University of Kent

Digital Preservation Strategy

1. Vision

The Digital Preservation Strategy defines how Special Collections and Archives will work to preserve the digital objects and records within their care. It is part of an integrated preservation approach that will ensure Kent's digital collections are authentic, reliable, useable and accessible for the long-term.

Special Collections and Archives will implement and embed this approach to digital preservation by 2026. It will enable Kent's unique and distinctive digital collections to be preserved for the benefit of scholarship and society.

2. Scope

This strategy applies to digital objects deposited with, donated to, or created by Special Collections and Archives and held for the purpose of long-term preservation. This includes:

- Objects that are born-digital, created in a 'native' digital format.
- Materials that are digitised, digital surrogates created from the original non-digital format. To preserve and/or for access purposes.

3. Context

3.1 This strategy aligns with the Digital Preservation Policy.

3.2 Special Collections and Archives hold digital objects in various file formats. These objects may form a complete digital collection, or be part of a hybrid collection with analogue material.

3.3 Special Collections and Archives accepts new accessions of digital objects from depositors.

3.4 Digital objects are stored on the University of Kent network in a hierarchical file structure. This includes master copies of digital objects and access copies of digital objects, as well as some administrative and procedural documentation.

3.5 Special Collections and Archives work with colleagues to provide high quality digitisations of analogue material as supported by current capabilities and resource.

4. Activities

We aim to operate at Level 2 of the NDSA Levels of Preservation[1] by 2026. The following activities provide a framework for how we will achieve this. A breakdown of tasks aligned with each functional area of the NDSA levels is included in Appendix 2.

These activities are organised based on the DCC Digital Curation Lifecycle Model[2].

4.1 Create or receive

4.1.1 Collect and receive digital objects in accordance with the Special Collections and Archives Collection Management Policy.

4.1.2 Create guidance regarding file formats, file arrangement and transfer of digital objects to be shared with depositors pre-transfer of digital objects.

4.1.3 Regularly review the Special Collections and Archives deposit form to ensure it accommodates current information pertaining to digital object deposit.

4.1.4 Ensure potential technology restraints, requirements or data protection issues of digital objects are identified before deposit. Any such issues and the decisions associated with them will be recorded in the collection deposit form.

4.1.5 Encourage depositors of born-digital objects to record information regarding the context and environment of the object's creation in the collection deposit form.

4.1.6 Encourage depositors to supply manifests with any digital objects deposited.

4.1.7 Share a list of recommended file formats with depositors before transfer of material to the archive.

4.1.8 Ensure that digital objects created by Special Collections and Archives are digitised following agreed workflow processes and adhere to relevant standards (see appendix 3).

4.1.9 Ensure appropriate administrative, descriptive, technical, structural and preservation metadata is created for digitised objects in line with current digital collections cataloguing guidance. This may be in the form of metadata embedded in the digital object, or metadata within a catalogue record in the archival management system (currently Axiell Calm).

[1] https://ndsa.org/publications/levels-of-digital-preservation/
[2] https://www.dcc.ac.uk/guidance/curation-lifecycle-model

4.2 Appraise and select

4.2.1 Appraise and select digital objects in accordance with the Special Collections and Archives Collection Management Plan.

4.2.2 Periodically reassess digital objects to ensure they still meet our collecting criteria.

4.2.3 Periodically reappraise digital objects to identify any emerging digital preservation risks using the Digital Asset Register.

4.3 Ingest

4.3.1 Use a "forensic" PC[3] with virus checker and write blocker capability to ingest born-digital objects stored on external media.

4.3.2 Create a disk image of all born-digital accessions transferred using external media.

4.3.3 Characterise born-digital objects on ingest, through use of the file format identification tool *Droid,* to ensure the content is fully understood and to validate authenticity of the file.

4.3.4 Create a hash value on ingest of all born-digital accessions, through use of the file format identification tool *Droid*, and record this in the metadata.

4.3.5 Verify the contents of digital objects using the supplied manifest or listing provided by the depositor.

4.3.6 Store master copies of digital objects in file share \\files.kent.ac.uk\Shared\ARCHIVES on the University of Kent storage area network (SAN).

4.3.7 Migrate files for digital objects transferred in a non-preferred, proprietary or obsolete format, to the recommended format at point of ingest, wherever possible for preservation and access purposes. Record decisions and the rational to migrate formats in the metadata and ensure significant properties remain intact.

4.3.8 Where file formats are migrated, we will retain the original file format alongside the master copy, in addition to the new format.

4.3.9 Assign an appropriate reference code to all new accessions according to Special Collections and Archives current policy.

4.3.10 Create access copies of all master files on ingest and store in the appropriate file share on the University of Kent SAN.

[3] https://www.dpconline.org/handbook/technical-solutions-and-tools/digital-forensics

4.4 Preservation action

4.4.1 Ensure all future accessions of digital objects have, as a minimum, an accession record and a collection descriptive record in the archive management system (currently Axiell Calm). This should include the storage location.

4.4.2 Ensure all digital objects have sufficient administrative, descriptive, technical, structural and preservation metadata at an agreed standard[4]. See Cataloguing of born digital archives (appendix 4) for information on cataloguing practices.

4.4.3 Ensure metadata created in the archive management system (currently Axiell Calm) adheres to the recognised international metadata standard *ISAD(G)*[5].

4.4.4 Ensure file formats chosen for digitisation and migration are open-source, non-proprietary, uncompressed formats that are widespread in use wherever possible (see appendix 3).

4.4.5 Define a list of recommended file formats accepted by Special Collections and Archives to ensure that any forward migrations are straightforward and that value for money is obtained from investment in equipment and skills. However, these recommendations will not be prescriptive and we will work with depositors to ensure transfer of their digital objects in the most appropriate file format for the material (see appendix 3).

4.4.6 Assess file fixity of all digital objects on a regular, scheduled basis by generating and comparing hash values.

4.4.7 Replace any files found to have been damaged or changed with back-up copies.

4.4.8 Store fixity data for any new accessions in descriptive catalogue records.

4.4.9 Create and maintain a Digital Asset Register for all digital object collections.

4.4.10 Identify and monitor any risks associated with digital objects using a digital preservation risk assessment model[6].

