Archives West Midlands

Accessioning Born-digital material

E: archiveswestmidlands@gmail.com

www.archiveswestmidlands.wordpress.com

Policy Owner(s)	Archives West Midlands Board of Trustees
Version	1.2
Prepared by	Emma Hancox
Approved by	
Date approved	
Review date	31st October 2020

Contents

Acknowledgements

The preparation of this policy was made possible thanks to funding provided by The National Archives' Sustainability Fund and Archives West Midlands.

Elements of this work draw upon existing examples of documentation from across the sector. It has also benefitted greatly from guidance provided by The National Archives and the assistance of Archives West Midlands member services in sharing their existing guidance.

Aims

The objectives given for the project were:

1. Metadata and file format guides.
2. Review of accessioning policies with a particular focus on strengthening policies around the appraisal of digital material.
3. Production of guidance for donors and depositors around the care of their digital records (in preparation for transfer).

Approach

This report has been created as part of a project funded through The National Archives' Sustainability Fund and Archives West Midlands. It should be read in conjunction with the Digital Preservation Policy template created by Charlene Taylor which was another component of the project.

The approach taken was to review documentation produced by other archives and also to read advice provided online to ascertain best practice. The main source used was the Digital Preservation Coalition's handbook which is freely available online.[1] The Paradigm Workbook on Digital Private Papers was also used.[2]

An email was sent to Archives West Midlands member services asking them to share guidance they have for donors and depositors about looking after born-digital material ready for transfer and also asking for any statements from their accessioning policies regarding born-digital material especially around appraisal.

Three other archives were contacted who undertake digital preservation activities. These were from different types of organisations; the Bodleian Libraries, the Borthwick Institute for Archives and Gloucestershire Archives.

Two archives from the West Midlands region responded to say they do have documentation (some of it in draft form) and were willing to share this.

Documentation from the following five services was reviewed:

Staffordshire and Stoke on Trent archive service
Gloucestershire Archives
Warwickshire County Record Office
Bodleian Libraries, the University of Oxford
Borthwick Institute for Archives, the University of York

[1] Digital Preservation Handbook, 2nd Edition, **http://handbook.dpconline.org/**, Digital Preservation Coalition © 2015. Date accessed 24/08/2018.
[2] Paradigm: Workbook on Digital Private Papers. http://www.paradigm.ac.uk/workbook/ Date accessed 24/08/2018.

OAIS reference model

When embarking on digital preservation work it is useful to gain a basic grasp of the OAIS (Open Archival Information System) model. The model provides a shared vocabulary that can be used by information professionals and IT staff to communicate when working on digital preservation projects and is adopted by digital preservation systems such as Preservica. A brief introduction is available in the Digital Preservation Handbook[3] and the Paradigm Workbook[4]. A more detailed introduction is given in a Technology Watch Report produced by the Digital Preservation Coalition.[5]

Review of accessioning policies

When reviewing the documentation and reading online guidance, a number of areas came to the fore as being important for consideration and these are highlighted in the report.

Identifying the service's Donors and Depositors

Who the key depositing and donating groups are for each repository will affect the policies and approach adopted helping to plan ahead. Possibilities include the parent organisation or authority (if working in a local authority), external organisations, the Diocese, the Coroner, Magistrates' Courts, Parish Councils, Schools, Businesses and individuals. This information can be included in a digital preservation policy.

A key point that came up is that if records are being taken in from an organisation, it may be necessary to liaise with an IT contact or Records Manager. Their name and contact details can be requested on a digital deposit form.

Advice for Donors and Depositors in caring for their records

A number of archives provide advice to donors and depositors on how to organise and look after their digital files in preparation for transfer. This assists the acquisition process as the files received will be easier to manage. An example document, based on advice created by other archive repositories, has been created to show the types of information that can be included **(see Appendix 1).**

---

[3] Standards and Best Practice, Digital Preservation Handbook https://www.dpconline.org/handbook/institutional-strategies/standards-and-best-practice, Date accessed 19/11/2018.

