e.g. save any report the software generates in the "metadata folder").

Software	**Further guidance**
Use the anti-virus software your organisation subscribes to or use free anti-virus software such as ClamAV or AVG	Dealing with computer viruses in digital collections (British Library blog) Do not try this at home (blog by Rachel MacGregor on the practical elements of quarantine)

1.4 Transfer — Essential

- Transfer the digital content from the media to the "content" folder on the PC using copying software (see below). This software helps ensure important information such as dates are not changed. Some software will also check identical complete copies were made.
- Some archives ask the depositor to use software such as Bagger or Exactly to transfer content over the internet or on a portable storage device.
- Disk imaging is an alternative to copying the content. Software, such as FTK Imager Lite, can create an exact copy of the contents of the media, including original metadata.

Copying software

- Teracopy (copies content and checks complete identical copies were made)
- Data Accessioner (for migrating content between media and also creating and checking checksums)
- Robocopy command line (for copying)

Disk imaging software

- FTK Imager Lite
- BitCurator (includes disk imaging tools)

Content transfer software

- Bagger
- Exactly

Further guidance

- Teracopy User Manual
- Data Accessioner (video)
- Running the robocopy command (Canadian Heritage Information Network)
- University of Hull Idiot Guide No. 3: FTK Imager
- BitCurator Quick Start Guide and other documentation
- Bagit: Transferring Content (video)
- Bagger Tutorial (State Archives of Carolina videos)
- Digital Content Transfer (Library of Congress)
- Guidance for Donors and Depositors Using Bagger (Gloucestershire Archives)
- AVP's Exactly Tool Webinar Recording (video)

1.5 Create checksums — Essential

- If you or the depositor created checksums before the transfer then they should be checked afterwards to ensure they remained the same.
- If not, use software (see below) to create checksums and if possible save them with the content (e.g. in the "metadata" folder for the accession).
- The National Archives currently uses a type of checksum called SHA-256. However, other archives use a MD5 checksum.
- At this point you may want to create a copy of the content that will be used for the steps outlined in section 2 (sometimes called a "working copy"). This will reduce the risk of the content being changed.

Checksum software

- DROID and Karen's Directory Printer (will create checksums - see step 2.1)
- CSV Validator and integrity schema (can be used to check checksums created in DROID)
- AVP Fixity
- Jacksum

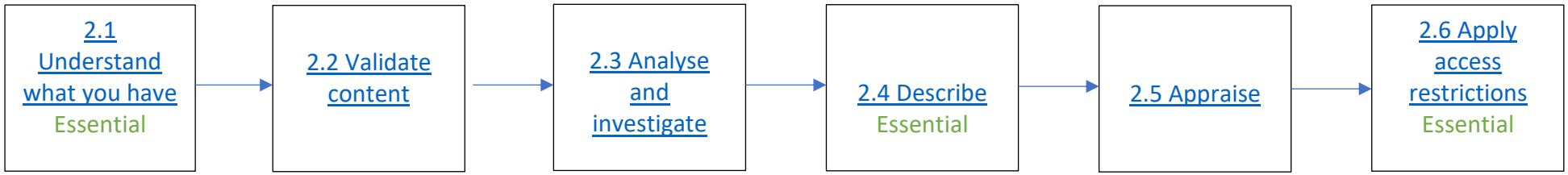
Further guidance

- Checksum or Fixity? Which tool is for me? (University of York blog)
- DPC Handbook: Fixity and checksums
- DROID report as basis for collection integrity checks (article by The National Archives)
- AVP's Fixity Tutorial (video)

2. Ingest

This workflow describes how you prepare the content so it is ready for preserving at the next stage.

Summary



2.1 Understand what you have — Essential

- Use software, such as DROID, to identify what you have and create a list of the content. This should include file names, file paths, sizes, file format, last modified date etc.
- Identifying the file formats accurately is particularly important.
- Save the list in an open format (e.g. CSV or XML) and store in the "metadata" folder you created in step 1.2.

Software	**Further guidance**
DROID (identifies file format and other information) Fido (identifies file format only) MediaInfo (useful for identifying audiovisual files) Karen's Directory Printer (useful for creating lists of files but does not carry out identify file formats with the same degree of certainty as DROID or Fido)	DROID: User Guide DROID Video Demo University of Hull Idiot Guide No. 4: Karen's Directory Printer University of Hull Idiot Guide No. 5: Droid Fido for format identification, and why it matters (Open Preservation Foundation Webinar) Bodleian Libraries: Introduction to Digital Preservation: Identification

2.2 Validate content

- Validation software checks whether the content conforms to their file format specification. In some cases it can also fix issues.
- It is not always seen as an essential step but can help flag issues. For example, if the content does not conform to this specification then it may be more difficult to read or manage in the future.
- It can also be useful for checking the quality of digitised content.