---

[4] http://www.loc.gov/standards/premis/v3/
[5] https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf
[6] https://wiki.dpconline.org/index.php?title=Digital_preservation_risks

4.5 Storage

4.5.1 Maintain a network storage area with a suitable file structure for digital object master copies (\\files.kent.ac.uk\Shared\ARCHIVES\Master).

4.5.2 Maintain a network storage area with a corresponding file structure for digital object access copies (\\files.kent.ac.uk\Shared\ARCHIVES\Access).

4.5.3 Define and control access requirements for both the master and access network storage areas. Ensure network storage is secure and that access is limited to those working directly with the digital collections to prevent malicious or accidental damage.

4.5.4 Work with IT colleagues to ensure network storage meets digital preservation requirements outlined below:

- Three complete copies of each digital object.
- Each copy of the digital object is stored in separate location, including one copy in a geographically separate location.
- At least one copy is stored on a different storage media.
- Storage media is regularly and proactively refreshed according to industry standards.
- Any transfer of digital objects to new media is carried out using robust and proven methods to avoid damage or data loss.
- Define and document back up practices in the event that a digital object needs to be replaced.

4.5.5 Document these storage systems and measures, and maintain clear procedures on using them.

4.5.6 Minimise the use of legacy storage media wherever possible.

4.6 Access, Use and Reuse

4.6.1 Ensure licence and rights information is captured in the descriptive catalogue record of all digital objects, and in embedded metadata where appropriate.

4.6.2 Digital objects will be made as openly available as possible, subject to copyright restrictions and according to the agreements made with the depositor.

4.6.3 Digital objects with access restrictions will only be made available under supervision on a secure, un-networked PC in the Special Collections and Archives reading room.

4.6.4 Requests to access digitised material will be managed by the Special Collections and Archives team, and assessed on a case-by-case basis.

5. Document review date

Digital Preservation standards and practices change regularly, and this strategy will be monitored and reviewed regularly to ensure it reflects current best practice.

Appendix 1 – Glossary of terms

Accessible: continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for. Digital Preservation Handbook

Accession: to take intellectual and physical custody of materials, often under legal or policy authority. Dictionary of Archives Terminology

Authentic: The digital material is what it purports to be. In the case of electronic records, it refers to the trustworthiness of the electronic record as a record. In the case of "born digital" and digitised materials, it refers to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. Confidence in the authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made. Digital Preservation Handbook

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 Handbook

Digital object: An object composed of a set of bit sequences. Reference Model for an Open Archival Information System (OAIS)

Discovery: The process of searching for and identifying potentially relevant materials. Dictionary of Archives Terminology

Format obsolescence: Formats evolve as users and developers identify and incorporate new functionality. New formats, or versions of formats, may introduce file format obsolescence as newer generations of software phase out support for older formats. When software does not provide for backwards compatibility with older file formats, data may become unusable. Digital Preservation Handbook

Ingest: To accept content or metadata into a repository or database. Dictionary of Archives Terminology

Integrity: The quality of being whole and unaltered through loss, tampering, or corruption. Dictionary of Archives Terminology

Long-term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, on the information being held in [a repository]. This period extends into the indefinite future. Reference Model for an Open Archival Information System (OAIS)

Metadata: Information which describes significant aspects of a resource. Metadata are required to successfully manage and preserve digital materials over time and to ensure essential contextual, historical, and technical information are preserved along with the digital object. Digital Preservation Handbook

Preservation: The professional discipline of protecting materials by minimizing chemical and physical deterioration and damage to minimize the loss of information and to extend the life of cultural property. Dictionary of Archives Terminology

Reliable: The quality of being dependable and worthy of trust; the quality of being consistent and undeviating. https://dictionary.archivists.org/entry/reliability.html

Useable: Being in an accessible location and usable condition. International Council on Archives

Appendix 2 – Actions aligned to NDSA Level 2

NDSA functional area	Level 2 description	Activities associated
Storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function.	4.5.4 Work with IT colleagues to ensure network storage meets digital preservation requirements. 4.5.5 Document these storage systems and measures, and maintain clear procedures on using them.
Integrity	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content.	4.3.1 Use a forensic PC with virus checker and write blocker capability for ingesting born-digital objects stored on external media in to the archive 4.3.4 Create a hash value on ingest of all born-digital accessions using the file format identification tool Droid. 4.3.5 Verify the contents of the digital objects using the supplied manifest or listing provided by the depositor 4.4.6 Assess file fixity of all digital objects on a regular, scheduled basis by generating and comparing hash values. 4.4.8 Store fixity data for any new accessions in descriptive catalogue records and maintain a separate database of fixity data for all digital objects held by Special Collections & Archives. 4.5.4 … Any transfer of digital objects to new media is carried out using robust and proven methods to avoid damage or data loss
Control	Document the human and software agents authorized to read, write, move, and delete content and apply these.	4.5.3 Define access requirements for both the master and access network storage areas. Ensure network storage is secure and that access is limited to those working directly with the digital collections to prevent malicious or accidental damage. No one member of staff should be able to access all copies of data.
Metadata	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural).	4.1.9 Ensure any digital objects created through digitisation activity have sufficient administrative, descriptive, technical, structural and preservation metadata associated with them, as per internal digital collections cataloguing guidance. This may be in the form of metadata embedded in the digital object itself, or metadata stored in the catalogue record in the archival management system (currently Axiell Calm).