[4] *Introduction to OAIS*, Paradigm, http://www.paradigm.ac.uk/workbook/introduction/oais.html, Date accessed 19/11/2018.

[5] Brian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)*, DPC Technology Watch Report 14, 02 October 2014, https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file. Date accessed 19/11/2018.

Advising on accepted file formats

Different approaches were taken by archive services with regard to acceptable file formats. For example, one archive service mentioned that they preferred proprietary formats and that their approach was to define a list of acceptable formats for long term preservation, although they had not managed to do this yet.

Another archive service gave an overview of the types of files they could preserve in terms of general categories such as:

Image files
Word processing files
Simple spreadsheet files
Sound files

The service mentioned that they are not able to preserve email in-boxes, websites, podcasts/live streamed content, complex spreadsheets and databases at present. They recommend that a potential donor or depositor contacts them if they have digital content in any of these more problematic formats. The same service mentioned that if a depositor wishes to transfer files that were created in a format that is now proprietary or obsolete, they should also submit a copy in an open format.

An alternative approach was shown by a university archive service; *'We've never really specified acceptable file formats and metadata requirements and I'm not sure we would want to in the future. We tend to work with personal archives and the variety of formats is usually too great for us to do so. We'd also expect file format normalisation to be something that we would need to carry out rather than donors, many of whom won't be the original creators of the material or have the technical expertise to do so.'*

What can be taken from these examples is the fact that an archive service should not feel obliged to take file formats that it cannot confidently preserve at the current time and should be honest with donors and depositors about the types of files that they can accept. It is also important to assess material on a case by case basis, for example, if a depositor is able to provide sufficient support or funding, this might enable an archive to develop ways of dealing with challenging formats.

In terms of whether to accept only open or proprietary formats and whether to migrate proprietary formats to open formats, the Digital Preservation handbook provides a useful summary of the issues inherent in this.[6]

[6] *File Formats and Standards*, Digital Preservation Handbook, https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards Date accessed 24/08/2018.

The acquisition process

Digital Deposit Form

Some of the archives had a specific digital deposit form to be completed and signed by the depositor when digital material is given to the archive. The two areas of importance this covers are the collection of metadata to aid future management of the materials and the transfer of any rights that need to be obtained for the preservation of the material. Metadata components used need to be selected by the individual archive service bearing in mind their needs and what they can reasonably ask or expect the donor or depositor to provide.

Guidance from the DPC handbook emphasises this; 'alongside the standard procedural documents an organisation may wish to create a suite of standard depositor agreements and licences to aid in the negotiation process. These will be particularly useful in ensuring that the minimum permissions and intellectual property rights required for preservation are granted. Without sufficient licence agreements an organisation may find itself in possession of digital collections that it does not hold the rights to actively preserve or provide access to. These may also be complemented by guidance notes for depositors that set out requirements for material to be transferred and accompanying documentation.'[7]

The handbook also raises an important point regarding the deletion of supporting files that do not form part of the supporting record. 'If the transfer includes content that is essential to the understanding of the records but does not constitute a record itself there should be an agreement that the organisation can delete those files when their content has been captured for use elsewhere (for example as metadata for the records).'[8]

Metadata at the point of acquisition and best practice

The DPC handbook emphasises the importance of metadata describing the records which will be of use in the future.

The key elements it specifies should be included are:

- a verifiable manifest consisting of a list of the file and folder names and checksums/fixity values for each file.
- the size of the files (with a total volume)
- a list of the file formats

[7] *Acquisition and Appraisal*, Digital Preservation Handbook, https://www.dpconline.org/handbook/organisational-activities/acquisition-and-appraisal Date accessed 24/08/2018.
[8] *Acquisition and Appraisal*, Digital Preservation Handbook, https://www.dpconline.org/handbook/organisational-activities/acquisition-and-appraisal Date accessed 24/08/2018.