Validation software	**Further guidance**
JHOVE (validates certain file formats and also carries out identification) Jpylyzer (validates JPEG 2000 Part 1) veraPDF (validates PDF/A) MediaConch (validates audiovisual files)	Bodleian Libraries: Introduction to Digital Preservation: Validation (includes good links to various open source tools) JHOVE Documentation

2.3 Analyse and investigate

- You may wish to analyse the metadata you captured during steps 2.1-2.2 and flag any issues for investigation.
- This includes looking out for corrupt files, compressed files, encrypted files and password protected files. You will probably need to go back to the depositor to resolve these.
- It can also flag unidentified formats which could require further research.
- Some archives also convert file formats to a preferred file format for preservation (see step 3.5).

Software	**Further guidance**
- Freud (used by The National Archives to analyse a DROID export and pick up common issues to mark for investigation) - HxD Hex Editor (displays the bytes of a file and helps with file format research)	- How to research and develop signatures for file format identification (The National Archives) - My first file format signature (University of York blog)

2.4 Describe

Essential

- As a minimum create a high-level description of the content.
- You may decide to do more detailed cataloguing in accordance with your organisation's cataloguing standards (either now or at a later date).
- You can add the descriptions to the list you created in step 2.2 or create them in a CSV or XML file.
- If you use a collection management system you may wish to record the descriptions there (e.g. the accession record or catalogue).

Software

- Quick View Plus (allows you to view over 300+ file formats. $99 per year)
- VLC (for playing audio and video files)

Further guidance

- Levels of Born-Digital Access (pp.10-13 cover description)
- Paradigm - Arranging and cataloguing digital and hybrid archives
- Digital Cataloguing Practices at The National Archives
- Quick View Plus Product Fact Sheet and Supported File Format List

2.5 Appraise

- You may have already carried out appraisal at step 1.2. At this stage you may wish to carry out further appraisal.
- As a minimum you could consider identifying and removing duplicates by comparing the checksums of the content. There is software that can help you do this (see below).
- However, you may decide to keep duplicates if they have useful contextual information (e.g. file name).

De-duplication software

- CSV Validator and deduplication schema (can be used for de-duplication)
- Seeing double (Blog by Rachel MacGregor on deduplication using the CSV Validator)
- TreeSize Free
- ePADD (supports the appraisal of email archives as well as processing, preservation, and discovery)

Further guidance

- Paradigm Project – Appraisal and Disposal
- DPC Handbook: Acquisition and Appraisal
- Susanne Belovari (2017) Expedited digital appraisal for regular archivists: an MPLP-type approach, *Journal of Archival Organization*, 14:1-2, 55-77
- Victoria Sloyan (2016) Born-digital archives at the Wellcome Library: appraisal and sensitivity review of two hard drives, *Archives and Records*, 37:1, 20-36

2.6 Apply access restrictions — Essential

- Some of the content may contain personal, sensitive or confidential information.
- If the content is subject to the Freedom of Information Act then you will need to use the Act's exemptions to inform any restrictions.
- The depositor should help you identify this during transfer at step 1.2. Cataloguing at step 2.4 can also help with this.
- There is software that can help you identify personal information. Some of it is commercial and expensive, but a list of free software can be found below.
- Access restrictions or any risks should be recorded somewhere (e.g. in the list you created in step 2.2 and/or in any collection management system).

Software

- Bulk Extractor
- BitCurator (digital forensics tools for digital preservation including Bulk Extractor)
- ePADD (can help identify sensitive information in email archives)

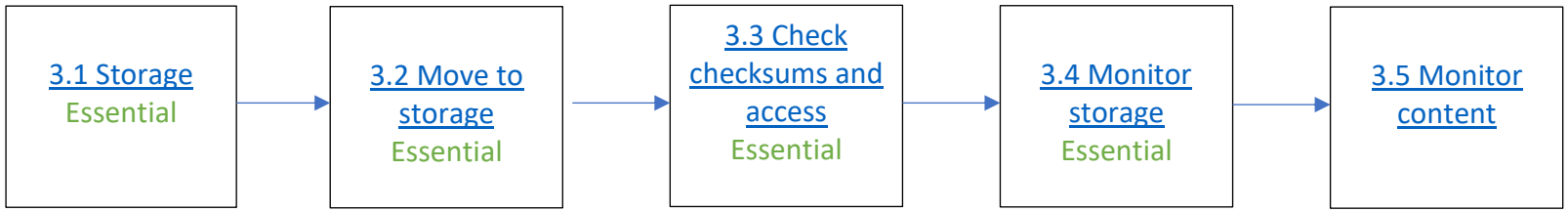
Further guidance

- Victoria Sloyan (2016) Born-digital archives at the Wellcome Library: appraisal and sensitivity review of two hard drives, *Archives and Records*, 37:1, 20-36
- BitCurator: Using Bulk Extractor to Locate Potentially Sensitive Information (video)

3. Preserve

This workflow describes the process of transferring content to secure storage and preserving it.