		4.4.1 Ensure all future accessions of digital objects have, as a minimum, an accession record and a collection descriptive record in the archive management system (currently Axiell Calm). This should include the objects storage location. 4.4.2 Ensure all digital objects have sufficient administrative, descriptive, technical, structural and preservation metadata at an agreed standard . 4.4.3 Ensure metadata created in the archive management system (currently Axiell Calm) adheres to the recognised international metadata standard ISAD(G) . 4.4.9 Create and maintain a Digital Asset Register for all digital object collections. 4.6.1 Ensure licence and rights information is captured in the descriptive catalogue record of all digital objects.
Content	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices.	4.1.2 Create guidance regarding file formats, file arrangement and transfer of digital objects to be shared with depositors pre-transfer of digital objects. 4.1.3 Review the current Special Collections & Archives deposit form and ensure it accommodates information pertaining to digital object deposit. 4.1.7 Share a list of recommended file formats with depositors before transfer of material to the archive 4.3.3 Characterise all digital objects using the file format identification tool Droid on ingest to ensure the content is fully understood and to validate authenticity of the file. 4.3.5 Verify the contents of the digital objects using the supplied manifest or listing provided by the depositor. 4.3.7 Where digital objects are transferred in a non-preferred, proprietary or obsolete format, migrate the file to the recommended format on ingest wherever possible for preservation and access purposes, ensuring all significant properties remain intact. Record any decisions made to migrate formats, and the rationale for the decision, in the objects metadata. 4.4.5 Define a list of recommended file formats accepted by Special Collections & Archives to ensure that any forward migrations are straightforward and that value for money is obtained from any investment in equipment and skills wherever possible. However, these recommendations will not be prescriptive and we will work with depositors to ensure transfer of their digital objects in the most appropriate file format for the material.

Appendix 3 – Digitisation technical implementation strategy

1. Purpose and scope of this document

This document sets out:

- Definitions
- Digitisation and digital preservation principles adopted
- Types of file storage used
- Digital file formats used for digitisation and digital transfer
- File naming
- Metadata
- Equipment used for digitisation and digital transfer
- Basic workflows for digitisation and digital transfer
- Appendix – Audio and moving image formats, risk, and file format normalisation

This document was initially produced in P1397 Digital Asset Management. Background work was carried out as part of P1228 Digital Archiving and Preservation Strategy and P1396 Digital Asset Management Review. It has been reviewed here in line with the Digital Preservation Policy (2022) and Digital Preservation Strategy (2022).

2. Definitions

Digital preservation is about preserving access over time. It is a proactive task, rather than being reactive, and involves policies, strategies and procedures, to ensure the continued preservation and access of digital (both born-digital and digitised) content, even with the challenges of media failure and technological change[7]. Digital preservation requires organisational commitment, setting responsibilities, and resources.

Digitisation is the process of converting information from a paper or analogue source media into a digital file for storage, accessibility and preservation. There are also instances where digital files (often on physical items) require reformatting due to obsolesce of technology or physical degradation (such as the audio carriers and formats MiniDisc and DAT).

'Born-digital' files are those without an analogue equivalent; they are created and managed in digital form[8]. They can include digital photographs, documents (such as MS Word and PDF), electronic records (held in databases, spreadsheets), digital audio and video.

[7] Association for Library Collections and Technical services, 'Definitions of Digital Preservation', accessed online 12 January 2015, <http://www.ala.org/alcts/resources/preserv/defdigpres0408>
[8] Defining "Born Digital" An Essay by Ricky Erway, OCLC Research, accessed online <http://www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf>

3. Digitisation and digital preservation principles adopted

- Where possible, file formats chosen for digitisation and migration will be open-source, non-proprietary uncompressed formats, which are widespread in use.
- The range of formats will be narrow to ensure that any forward migration is easier and to ensure that value for money is obtained from any investment in equipment and skills.
- We will digitise to the highest quality 'uncompressed' format possible in order to minimise any loss in conversion from analogue sources.
- A fundamental goal of digital preservation is to ensure the integrity of the files is not compromised (of both the masters and copies) and, thus, the 'fixity' of the files. One of the easiest ways to determine whether digital records have become corrupt is to use a Checksum. A Checksum is a small data value calculated from a digital file and is used to validate the file; a checksum can be generated upon accession into the archive, and then recomputed at later dates (and certainly every time the record is moved, for example). If the checksums match then the digital record has not been altered. The MD5 checksum/cryptographic hash is one of the most widely used.

From a preservation perspective, material within Special Collections and Archives is of mixed risk, with the paper records generally of low risk and audiovisual material at high risk.

4. Types of storage

4.1 Archival storage

Information Services digital files (both preservation and access) are stored on the University's SAN (Storage Area Network) on the server \\files.kent.ac.uk\Shared\ARCHIVES\Master.

4.2 Access file storage

The storage of the access digital files will depend on the historic arrangements of creating access images as well as the file format and copyright permissions.

Special Collections and Archives (including British Stand-Up Comedy Archive)

Access files (typically jpegs, mp3s and mp4s) will be generated. Access jpegs will be created locally, and stored on Calm servers, where rights allow, to be made available through the Calmview[9], or made available via the SC&A reading room PC terminal. Access audio and moving images files will be stored locally on \\files.kent.ac.uk\Shared\ARCHIVES\Access. Access to audio and moving images files will be through an SC&A reading room PC terminal.

British Cartoon Archive

Large size jpeg access files, for example, created by the British Cartoon Archive through digitisation processes, will be stored locally on \\files.kent.ac.uk\Shared\ARCHIVES\Access. Smaller 'access' jpeg files will be created locally, and transferred to the Calm server so that access can be provided via the British Cartoon Archive catalogue (http://archive.cartoons.ac.uk).

4.3 Storage systems

IS is responsible for the curation of digital collections specified in the IS Library Collections' Collections Policies. We use a modular system to provide for robust and sustainable curation during the lifecycle of digital assets, using OAIS as our reference model. The Research Technologies Development Group monitors systems and services. Business and Technical owners are named in the IS Service Catalogue.

OAIS Function	Systems	Purpose	Location
Ingest	Calm for Archives Symphony for library materials RAID 1 array	Interface for creators or library to deposit metadata descriptions and associated files	Hosted: Axiell Hosted: SirsiDynix Local: RAID 1array
Data management	Calm (ISAD (G)) Symphony (MARC, RDA, DCRM(B))	Holds metadata in various required schema appropriate to type of collection	Hosted: Axiell Hosted: SirsiDynix
Archival storage	Calm server SAN - ARCHIVES	Holds digital objects (metadata linked together with digital files) for specified storage	Hosted: Calm-access file copies, derivative \\files.kent.ac.uk\Shared\ARCHIVES.
Access	LibrarySearch (Sirsi Dynix Enterprise) Calmview	Deliver discovery and access, serve files via the web	https://librarysearch.kent.ac.uk https://archive.cartoons.ac.uk/ https://archive.kent.ac.uk/

[9] Either http://archive.kent.ac.uk or via the catalogue searchable at https://www.kent.ac.uk/library/specialcollections/index.html until the data has been transferred to Calm.