- a statement detailing any IPR associated with the records

The Paradigm project, which was specifically related to private papers, included an example transfer form **(see Appendix 2)**. This form includes the most important elements but is also relatively straightforward. It is recommended that Archives West Midlands member services use this as a basis and augment it with other metadata elements which meet their own requirements.

Metadata standards

There are a number of standards that can be drawn upon for recording metadata during the process of acquiring and preserving born-digital materials. For descriptive metadata these include Dublin Core and Metadata Object Description Schema (MODS) and for administrative metadata these include PREMIS and Schema for Rights Declaration METSRights. The Paradigm workbook provides a comprehensive overview of the different metadata options[9], however it has not been possible to explore these further during the time available for this project.

In future work it is suggested that a metadata standard is drawn upon to promote consistency and interoperability between Archives West Midlands member services. Bearing in mind the different archives represented, it would be difficult to select metadata elements that would be applicable and suitable for use by all services. It would be beneficial for member services to have some involvement in deciding which metadata elements would be applicable to their institution and collections.

Metadata for authenticity

One of the elements mentioned in the Digital Preservation Handbook that should be collected on acquisition is a checksums or fixity value for each individual file. Ideally this should be provided by the donor or depositor at the point of transfer. As stated in the handbook 'if no manifest was made prior to transfer it may be impossible to check that the files have retained their integrity.'[10] One way of obtaining this information is to ask donors or depositors to download DROID (The National Archives' file profiling tool)[11]. One of the archive services whose documentation was reviewed indicated on their digital deposit form that they would like donors and depositors to download DROID, provide the archive service with the name of the DROID report and send it to them with the form. As well as creating a list of useful information about files, DROID has the added capacity of

[9] *Rights Metadata for personal archives*, Paradigm, http://www.paradigm.ac.uk/workbook/metadata/rights-metadata.html, Date accessed 24/08/2018.
[10] *Acquisition and Appraisal*, Digital Preservation Handbook, https://www.dpconline.org/handbook/organisational-activities/acquisition-and-appraisal Date accessed 24/08/2018.
[11] *File Profiling Tool (DROID)*, The National Archives, http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/, Date accessed 24/08/2018.

creating a hash that can be used to verify the files once they are moved around, however this function needs to be enabled.

Obtaining metadata in this way has strong benefits for the archive as it does not have to be captured manually. Although many donors and depositors might find the use of DROID challenging, it is recommended that Archives West Midlands member services promote its use in situations where it is judged that the donor or depositor has the skills, or could learn to use it.

Use of tools

Gloucestershire Archives uses the freely available Bagger tool developed by the Library of Congress so that depositors can create their own submission information package[12]. There are many advantages to using this tool as it allows the depositor to include delivery information, it includes a list of the contents and it has the capacity to validate the package automatically providing a checksum that the material can be compared against when it arrives. Bagger is also compatible with Archivematica.[13] Use of this tool at Gloucestershire Archives has been more of an aspiration than something that has been actively used by depositors (beyond initial tests). Its intended use is for County Council employees transferring organisational records in the future where the volume of records is likely to be larger.

If there are no institutional restrictions on downloading software, it is recommended that services download and experiment with tools such as DROID and Bagger to gain familiarity with them in order to establish whether they are suitable for donors and depositors to use and if so, so that they can be explained to these groups who may have questions when using them.

Legal Considerations and negotiating rights

There is a useful summary of the legal issues involved in digital preservation activities in the Digital Preservation handbook which it is advisable to read.[14] It is important to ensure that permission is granted for the files to be copied both for preservation and access. Hardware and software needed to render the files also have their own IPR considerations.

Deletion of the original files by the donor/depositor

An important point that came up during the research is how to advise the donor or depositor on deletion of copies of the records that they hold.