Summary



3.1 Storage — Essential

- The NDSA Levels of Digital Preservation are useful for planning storage - in particular the sections on 'storage' and 'control'. Think about creating several copies, in different physical locations and using different storage technologies.
- If you currently have no storage, think about some practical solutions. For example, as an interim approach you could use your organisation's storage network (see the guidance below for more information).
- Think carefully about who in your organisation is allowed to access the digital content and the type of access that they have (e.g. read, write, move, delete). Keep a record of who has access.

Software	**Further guidance**
-	- DPC Handbook: Storage - Bit by bit: Processing Born Digital Accessions at National Records of Scotland (includes a description of their interim solution) - Digital Preservation Guidance for Scottish Local Authorities (includes an overview of storage options) - Digital preservation recommendations for small museums (Canadian Heritage Information Network – includes practical storage solutions) - Building Wellcome Collection's new archival storage service (blog) - Protecting Your Data: Backups, Archives & Data Preservation (DataONE presentation)

3.2 Move to storage — Essential

- Before moving the content to the storage, check the documentation you created in steps 1-2 has been saved in the “metadata” folder.
- If you use a collection management system, you may wish to add some of this documentation to it (e.g. the accession record or catalogue) or record where it is stored.
- Some archives will package the content and metadata in a ‘bag’ using software such as Bagger.
- Move the content to the storage. You could use copying software (see below) to do this to ensure date information and other file attributes are preserved.
- Some software will also check the copied content to ensure it is identical. If not, use checksum software to check this (see step 1.4).
- If applicable, you may decide to keep the original storage media or photograph it.

Software

- Bagger (packages files for transfer and storage)
- Teracopy (copies content and ensures that they are identical)
- Data Accessioner (for migrating content between media and also creating and checking checksums)
- Robocopy command line (for copying)

Further guidance

- Bagger Tutorial (State Archives of Carolina videos)
- Teracopy User Manual
- Data Accessioner (video)
- Running the robocopy command (Canadian Heritage Information Network)

3.3 Check checksums and access — Essential

- Use checksum software (see below) to carry out regular integrity checks of the content.
- Keep a record of when you carry these out.
- If checksums of content do change then investigate. For example, if the content is corrupt or has been accidently changed, it may need to be replaced.
- Ideally, you should keep logs of actions performed on content and carry out periodic reviews of these logs.

Software

- AVP Fixity
- CSV Validator and integrity schema (can be used to check checksums created in DROID)
- Jacksum

Further guidance

- Checking Your Digital Content: How, What and When to Check Fixity? (NDSA – Draft Fact sheet)
- DPC Handbook: Fixity and checksums
- AVP's Fixity Tutorial (video)
- DROID report as basis for collection integrity checks (article by The National Archives)
- Checksum or Fixity? Which tool is for me? (University of York blog)

3.4 Monitor storage

Essential

- The lifetime of storage can be short - it can fail or corrupt the content.
- You will need to review your storage every 3-5 years and move content onto new storage.
- Create multiple copies and use a mix of different types of storage technologies if you can.
- For hard drives there is software that can help you with this (see below).

Software	**Further guidance**
Windows 10: Built in tools for hard drive health check (includes undertaking a 'S.M.A.R.T analysis' of hard drives)	DPC Handbook: Storage Digital preservation recommendations for small museum (Canadian Heritage Information Network. Includes section on refreshing storage including some practical tips) How to See If Your Hard Drive Is Dying with S.M.A.R.T. (blog)

3.5 Monitor content

- You should monitor your content to understand if any of the file formats you hold, or the software/technology needed o taccess them, are at risk of becoming obsolete (outdated or no longer used).
- One solution is format migration where a file format is converted into a new file format. However if you do this it is important to also keep the original content.
- Some archives undertake format migration during step 2 (Ingest) and convert particular types of content into a preferred file format (called normalisation). Others wait until the risk of the content becoming obsolete is high.
- One low-cost option is to only migrate the content when someone wants to access it.
- Several types of software can carry out format migration and some are listed below.
- Emulation is an alternative to format migration and attempts to recreate the functionality of the original software or technology.