5. File format evaluations for Special Collections and Archives

The following issues are recommended when selecting file formats for long-term preservation[10], and when selecting the formats SC&A collections will use, these issues have been considered. It should be noted that formats having all these attributes are not always available and substitutes may need to be found.

5.1 Sustainability considerations:

- *Disclosure*: does complete technical documentation exist for the format? Is the format a standard?
- *Quality*: is the quality high enough to anticipate future use and avoid risk of quality loss over time (e.g. through migration).
- *Transparency:* can the file be easily understood and identified by file analysis tools.
- *Self-documentation:* does the format have good metadata support? i.e. can it contain descriptive and technical metadata either generated by the creating application or entered manually? The inclusion of metadata will provide useful information about the file, and help to provide information about its context.
- *Non-proprietary origin*: often referred to as an 'open' format. Are there patents related to this format which could have a direct impact on the long-term sustainability of files produced in this format?
- *Adoption/ubiquity:* how widely has the format been adopted for digital preservation? If widely used then it is less likely to become obsolete quickly; there are also more likely to be commercial tools for capture, and if necessary future migration.

5.2 Cost considerations:

- What costs are incurred in implementing and maintaining access to the files?
- What are the storage and network costs?
- What are the required software and hardware costs?

5.3 Implementation considerations

- How hard is it (technically) to implement the format?
- Is there wide availability of tools for implementation? Are they interoperable?
- *Implementation costs*: what are the costs associated with the software tools and equipment required to produce the files? What are the costs associated with storing

[10] Criteria selected based on the following: FADGI (Federal Agencies Digitization Guidelines Initiative) (e.g. as seen at http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_pt3_20141202.pdf), http://www.paradigm.ac.uk/workbook/preservation-strategies/file-preservation.html, http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf, http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf

the files and making them available over internal networks?

5.6 Audio files

Collections cared for by Special Collections and Archives contain audio material on a number of formats on magnetic and optical media, currently: audio cassette, MiniDisc, DAT (digital audio tape), and audio CD.

At the time of writing there is a substantial amount of audiovisual material which is generally more endangered than conventional paper documents due to the deterioration of the audio carrier and obsolescence of both the carriers and the replay equipment[11]. These include analogue magnetic carriers, such as sound cassette tapes and open reels, and digital audio carriers, such as DAT (Digital Audio Tapes), MiniDisc, audio CDs and CD-Rs. None of these carriers/formats are suitable for long-term preservation due to the obsolescence of the format and replay equipment, and deterioration inherent in the format (CD-Rs, for example).

The format recommend for capturing and digitising audio is the codec LPCM (Linear Pulse Code Modulation) in a BWF/Broadcast Wave format (.wav) wrapper[12]. The BWF/Broadcast Wave format (file extension also .wav) is an extension of the WAVE format which incorporates metadata in the file header (through the addition of a 'bext' chunk). The BWF format is recommended by the IASA[13] (the International Association of Sound and Audiovisual Archives) and has been adopted for Special Collections and Archives collections for the following reasons:

The minimum recommended archival standard for digitising analogue material is to sample at 48kHz and an encoding rate of at least 24 bit[14]. A higher sampling rate of 96kHz is also recommended but for spoken word recordings there is no advantage to sampling at this higher rate and it would inevitably increase the size of files to no advantage.

---

[11] Joshua Ranger, 'For God's Sake, Stop Digitizing Paper', < https://www.avpreserve.com/blog/for-gods-sake-stop-digitizing-paper-2/>, accessed 2 December 2015. Dietrich Schüller, 'Audio and Video Carriers: Recording principles, storage and handling, maintenance of equipment, format and equipment obsolescence', TAPE (Training for Audiovisual Preservation in Europe), < http://www.tape-online.net/docs/audio_and_video_carriers.pdf>, accessed 2 December 2015.

[12] This format is recommended by the International Association of Sound and Audiovisual Archives *Guidelines on the Production and Preservation of Digital Audio Objects*, IASA-TC 04, 2.8.1 <http://www.iasa-web.org/tc04/key-digital-principles>

[13] The BWF format is recommended by the IASA as long as the limitations of embedded metadata are kept in mind. *Guidelines on the Production and Preservation of Digital Audio Objects*, IASA-TC 04, 2.8.2 <http://www.iasa-web.org/tc04/key-digital-principles>. The BWF format was standardised by the EBU (European Broadcasting Union), including the 'broadcast extension' chunk (bext) which contains the minimum information expected to be needed for broadcast production.

[14] IASA Technical Committee, *Guidelines on the Production and Preservation of Digital Audio Objects*, ed. by Kevin Bradley. Second edition 2009. (= Standards, Recommended Practices and Strategies, IASA-TC 04), Chapter 2, 'Key Digital Principles', <http://www.iasa-web.org/tc04/key-digital-principles>.

Some material accessioned into Special Collections and Archives will already exist in a digital form, either in files or on physical carriers. Depending on the format, this material will need to be normalised (transferred) to an archival digital file format, either due to obsolescence of the playback equipment or obsolescence of the format itself. Some digital formats will be captured/transferred at lower sampling and encoding rates than the recommended archival standard as the format has already been captured in a discrete signal. Increasing the sample or encoding rate during digital transfer will only increase 'noise' and the file size, but not increase quality.

See Appendix A for a table of audio formats within collections cared for by Special Collections and Archives and the normalisation undertaken per format.

5.7 Moving image files

Special Collections and Archives has moving image records held on analogue magnetic tape carriers (currently VHS, U-matic,) digital magnetic tape carriers (DVCAM, MiniDV) and optical media (DVD, CD-R). None of these formats/mediums offer an archival solution and all content will need to be digitised or digitally transferred.