---

[12] *Library of Congress/Bagger*, GitHub, https://github.com/LibraryOfCongress/bagger, Date accessed 24/08/2018.
[13] *Bag Ingest*, Archivematica, https://wiki.archivematica.org/Bag_ingest, Date accessed 24/08/2018.
[14] *Legal Compliance*, Digital Preservation Handbook , https://www.dpconline.org/handbook/institutional-strategies/legal-compliance, Date accessed 24/08/2018.

One of the archive services actively involved in digital preservation specifies that donors or depositors should not delete their copies of the files until the ingest process has been completed (e.g. until it has been verified that there are no problems with the copies and they have been successfully transferred to the digital archive environment). They say that they will aim to do this within 20 working days.

Another service also emphasised the need for donors and depositors not to delete files until confirmation has been received that the files have been received successfully. This follows guidance given in the DPC handbook.

Appraisal

Selection and appraisal was mentioned by several archives in relation to digital material, but in all cases was viewed as the same as for paper material. Some statements are given below:

'The process of selecting digital records for preservation is the same as that for paper records, that is, it is carried out in accordance with our collecting and appraisal policies.'

'We don't have specific guidance on appraisal for digital material, we would base appraisal on content (same as paper) rather than format.'

'I don't think we actually do appraisal of digital before it is accepted - or at least not in any way that is different to physical material - so there may be a discussion with the donor to talk about the types of content and then we will inform them which bits we consider to have longer term archival value and only ask for those to be deposited.'

In guidance documentation given to donors and depositors on the management of their files, accurate file naming, version control and well-organised folder structures was promoted. It is important to offer this guidance as it assists with the appraisal process.

Although in practice archivists do not seem to be differentiating between digital and traditional records in terms of appraisal, the DPC handbook states that digital material should be assessed differently and that the 'preservation implications' of formats etc. should be taken in to account. It says that *'while many of the same principles from the traditional preservation environment can usefully be applied, policies and procedures will need to be adapted to the digital environment.'*

There is a very detailed and useful interactive Decision Tree for the appraisal of born-digital material in the Digital Preservation Handbook.[15] The Decision Tree demonstrates some of the issues that are specifically relevant to digital material and should be considered. It is recommended that services make use of the Decision Tree when appraising digital material.

[15] *Decision Tree*, Digital Preservation Handbook, https://www.dpconline.org/handbook/organisational-activities/decision-tree. Date accessed 24/08/2018.

Information in the handbook on ingest suggests that appraisal may take place at a significantly later date than the transfer of material to the archive and it suggests that this can be particularly delayed for large born-digital collections which suggests collections of organisational records.

It also suggests that during the process of ingest items can be reviewed as to whether they conform to file format specifications or whether they meet the organisation's collecting policy. Assessing material at this later date means that permission to dispose of files should be obtained at the point of transfer to allow for later disposal if necessary.

Methods of Transfer

Archives had a variety of methods for accepting digital records. Methods included on a CD/DVD, USB or via email (recommendations on size limitations were given).

One archive service mentioned that files attached to an email should be zipped as this allows important information such as file creation dates to be retained. Some services used and an in-house FTP service, but this appeared to only be the case with archives based in universities.

Another option for transfer, mentioned by the DPC handbook, is via a paid for third party solution such as a cloud-based file sharing solution.

One archive service suggested that donors or depositors may wish to encrypt or password protect digital media before sending it and then supply the password separately if the files contained sensitive data. Alternatively another service requested that any password protection be removed prior to donating or depositing digital records.

As mentioned already one archive service encourages donors and depositors to use the Bagger transfer tool to create a submission information package (SIP). This involves creating a check-sum that can be validated when the SIP is received.[16]

Summary

This report has highlighted some initial steps that can be adopted by member services to take steps towards accepting born-digital material in to their collections. The advice document in Appendix 1 for donors and depositors who may be considering giving digital records to an archive service can be used before further procedures or processes have been put in place. An initial step that can also be taken is to create a digital deposit transfer agreement by adapting the one from the Paradigm project included in Appendix 2. If there are no institutional restrictions on downloading software, then it would also be beneficial to trial open source software tools such as DROID to assist with collecting information on file formats at the point of accession.