Software

- FFmpeg (for audio and video)
- ImageMagick (for images)
- Ghostscript (for pdfs)
- MIXED (converts some spreadsheets and databases to XML)
- LibreOffice or Apache Open Office (for word processing documents)
- Joyce (an Amstrad emulator)

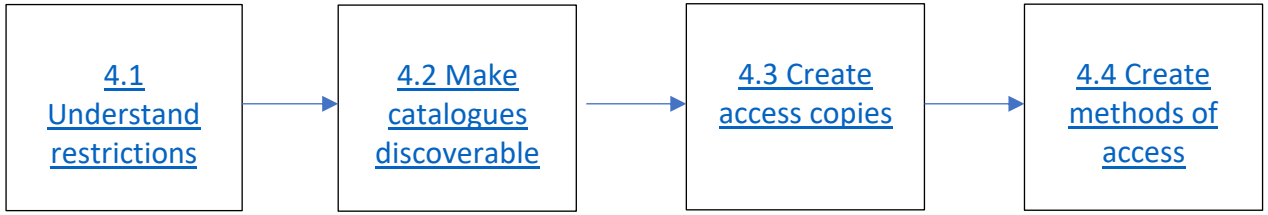
Further guidance

- DPC Handbook: Preservation Actions
- Library of Congress Recommended Formats Statement
- Archivematica format policies
- Practical Digital Preservation (blog by Tyler McNally which includes a section on normalisation software)
- File migration formats (blog by University of Glasgow)
- Bodleian Libraries: Introduction to Digital Preservation: Emulation

4. Access

This workflow describes the process of making content easy to find and accessible for users. The 'Levels of Born-Digital Access' by DLF Born-Digital Access Working Group is also a useful tool for planning access.

Summary



4.1 Understand restrictions

- In step 1.2 you should have captured any Intellectual Property Rights about the content. This will determine what users can or cannot do with the content and where they can view it.
- In step 1.2 and step 2.6 you may have captured other access restrictions including sensitive and personal information. This will also determine how the content can be accessed.
- The Levels of Born-Digital Access recommends using terms such as 'Closed', 'Open', 'Conditional Access' or 'Sensitive'.
- Redaction may be needed before access is provided to users (see guidance below).

Software	**Further guidance**
List of redaction software (DigiPres Commons)	Levels of Born-Digital Access (pp.25-26) The National Archives - Redaction Toolkit

4.2 Make catalogues discoverable

- Make your catalogue easy to find and access online (e.g. online catalogue).
- The catalogue should include access arrangements and restrictions.
- If you do not have an online catalogue then consider using Manage Your Collections or Archives Hub.
- At this point it simply could be a collection level catalogue.
- Also think about taking part in the Accessions to Repositories survey.

Catalogue portals	**Further guidance**
Discovery Archives Hub Aim25 (Greater London) SCAN (Scotland) ANW (Wales)	Manage Your Collections Help Accessions to Repositories

4.3 Create access copies

- You may wish to create access copies of the content.
- This can include converting the content to a different file format to reduce their size (e.g. MP3 for audio content, JPG for images) and/or to make the content more accessible because free viewers are available (e.g. PDF).
- Some archives create access copies during step 2 (Ingest) or step 3.5 (Monitor content).
- Alternatively, you may decide to only create access copies when someone requests access.

Software	**Further guidance**
- FFmpeg (for audio and video) - ImageMagick (for images) - Ghostscript (for pdfs) - LibreOffice or Apache Open Office (for word processing documents)	- Archivematica format policies (includes access formats) - Shein, Cyndi (2014) "From Accession to Access: A Born-Digital Materials Case Study," *Journal of Western Archives*: Vol. 5 : Iss. 1, Article 1. (p.20 includes section on creating access copies)

4.4 Create methods of access

- Ideally, you want a web interface which enables users to access both the catalogue and the content online. Although some content may have to be viewed onsite due to Intellectual Property Restrictions or access restrictions (see step 4.1).
- If you have a collection management system or online catalogue they may allow you to provide access to content such as images, audio-visual and PDFs.
- However, in practice many archives will have to develop practical interim solutions. These may include:
  - Providing access to the content at a dedicated secure PC with viewing software (see further guidance below) at the archive.
  - If no restrictions apply, sending (via portable media or a download) users a copy of the content and any supporting documentation – either in its original format or as access copies.