The codec and wrapper chosen will be dependent on the format of the original carrier.

Moving image material on an analogue magnetic tape carrier

Archivally, the best way to digitise moving image material would be by capturing as uncompressed 4:2:2 10-bit codec[15]; in an uncompressed video the entire information of the undigitised source is captured. However uncompressed video results in vast file sizes (c. 100GB per hour of video), which not only impacts on storage costs but also requires computers with adequate processing power to open and view files, and requires transfer of the files across the network (so a lot of network bandwidth)[16]. Some large institutions are digitising their video as uncompressed (the BBC captures uncompressed files and wraps as MXF) but many broadcast and heritage institutions have made the decision to use a lossy or lossless compression codec.

Special Collections and Archives evaluated possible codecs through online research and from discussions with peers at other heritage institutions in the UK and abroad. The favoured codec is FFv1.3[17]; this open source codec offers lossless compression but, unlike other lossless compression codecs (such as jpeg2000) there are fewer compatibility issues with software[18]. FFV1.3 can also be captured directly, without transcoding, and is currently the codec chosen by Archivematica, a digital preservation system, as the 'preservation' format for moving image content[19]. FFv1.3 supports technical metadata which is required to read and play the file, but descriptive metadata would be carried by the wrapper.

---

[15] 'Digitizing Video for Long-Term Preservation: An RFP Guide and Template' http://library.nyu.edu/preservation/VARRFP.pdf

[16] http://download.das-werkstatt.com/pb/mthk/info/comparison_video_codecs_containers.html#codec_uncompressed

[17] FFv1 entry on Wikipedia, <https://en.wikipedia.org/wiki/FFV1>

[18] SC&A had already come up against jpeg2000 incompatibilities with image jpeg2000 files.

[19] Any moving image content ingested with Archivematica is normalised as FFv1/LPCM in a Matroska (MKV) wrapper (the original format of the material is also maintained).

The two wrappers considered for Special Collections and Archives digitised collections were AVI and Matroska (MKV). AVI is a proprietary format but is very widely supported and documented; the 'self-documentation' facility is described as 'acceptable' by the Library of Congress[20] . One of the main proponents of FFv1, the Austrian Mediathek, wraps their files in AVI as they believe it to be the best supported container for video[21]. Matroska is an open source wrapper. It has been adopted as the wrapper for digital video preservation by The National Archives (UK) and also the City of Vancouver Archives. The 'self-documentation' facility is described as 'good' by the Library of Congress[22], and one of the reasons for its adoption by TNA was the extended tag support for metadata[23].

Matroska has been chosen by Special Collections and Archives, to continue using open formats, because of its ability to provide support for metadata, and because of the traction the wrapper is gaining within heritage and memory institutions[24].

Moving image material on a digital magnetic tape carrier

Magnetic tapes containing digitally stored materials can present in a variety of physical formats (DVCAM, MiniDV, DVCPRO-M etc) and contain moving image material encoded in a variety of digital formats (DV, DVCPRO-HD, HDV etc). In these cases, wherever possible, we will retain the original file format and capture data with a direct data transfer from the tape to avoid degrading the material in any way through a format switch. We will however rewrap the file as an MKV. There is no advantage in format shifting or converting the footage away from the original, indeed this would actually increase the file sizes and therefore storage burden. The only circumstances where we would consider conversion is if the original format is overly specialised or for some other reason we feel it will be a compatibility risk in the short or long term.

5.8 Image files

Documents, photographs, flyers, pamphlets, posters and negatives within Special Collections and Archives are being captured as uncompressed Tiff files.

Material within Special Collections and Archives is captured as uncompressed Tiff files. When the British Stand-Up Comedy Archive was established as a Project in 2015 the JPEG2000 file formats was evaluated as a possible file format for digitising photographs and paper document. The advantages of JPEG2000 include the fact that it is an ISO standard (ISO/IEC 15444-1), that it uses compression so much less storage space is required (and therefore associated network costs should be lower and access to files quicker), and also that JPEG2000 has built-in thumbnailing and tiling (so the same file can be used for both preservation and access).

[20] http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_pt1_20141202.pdf
[21] http://download.das-werkstatt.com/pb/mthk/info/FAQ-digital_video_archiving.html. The US's National Archives and Records Administration (NARA) also use AVI as their wrapper for digital video preservation.
[22] http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_pt1_20141202.pdf
[23] Email exchange with Ian Henderson, TNA, 6 February 2015
[24] The papers and presentations from a symposium held in Berlin in August 2016 (No Time To Wait)

After initial tests on JPEG2000 with the software which was available to IS, images can be scanned as JPEG2000 but not viewed as additional software, or software plugins, are required. There are also wider industry concerns with the supportable future of JPEG2000 for smaller institutions, there are few tools available for generation of JPEG2000 files suitable for archives and those that are, have unique challenges.

Therefore, for the present, images (documents, photographs, flyers, pamphlets, posters and negatives) are being created as uncompressed Tiff files.

6. Filenaming

The filenaming of digital assets has varied according to the collection or overall collecting Archive (for example, British Cartoon Archive, BSUCA). Across all collections filenames have been 'descriptive' (they bear some relation to the content) and have generally relied on the catalogue or accession number.

Records digitised more recently (after 2015) typically have a suffix of '-M' or '-A' to denote whether they are a Master or Access file. Where a single item has multiple corresponding files (such as two sides of a cassette tape) the filename will include an A or B.

6.1 Special Collections and Archives

Digital files within SC&A collections (such as the Playbills and Windmill collections) have been filenamed according to the final numeric element of the catalogue reference. For example:

UKC-MIL-MUG-BW.537289 – the digital filename is 537289.tif (or .jpg)

UKC-POS-BDGN.0593995 – the digital filename is 0593995.tif

The numeric elements of the catalogue reference are unique so there should be no duplication of filenames across collections.

For newer archival material, where barcodes are not typically applied, collections use the reference code from the catalogue record. For example, MWT-MUS-1-0039.