[16] *Bagit: Transferring Content for Digital Preservation*, Library of Congress, http://www.digitalpreservation.gov/multimedia/videos/bagit0609.html. Date accessed 24/08/2018.

Regarding the other project objectives, the Digital Preservation Coalition's interactive Decision Tree is a useful existing tool that would be useful to consider new digital accessions against. It has not been possible to recommend specific file formats, however, member services can learn from the approach taken by other archives in being specific about groups of formats that can or cannot be taken and also assessing new accessions on a case by case basis when more challenging formats are offered. It is suggested that interoperability is prioritised with regard to the use of metadata elements and that all archive services have input in deciding which metadata elements would be most suitable for their collections.

Appendix- 1: Managing Digital Records (advice for donors and depositors)

This guidance is to help you ensure that your digital records are well-managed and are suitable for long-term preservation in an archive.

1. **Name your files clearly**

   Like the file title on a paper folder, file names act as a short descriptive caption for the content of your digital records. Clear naming will help you and archivists who may look after the files in the future, to understand them.

2. **Use the yyyy-mm-dd format for dates**

   It is recommended that you use the format yyyy-mm-dd for recording dates. For example 1st April 2005 would be 2005-04-01. File management tools understand this date format and will be able to list your files accordingly.

3. **Create a meaningful directory structure**

   Use directories and folders to organise your files in a logical manner and use the structure consistently.

4. **Only use alphanumeric characters in filenames**

   The only characters you should use in your filenames are lowercase a-z, 0-9, hyphens (-) and underscores (_).

   Full stops should only be present to separate a filename from its file extension e.g. 20170302meetingminutes.doc

   Spaces can cause problems when files are moved between operating systems and when some digital preservation tools are used so it is good practice to replace spaces with an underscore (_).

   Although Windows is not, some operating systems are case sensitive, so it is advisable to always use the lower case in file names.

5. **Weed your files regularly**

   Go through your files on a regular basis to get rid of duplicate or unnecessary files.

   There may be cases where previous versions of a document need to be consulted, but in many cases only the final version will need to be kept.

6. **Use version control**

   Where several versions of a document do need to be kept, use version control. This means including a version or draft number in the file name so that it is clear from looking at the files which are the earliest and latest versions.

7. **Make sure your files are 'self-documenting'**

   All files should have a title, author and date but document history, purpose and status are also useful.

   If possible, include this information on the first page of the document. This will make it much easier for you and anyone preserving the document in the future to understand it.

   If you are using a database or spreadsheet you may need to include interpretation of codes or abbreviations and relationships between data tables may need to be described.

   If your files include pictures, you may need to save images, captions or titles separately.

   It is easier to record this type of information at the point of creation (you may forget important details if you leave it until a later date).

   Data about your documents can sometimes be stored in the files themselves. Many desktop applications include a properties option where you can record title, author, keywords and comments about your files.

8. **Save your files in the original format that you created them in**

   Files that are in the original format they were created in are easier to preserve for the long term. 'Derived' files may be poorer quality and have lost some of their original qualities.

   For example if you are giving photographs to an archive it is preferable for them to be transferred in their original format rather than a different format which may be at a lower resolution.

9. **Use file types that are widely used and supported**

   These will be easier to preserve in the long term than obscure or less well-documented file types.

   The best file types to use are those that are open, well-documented and uncompressed.

10. **Store your records safely**

    Store your files on your hard drive. Your computer should run regular background checks on data integrity and make you aware of any problems.

    You can 'back your data up' by storing it on static storage media (such as CDs or USB). However, it is not advisable to store your files on static storage alone as this type of media is liable to deteriorate, may become obsolete and you may not be able to detect problems until it is too late.

    Other ways of backing-up your data include 'cloud storage' and a portable hard drive. It is important to store your back-ups in a different location to the original copies of your files.