Software

- SCOPE (a free digital archives access interface)
- ePADD (supports the appraisal of email archives as well as processing, preservation and discovery)
- Quick View Plus (allows you to view over 300+ file formats. $99 per year)
- VLC (for playing audio and video files)
- LibreOffice or Apache Open Office (for word processing documents)
- BitCurator (access tools)

Further guidance

- Levels of Born-Digital Access (especially the sections on security, tools and mediation)
- DPC Handbook: Access
- SCOPE: A digital archives access interface (article by Kelly Stewart & Stefana Breitwieser)
- New Shared Born Digital Access Solution at Yale University Library (blog on creating a workstation for access)

Appendix A - Glossary

- Accession – “Material that comes into an archive as a single acquisition is described as an accession. A number of accessions may form one single collection with shared provenance, e.g. the records of a business may be transferred to an archive over time.” (ArchivesHub)

- Appraisal - the process of identifying which content has continuing value and which content can be disposed. See The National Archives – What is Appraisal? for more information.

- Born digital – “Digital materials which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form.” (Digital Preservation Coalition Handbook)

- Checksum – “A checksum is a string of characters that relate to a digital object, and which act as the object’s unique signature or digital finger print. Checksums can be used for checking the integrity of a digital object through comparison of the checksum over time.” (Community Archives and Heritage Group - Digital Preservation for Community Archives)

- Catalogue / Description – “A description of the material within an archival collection, providing essential information about the collection. Often also called an archival description, a catalogue, or a finding aid.” (ArchivesHub)

- Depositor – the person or organisation donating or depositing the content to the archive.

- Digital objects – “describes an aggregated unit of digital content comprised of one or more related digital files. These related files might include metadata, master files and/or a wrapper to bind the pieces together.” (Bodleian Libraries - Introduction to Digital Preservation)

- Disk image – “A disk image is a file containing an exact copy of the entire contents of an electronic storage device.” (Community Archives and Heritage Group - Digital Preservation for Community Archives)

- Emulation – the use of software (an emulator) to recreate an obsolete software and hardware environment, allowing access to original digital content and providing an authentic user experience.

- Encryption software – a security tool to prevent unauthorised access to digital content.

- Format migration – “A means of overcoming technical obsolescence by preserving digital content in a succession of current formats or in the original format that is transformed into the current format for presentation. The purpose of format migration is to preserve the digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.” (NDSA)

- Integrity Checking - a process that uses checksums to ensure that digital content has not been altered, lost, or damaged over time.

- Intellectual Property Rights - any rights an individual or organisations hold in the content including copyright and design rights.

- Metadata - data about data. It is required “to manage and preserve digital materials over time and…..assist in ensuring essential contextual, historical, and technical information are preserved along with the digital object.” (Digital Preservation Coalition Handbook)

- Normalisation – “Some digital repositories will place a limit on the number of formats which they will support, and as such may only support the formats which most best overall promote functionality, longevity and preservability. Normalization, in this instance, is the process of converting a digital object from its original format to an accepted format, so that a repository can ingest and support the object.” (Community Archives and Heritage Group - Digital Preservation for Community Archives)

- Open Source - software for which the original source code is made available and may be redistributed and modified by users in accordance with an approved open source license.

- Package – “any arbitrary container of digital data” and "the act of creating an arbitrary container of digital data.” (NDSA)

- Redaction – “the separation of disclosable from non-disclosable information by blocking out individual words, sentences or paragraphs or the removal of whole pages or sections prior to the release of the document.” (The National Archives – The Redaction Toolkit)

- Refreshing – “Copying information content from one storage media to the same or another storage media.” (Bodleian Libraries - Introduction to Digital Preservation)

- Selection – a decision making process to decide which content is transferred to the archive.

- Storage media – devices that store the original digital content e.g. CDs, DVDs, floppy disks and hard drives.

- Validation - checks whether the digital content conforms to their file format specification.

- Write blocker – “an electronic device which prevents the ability for digital objects to be changed or altered during the process of transfer from one storage device to another. Write blockers were developed as a digital forensics tool, but can be used for digital preservation purposes during the ingest of digital objects into a repository.” (Community Archives and Heritage Group - Digital Preservation for Community Archives)

Appendix B – Business cases for software

These templates can be presented to your IT departments to make a case for installation of some of the key pieces of software. For creating business cases to fund digital preservation activities you may find the Digital Preservation Business Case Toolkit (Digital Preservation Coalition) useful.