6.2 British Cartoon Archive

Digtial and digtiised files with the British Cartoon Archives have been filenamed according to their unique reference code. This consists of the accession number, which acts as a prefix, followed by a sequential number. For example:
Ben Jennings: accession number BJD. Digital files are BJD0001, BJD0002, etc.
Ernest H. Shepard: accession number ES. Digitised files are ES0001, ES0002, etc.

6.3 Born-digital records

Born-digital records are filenamed using the reference number followed by '-D' to denote they are a born-digital record.
Example:. MER-002-001-D-A

6.4 Uncatalogued material

Typically, material will not be scheduled for digitisation until it has been catalogued or a unique reference number has been applied. However, in a situation where uncatalogued material needs to be digitised, the following should be used to create the filename:

- The collection REF number
- Plus a sequential number in 4 digit format (e.g. 0001) - to ensure this number is unique, check which number should be used next by looking at the collection folder on the ARCHIVES file share
- Use a numerical suffix if multiple pages/leaves, separated with an underscore.

7. Metadata for collections

7.1 Metadata held in catalogue systems

Metadata about Special Collections and Archives material is currently held in a variety of catalogue systems. The storage of metadata depends on historic cataloguing arrangements.

British Cartoon Archive – Axiell Calm (Client 1)

Special Collections and Archives (archival collections) - Axiell Calm (Client 2)

Special Collections and Archives (book Collections) - SirsiDynix Symphony

7.2 Embedded metadata

Metadata about individual digital files (or items) will also be embedded within the files, for a number of reasons, including ones of digital preservation, accessibility, and to assist with cataloguing. The benefits of embedded metadata include:

- Embedded metadata makes a file 'self describing'[25]; it provides data to the systems that will preserve the digital files, and to those systems which provide access to the digital files;
- It can assist in disaster recovery and removes the risk of losing the link between metadata and the digital file;
- It provides information for the people involved in the content's life cycle.
- Metadata embedded in born-digital master and digitised master files can also be transferred to 'access' copies (generated at a lower specification/resolution) which assist in user accessibility.

Embedded metadata does have its limitations and it is not that flexible (there is a fixed number of fields, often with fixed character limits), which is why we are using a dual approach of embedding some metadata, but also keeping this information (and in fact more metadata) externally in catalogues.

[25] Stephanie Christensen and Doug Dunlop, ' The Case for Implementing Core Descriptive Embedded Metadata at the Smithsonian', p.10

As a minimum, all digitsed material should include the following information in the IPTC profile:

IPTC – Contact

- Creator: University of Kent
- Headline: [Artist/Creator], [?] Collection
- Description: [same as above, or complete reference (time allowing)]

IPTC – Status

- Source: Special Collections & Archives, University of Kent
- Copyright notice: [© copyright holder information (where known/applicable)]

7.3 Audio files

There is no official metadata standard for embedded metadata in audio files, although there are general recommendations and popularly used standards, such as the bext chunk and RIFF LIST INFO chunks for wave files[26], and the ID3 metadata type for MP3 files (defined as the 'de facto' standard)[27].

When digitising audio material, BWF MetaEdit[28], a free open-source tool developed by AV Preserve and the Federal Agencies Digitization Guidelines Initiative, will be used to embed metadata. BWF MetaEdit adds metadata to two chunks within the .wav file: in the bext chunk (which records information such as a description, information about the origination of the file, and the coding history) and in the RIFF LIST INFO chunk (which records information such as title, archival location, copyright, genre, keywords, software used). The advantage of BWF MetaEdit for wav files is that it is free and open-source, and also has other functions such as batch editing (where you can edit multiple .wav files at once) and batch import and export functions to/from Excel and CSV files, which has the potential to increase the speed at which these recordings can be catalogued on CALM.

When our archival master is a compressed format (such as an MP3 on a data or audio CD which has been deposited), we are using tools such as Adobe Bridge[29] (or Adobe Audition's metadata tools[30] as we transfer audio) to embed metadata in the ID3 format.

For access surrogate files we are creating MP3 files from wav/BWF .wav masters. Although wave files do not support ID3 tags, upon the creation of access MP3 files from BWF masters, metadata is automatically mapped from LIST INFO chunks to ID3 tags in the MP3 files.

---

[26] A 2009 report produced by AVPreserve points out that bext and LIST INFO chunks are not flexible enough or sufficient, but that their adoption makes them the best option for now: http://www.digitizationguidelines.gov/audio-visual/documents/AVPS_Audio_Metadata_Overview_090612.pdf

[27] *Encyclopaedia of Multimedia,* ed.Borko Furht, p 574 (https://books.google.co.uk/books?isbn=0387747249)

[28] BWF MetaEdit is a free, open source tool developed by the Federal Agencies Digitization Guidelines Initiative (FADGI, a 'a collaborative effort by federal agencies to define common guidelines, methods, and practices for digitizing historical content') and AVPreserve (a US consultancy firm which provides support for audio digitisation and digital preservation): http://www.digitizationguidelines.gov/audio-visual/documents/help_home.html.

[29] Adobe Bridge has the advantage of being able to batch edit metadata to multiple files at once.

[30] The software programme used for recording and editing audio files; this is included within the Adobe Creative Cloud (CC) license.

7.4 Image files

Upon capturing, technical metadata is automatically embedded in the digital file (this is called Exif data) by the scanner or digital camera. Exif data is technical metadata and includes information on the scanner/digital camera used, the date captured, the number of pixels, the colour space, etc.

Exif metadata tags are mainly employed to record technical metadata. We also embed basic descriptive metadata into digital files using the IPTC Core schema (an XML based schema packaged in the XMP (Extensible Metadata Platform) format[31]) which is a combination of descriptive and rights based metadata. This data is embedded using Adobe Bridge[32], which has the advantage of being able to batch edit metadata to multiple files at once.

8. Equipment

Equipment for digitisation has been sourced in a number of ways. Equipment has been sourced externally and purchased (for example, from approved suppliers through UniBuy or in the cases of obsolete technology through online marketplaces). Equipment has also been sourced internally, particularly in the case of obsolete technology, from Schools within the University of Kent. Equipment has also been purchased using external funds. In 2021, Special Collections & Archives service was awarded over £110,000 to assist with the digitising of its collections, through the Capability for Collections (CapCo) fund, administered by the Arts and Humanities Research Council (AHRC), part of UK Research and Innovation (UKRI). This funding enabled us to purchase new high-end equipment from the Phase One Cultural Heritage solutions.