Ensure that anti-virus software on your computer is up-to-date so that any problems will be detected.

11. **From the moment you create your files, plan ahead for the future of your digital data! Don't leave it too late!**

Heavily relied on the Borthwick Institute's guidance available on their website:

Managing your digital material: some good practice guidelines for donors and depositors, Borthwick Institute for Archives, https://www.york.ac.uk/media/borthwick/documents/donors/Preparation_of_digital_material_for_deposit_revised.pdf, Date accessed 27/09/2018.

'Top Tips for Managing digital records' provided by Gloucestershire Archives were also referred to:

Preserving Digital Records, Guidance for Donors and Depositors, Gloucestershire Archives, https://www.gloucestershire.gov.uk/media/1946/guidelines_digitalrecods_donorsdepositor_v_1_2-49339.pdf, Date accessed 27/09/2018.

Appendix- 2: Example transfer list form from the Paradigm Project.

Transfer List, Paradigm, http://www.paradigm.ac.uk/workbook/appendices/transfer-list.html , Date accessed 27/09/2018

1. Transfer List

This transfer list is not a legal document. Its purpose is to record details about the material transferred to the archive to ensure that the authenticity of the material can be audited in the future; to collect details, such as usernames and passwords, needed to access the material; to record contextual information which will be used when compiling finding aids; and to record details about the suggested closure period of records series.

Owner details

Name	
Address	
Telephone	
Email	

Paradigm staff details

Name	
Position	
Address	
Telephone	
Email	

Terms of transfer:

The materials detailed in the schedule of transferred material below are transferred under the terms and conditions set out in the Paradigm project deposit agreement.

2. Schedule of Transferred Materials

Media ref. no. *[Ref. no of CD-R or USB stick, e.g. CD-R-1]*	
MD5 checksum(s) *[record values of MD5 checksum(s)]*	
Extent *[In bytes]*	
Technical description *[description of file formats, passwords]*	
Content Description *[Covering dates, subjects, record types, etc.]*	

Restrictions

Please specify any restrictions to access and use. Does the material contain any confidential items or personal data?

Signatures

Signature of owner or of owner's authorised representative	
Name of signatory	
Date	
Signature of archivist	
Name of signatory	
Date	

Appendix 3: Contents of a DROID report

DROID extracts metadata about your files and folders. A CSV export will contain the following columns and you will see the main sub-set of these on screen within the GUI:

1. ID
2. parent ID
3. unique resource identifier (URI)
4. file path (Resource column in the GUI)
5. filename
6. identification method (signature, container signature or extension)
7. status DROID: user guide July 2017 Page 9 of 21
8. file size (Size column in the GUI)
9. type (file, folder or container)
10. file extension (Extension column in the GUI)
11. last modified date (Last modified column in the GUI)
12. extension mismatch warning
13. hash (Hash column in the GUI)
14. file format count (Ids column in the GUI)
15. PRONOM unique identifier for the file format (PUID column in the GUI)
16. mime-type (Mime type column in the GUI)
17. file format name (Format column in the GUI)
18. file format version (Version column in the GUI)

3.11 Hash (checksum) DROID can optionally generate a content hash or checksum of the contents of each file and container file, using either the SHA 1, SHA 2 (256) or MD5 algorithms. A content hash is a long unique string of numbers and letters that can be used to uniquely identify the content of the file. It is extremely unlikely that two different files will have the same content hash. Content hashes can be used to detect files with duplicate content, or can be linked to forensic hash databases to find or exclude files which are widely used (and therefore not unique to your organisation) or which contain illegal content. Hashes can also be helpful when moving files around. Once moved, if you re-calculate the hashes and check against the previous values you can verify that the copying process has not caused any changes at the bit level.

For more information about DROID see *DROID: user guide*, The National Archives, http://www.nationalarchives.gov.uk/documents/information-management/droid-user-guide.pdf, Date accessed 27/09/2018.