DROID
About DROID is a software tool developed and used by The National Archives to perform automated batch identification of file formats. DROID is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies. DROID uses internal signatures to identify and report the specific file format and version of digital files. These signatures are stored in an XML signature file, generated from information recorded in the PRONOM technical registry. New and updated signatures are regularly added to PRONOM, and DROID can be configured to automatically download updated signature files.
Download link https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/
User guide http://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf
System requirements DROID requires Java 1.7 or 1.8 Standard Edition (SE) and should work on any platform which supports either of these.

Installation instructions can be found in Section 2 of the user guide.

DROID is built and tested on:

Red Hat Enterprise Linux Server 64 bit/OpenJDK
Ubuntu Desktop 64 bit/Oracle Java
Linux Mint
CentOS 64 bit/Oracle Java
Microsoft Windows 7 (32/64 bit)/Oracle Java
Microsoft Windows 10 (64 bit)
Microsoft Windows Server 2008 (64 bit)/Oracle Java
Mac OS Sierra/Java

Business case for installation

Digital Preservation looks specifically at the activities necessary to preserve, and to ensure continued long-term access to digital material. Passive preservation is not an option - allocating a priority to the preservation of digital material much more urgent than for paper archives. Unlike paper, a digital material which is not selected for active preservation treatment at an early stage in its existence will very likely be lost or unusable in a few years' time.

Increasingly we are creating and collecting digital material. This includes:

- Digitised content of collections where we hold the original;
- Digitised content of collections where we don't hold the originals;
- Born digital collections, which have been created and managed electronically. Common examples of born digital objects are photographs taken with a digital camera, an email or a text document.

A key element of digital preservation is understanding file formats. File formats and software used can become obsolete over time. Therefore, it is possible to have successfully preserved something but lack the means to access it. Strategies such as migration can help mitigate these risks. It is also important to capture contextual information is required to understand the digital material and for it to be useful.

DROID stands for Digital Record Object Identification. It's a free software tool developed by The National Archives that will help us to automatically profile a wide range of file formats. For example, it will tell us what versions we have, their age and size, and when they were last changed. It can also provide us with data to help you find duplicates.

Profiling file formats will help us to manage our information more effectively. It helps us to identify risks (and therefore plan mitigating actions). It can also help us to save money, for example by supporting data reduction.

AVP Fixity

About

Widely used in organisations large and small, Fixity is a free utility for automated monitoring and reporting on the data integrity of stored files. Fixity scans a folder or directory and creates a manifest of the files, including their file paths and their checksums, against which a regular comparative analysis can be run. Fixity monitors file integrity through the generation and validation of checksums, and file attendance through monitoring and reporting on new, missing, moved and renamed files.

Download link

https://www.weareavp.com/products/fixity/#fixity-download

User guide

https://www.weareavp.com/wp-content/uploads/2018/07/Fixity_v1.2_UserGuide.pdf

System requirements

Application requires Java 1.7 or higher installed (JRE or JDK).

Mac OS or Windows.

Business case for installation

Digital Preservation looks specifically at the activities necessary to preserve, and to ensure continued long-term access to digital material. Passive preservation is not an option - allocating a priority to the preservation of digital material much more urgent than for paper archives. Unlike paper, a digital material which is not selected for active preservation treatment at an early stage in its existence will very likely be lost or unusable in a few years' time.

Increasingly we are creating and collecting digital material. This includes:

- Digitised content of collections where we hold the original;
- Digitised content of collections where we don't hold the originals;
- Born digital collections, which have been created and managed electronically. Common examples of born digital objects are photographs taken with a digital camera, an email or a text document.

A key element of digital preservation is file fixity and data integrity. Fixity measures such as checksums can record and regularly monitor the integrity of each copy of the digital material. This helps detect corruption or loss. Systems are also needed to protect digital material from unauthorised or accidental change.

Widely used in organisations large and small, AVP Fixity is a free utility for automated monitoring and reporting on the data integrity of stored files. Fixity scans a folder or directory and creates a manifest of the files, including their file paths and their checksums, against which a regular comparative analysis can be run. Fixity monitors file integrity through the generation and validation of checksums, and file attendance through monitoring and reporting on new, missing, moved and renamed files.

CSV Validator

About

CSV Validator is a CSV validation and reporting tool developed by The National Archives which implements CSV Schema Language. Released as Open Source under the Mozilla Public Licence version 2.0. The CSV Validator will take a CSV Schema file and a CSV file, verify that the CSV Schema itself is syntactically correct and then assert that each rule in the CSV Schema holds true for the CSV file. It can also be combined with a CSV file exported from a DROID report to detect duplicate digital files based on their checksum and integrity checks.