Equipment being used at the time of writing is listed below.

8.1 Digital Image capture

Epson Perfection V850 Pro
Epson Expression 10000 XL
Phase One iXG Camera System (with 45mm, 55mm and 120mm lenses)
Phase One motorized copy stand
Capture One software
Phase One Film Capture Stage and carriers
Phase One DT V Book Cradle

8.2 Audio digitisation

Cassette deck: Denon DN-790R
Minidisc deck: Sony MDS-JE530

---

[31] XMP was introduced as a format by Adobe, but since 2012 has been an ISO standard; see http://www.iso.org/iso/home/news_index/news_archive/news.htm?refid=Ref1525. A useful guide to IPTC and XMP can be found on the JISC Digital Media website: http://www.jiscdigitalmedia.ac.uk/guide/putting-things-in-order-links-to-metadata-schemas-and-related-standards. XMP tags are based on Dublin Core tags; further information at http://www.photometadata.org/META-Resources-metadata-types-standards-XMP.

[32] Included within the Adobe Creative Cloud (CC) license

8.3 Moving image digitisation

Sony HVR-1500 (for DVCam and MiniDV)
Panasonic AG-7500 (for VHS)
DPS 575 (AV synchronizer)

Workflows

Workflows can be found on SharePoint: IS Learning and Research Resources: Special Collections & Archives - Digital workflows - All Documents (sharepoint.com)

Appendix: Audio and moving image formats, risk, and file format normalisation

The Preservation Self-Assessment Program (PASP) has been useful in assessing the risk of each format: https://psap.library.illinois.edu/format-id-guide.

Audio formats

Type	Format examples	Risk	Digitisation / transfer sample and encoding rates
Analogue magnetic audio tape formats	Compact audio cassettes	Medium (equipment is still readily available; prioritisation should be based on the tape condition and content)	**48kHz 24 bit**
	Reel-to-reel tapes	Medium (equipment is still available; prioritisation should be based on the tape condition and content)	**48kHz 24 bit**
Digital magnetic tape formats	Digital Audio Tape (DAT)	High (equipment is now obsolete and no longer manufactured)	DATs use lossless encoding. The audio information has already been captured in a discrete signal, so each DAT should be transferred at the original sampling and encoding rates (generally **16 bit** and **44.1kHz** or **48kHz**); as this was a format used for broadcast purposes the rates are generally

			recorded on the tape or tape box.
Digital	MiniDisc	High (equipment is now obsolete and no longer manufactured)	MiniDisc is a Sony magneto-optical disc format. They appeared in two forms: a mass replicated disc (which worked according to the principles of optical discs) and as a recordable and rewritable magneto-optical disc (both are now obsolete). All of the MiniDiscs in SC&A collections are the recordable and rewritable format. MiniDisc recordings employed ATRAC (Adaptive Transform Acoustic Coding) data compression (i.e. information encoded on MiniDiscs are compressed). ATRAC is a proprietary format, and minidisc players are now becoming obsolete. It is recommended that minidisc recordings are re-encoded as .wav files. Minidiscs sampled at **44.1kHz**, and each Minidisc recording should be captured at this rate.
	Audio deposited as files	Low	Some material will be deposited with SC&A in a digital file format. Files deposited in the WAV or MP3 format will be retained in their original file format. Although MP3 is not an archival format there is no immediate preservation risk; it is widely adopted and an open standard developed by the Motion Pictures Experts Group. Other digital audio formats deposited, such as WMA (Windows Media Audio), will be normalised to be WAV.
Optical media	Audio CD	Medium (playback equipment is readily available but optical media is subject to data loss through poor handling/storage or chemical	Audio CDs hold files at no higher than 44.1kHz 16 bit. The audio should be ripped and saved as an uncompressed WAV (at **44.1kHz 16 bit**).

		deterioration)	

Moving image formats

Type	Format examples	Risk	Digitisation / transfer sample and encoding rates
Analogue magnetic tape formats	VHS	Medium (playback equipment is readily available – although not at professional levels – but VHS can be susceptible to signal loss and damage from mould or binder deterioration; prioritisation should be based on the tape condition and content).	Losslessly compressed FFV1 1.3 /PCM audio
	U-matic	High (due to media and hardware obsolescence).	Losslessly compressed FFV1 1.3 /PCM audio. U-matic tapes cannot be digitised in-house due to a lack of equipment; digitisation is outsourced.
Digital magnetic tape formats	DVCAM	Medium (possible threat of obsolescence and DVCAM is subject to physical tape issues; prioritisation should be based on the tape condition and content).	Capture in DV (original format) wrap in MKV
	MiniDV	Medium (MiniDV is subject to physical tape issues but is also proprietary tape and is thinner	Capture in DV (original format) wrap in MKV

		tape than that in DVCAM).	
Optical media	DVDs	Low	With rewritable media (such as hard drives, floppy drives), or media which has inbuilt menu functionality (i.e. DVDs), we have created disc images, sector-by-sector copies, as part of the process of digitally preserving the original accession. A complete disk image (.iso file) serves as the preservation master, and from the iso file we have then created an access copy as an mp4 (h.264) file, using VLC, for use in our reading room.

Appendix 4 – Metadata for born-digital records

Special Collections & Archives:
Proposed approach for cataloguing born digital archives
Beth Astridge and Clair Waller – September 2020

In Special Collections & Archives we already have some born digital records that have been accessioned and catalogued as part of our collections. However, we have not used a consistent approach in how to catalogue born digital records which has resulted in some variations in the catalogue.

It is also clear that cataloguing born digital records requires a different approach and new consideration of some of our agreed processes for cataloguing records.

We might receive born digital records as part of deposited collections from external donors – and these might be entirely born digital collections or hybrid collections that include both physical and digital files. We might also increasingly receive digital collections from within the university as part of the University Archives.