Download link

http://digital-preservation.github.io/csv-validator/#toc3

User guide

http://digital-preservation.github.io/csv-validator/

https://openpreservation.org/blog/2019/05/28/droid-report-as-basis-for-collection-integrity-checks/ (on using it for integrity checks)

System requirements

CSV Validator is predominantly written in Scala 2.11 and runs on any platform with a Java Virtual Machine (JVM). The Validator toolset provides:

- A stand-alone command line tool.
- A desktop application (a simple Swing GUI).
- A library that can be embedded into your own Scala project.
- A library that can be embedded into your own Java project, as it also provides native Java 7 interfaces.

Business case for installation

Digital Preservation looks specifically at the activities necessary to preserve, and to ensure continued long-term access to digital material. Passive preservation is not an option - allocating a priority to the preservation of digital material much more urgent than for paper archives. Unlike paper, a digital material which is not selected for active preservation treatment at an early stage in its existence will very likely be lost or unusable in a few years' time.

Increasingly we are creating and collecting digital material. This includes:

- Digitised content of collections where we hold the original;
- Digitised content of collections where we don't hold the originals;
- Born digital collections, which have been created and managed electronically. Common examples of born digital objects are photographs taken with a digital camera, an email or a text document.

A key element of digital preservation is file fixity and data integrity. Fixity measures such as checksums can record and regularly monitor the integrity of each copy of the digital material. This helps detect corruption or loss. Systems are also needed to protect digital material from unauthorised or accidental change. CSV Validator can be combined with a CSV file exported from a DROID report to detect duplicate digital files based on their checksum and integrity checks.

Teracopy

About

TeraCopy is a file transfer utility designed as an alternative for the built-in Windows Explorer file transfer feature. Its focus is data integrity, file transfer reliability and the ability to pause or resume file transfers. It can copy or move files from one location to another without changing the created date or modified dates of the file. Various checksums can also be done to verify that the files are the same after copying.

Download link

http://www.codesector.com/teracopy

User guide

https://codesector.kayako.com/category/3-teracopy

System requirements

Mac OS:
TeraCopy 1.0 - Mac OS X 10.12 Sierra or later
Windows:
TeraCopy 3.26 - Windows Vista/7/8/10 and Windows Server 2008/2012/2016
TeraCopy 2.3 - Windows XP/Vista/7/8/10 and Windows Server 2003/2008/2012/2016

Business case for installation

Digital Preservation looks specifically at the activities necessary to preserve, and to ensure continued long-term access to digital material. Passive preservation is not an option - allocating a priority to the preservation of digital material much more urgent than for paper archives. Unlike paper, a digital material which is not selected for active preservation treatment at an early stage in its existence will very likely be lost or unusable in a few years' time.

Increasingly we are creating and collecting digital material. This includes:

- Digitised content of collections where we hold the original;
- Digitised content of collections where we don't hold the originals;
- Born digital collections, which have been created and managed electronically. Common examples of born digital objects are photographs taken with a digital camera, an email or a text document.

It is important that we can transfer digital content reliably. Teracopy will allow us to:

- Identify errors in transfers easily.
- Verify files after they have been copied to ensure that they are identical.
- Preserve date timestamps – Teracopy keeps the original time and date of files when copying.

Quick View Plus

About

Quick View Plus 2020 is a desktop file viewer which allows users to view 300+ file formats without the need for the applications they were created in. It costs $99 for an annual subscription.

Download link

https://avantstar.com/downloads

User guide

https://avantstar.com/ittrium/reference/A1xd7x1x66y1x33fx1x67y1x37ex1x67y8xee3x8x1/QuickViewPlus2020FactSheet.pdf

System requirements

Compatible Operating Systems:
Windows 10 (64- and 32-bit)
Windows 8 (64- and 32-bit)
Windows 7 (64- and 32-bit)
64-bit Integrations
Microsoft Internet Explorer 11
Microsoft Outlook 2010 - 2019
Microsoft Windows Explorer
32-bit Integrations
Microsoft Internet Explorer 11
Microsoft Outlook 2010 - 2019
Microsoft Explorer
Plug-ins:

Adobe Acrobat
Required Disk Space:
120 MB

Business case for installation

Digital Preservation looks specifically at the activities necessary to preserve, and to ensure continued long-term access to digital material. Passive preservation is not an option - allocating a priority to the preservation of digital material much more urgent than for paper archives. Unlike paper, a digital material which is not selected for active preservation treatment at an early stage in its existence will very likely be lost or unusable in a few years' time.

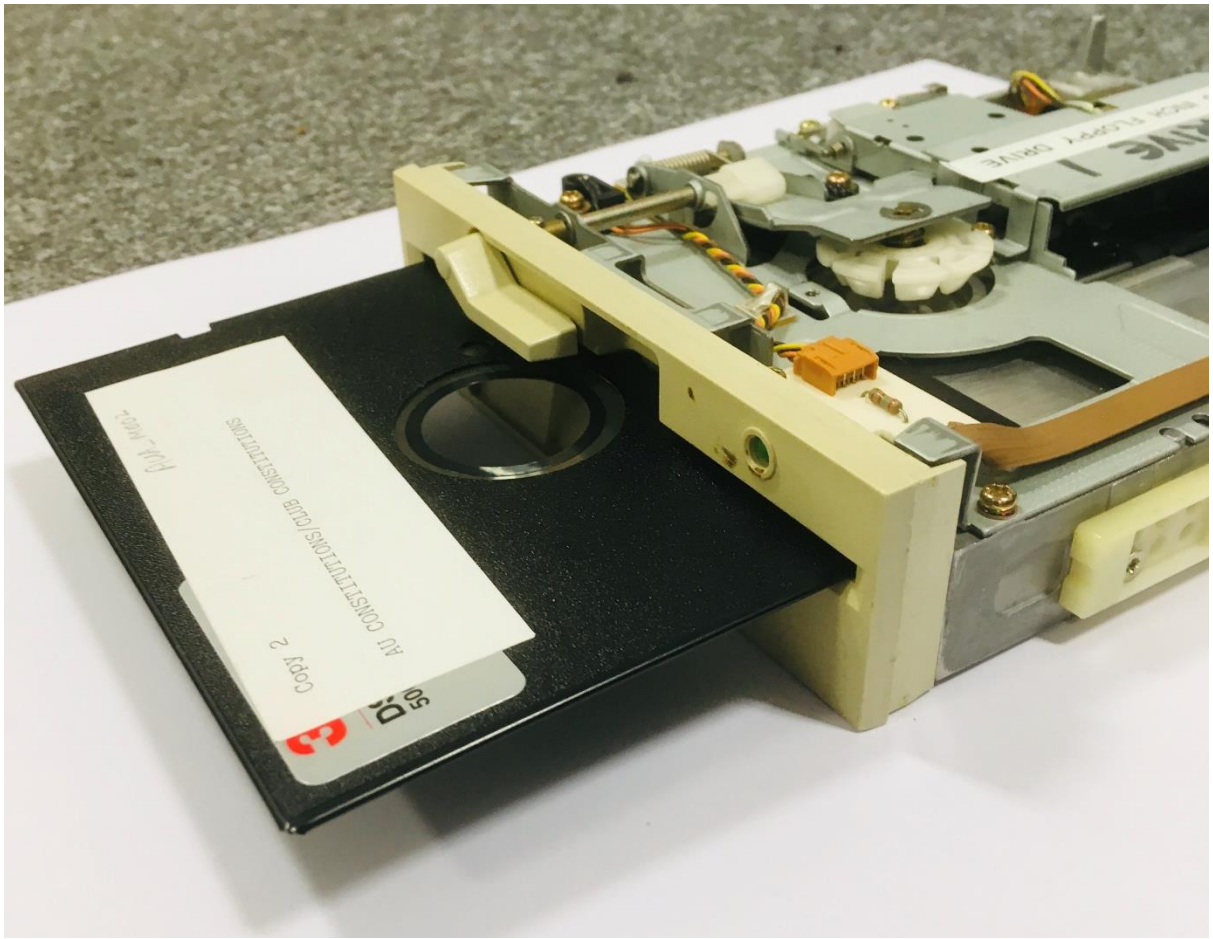
Increasingly we are creating and collecting digital material. This includes:

- Digitised content of collections where we hold the original;
- Digitised content of collections where we don't hold the originals;
- Born digital collections, which have been created and managed electronically. Common examples of born digital objects are photographs taken with a digital camera, an email or a text document.

Quick View Plus will allow us to view over 300 file formats from our born digital collections without having to purchase and install lots of different software. It could also be used by researchers as a single piece of software to access and view the majority of our born digital collections.

Appendix C – Photographs of hardware

USB Writeblocker. Inserted between the PC and a disk drive it will prevent the PC inadvertently altering the contents of the disk.



5 1/4" floppy disk in a disk drive. This can be attached to a PC via the Kryoflux which can read the disk and provide an image of it.

3 1/2" floppy disk in a disk drive connected to Kryoflux. The disk drive is powered from the mains, the Kryoflux is attached to a PC.

Hard Drive Caddy with hard drive inserted