This proposal is designed to explain why a different cataloguing approach is needed for born digital records, and set out some suggestions for how we could consistently approach cataloguing this type of material across all our collections.

Why are born digital records different?

- They are sometimes held in unstructured filing systems without any meaningful metadata structures
- They are sometimes held in filing systems that are multi-layered with folder, within folder, within folder, within folder, and then only one document in that folder!
- They might have been created in environments without any basic record-keeping practices applied
- They might contain a variety of datasets, computer code or algorithms, alongside more standard documents, spreadsheets, images and audio files
- We may not have access to any information about history, use and purpose of records and therefore we will be unable to communicate this to users as part of providing them with context about how and why the records were used and created
- There may be multiple versions of the same document and all of them need to be retained
- They might be corrupted or partially corrupted
- They might be huge in terms of size and number of records

The large size, variety and complexity makes it impractical to manage them in the same way we have previously with physical records.

Accessing born digital records

As well as receiving, storing and preserving these potentially vast collections of records – we also need to make them findable and understandable for our users.

The process by which users will want to search for this information will be different to how they have traditionally accessed archive collections. We usually curate the information for them – giving them context and structure – at collection and series level, and then the user will drill down to where they would like to look at the collection.

With born digital collections it is much more likely that users will carry out general text searches, find a record they are interested in and then look 'up' the structure to understand what it is and where it came from.

This means is it very important that our cataloguing and structure of the catalogue can support these different ways of accessing the collections – so that users can find things and understand more about the items as they search.

There is also a question about how they then access the record itself once they have found the reference in the catalogue entry – as with digital files there will be an expectation of a digital access method from wherever they are, as opposed to requesting to see items in the reading room. We need to plan for how we can deliver this form of access to improve the user experience.

A new approach

We need to consider how we can best provide information for the different cataloguing elements for born digital records to enable the user to understand the content and context of the collection. We have used the TNA position paper on cataloguing born digital records as a useful starter for how we might approach this. (https://www.nationalarchives.gov.uk/documents/digital-cataloguing-practices-march-2017.pdf)

The following contains a suggested approach for the different ISAD G cataloguing elements:

Level

We will provide contextual catalogue descriptions to series level. Then all individual digital objects (the record and its metadata) will be catalogued at file level (equivalent to the lowest reproducible format). This means that there is no folder structure below series level with born digital material catalogued at one flat level. The only exception for this may be for hybrid collections, where it may be necessary to add an additional sub-series level to give logical order to the collection (e.g. to ensure unique reference codes exist for physical and digital components).

For multi-level folder structures – we will record the original folder structure in the arrangement field, rather than replicate it digitally in the catalogue structure.

Reference

We need to consider how to continue to provide a unique reference to large numbers of records that still shows to which collection the items belong but does not necessarily reflect a folder structure in the same way we currently present. This could be that we use our standard collection abbreviation method alongside a randomly generated number for each digital file.

It would also be useful if we could indicate using the reference that the item is a digital file. We could do this by having -D (for digital!) at the end of the reference if it is a digital file. This approach is similar to that used at TNA who use /Z at the end for digital files. Calm interprets all '\' in the reference code field as breaks in the hierarchy, therefore we will use a hyphen here.

Collection/series/random reference/Digital file: SHIR/5/3432-D

Title

The title for each digital file would be the filename for that file.

The risk with this approach is that filenames often do not make logical sense – and therefore we would need to use the Scope and Content field to add extra information about the file where this is necessary.

Due to the potentially large number of files that could be ingested, individually curating the file titles may not be a sustainable practice. Additionally, changing the title could be seen as altering the original record and could impact the validation processes to check to deterioration over time. We therefore propose to not make changes to file titles, and use the original filename as the title in the catalogue.

Dates

Dates can be slightly complicated but ideally we would use the ‘last modified’ date as the date of creation in the catalogue. We might need to be careful to ensure that the process by which the depositor transfers the records to us does not alter the ‘last modified’ date.

We would explain in our catalogue descriptions using the ‘Note’ field, that for digital records the date field displays the ‘last modified’ date and that this is not necessarily the date the record was actually created, as this is not always possible to capture from the digital record.

Extent

In the extent field at series level we would like to identify not just the number of digital files in the series, but also the type of digital files there are. For example – rather than just noting 150 digital files, we should say 100 documents, 25 image files and 25 audio files. We feel that this would be better for the user who may be trying to find a specific type of file and this could help with their understanding of the content of the collection.

At file level – as it is a digital file and these are catalogued individually at file level – there will only be 1 file per catalogue entry, but it could be helpful here to include the size of the file and the file format.

Arrangement

This should be a mandatory field for all digital files to record the place of the digital file within its original folder structure. This context is extremely important for digital files, as the folder structure will not be replicated by the catalogue or in how it will be stored, to communicate the context in which the file was originally kept and help with searching and retrieval for the user.

For example - we could use the following description in the arrangement field:

This born digital record was arranged in the following file structure:

- Dame Stephanie/Speeches/1997-2000/1997

Scope and Content – Description

At fonds/sub fonds/series level we will use this field to provide the contextual information about the collection to enable the user to better understand what we have and decide if the content of the collection is of relevance to them.

At file level we can provide more information about the record – to support information already given in the title (filename) and arrangement (place in the structure) where necessary. This might be beneficial if the file title (filename) does not indicate the content of the file, and no further

Cataloguing of born digital archives

For rationale please see paper “Cataloguing born digital records” written by Beth Astridge and Clair Waller, September 2020

TNA position paper on cataloguing born digital records: https://www.nationalarchives.gov.uk/documents/digital-cataloguing-practices-march-2017.pdf https://www.nationalarchives.gov.uk/documents/digital-cataloguing-practices-march-2017.pdf

Standard ISAD(G) guidelines for archival cataloguing should be followed with the following local policies.

3.1 IDENTITY STATEMENT AREA

3.1.1 Reference code(s)

Use our standard collection abbreviation method alongside a randomly generated number for each digital file, followed by -D (for digital!) at the end. E.g. SHIR/5/3432-D

3.1.2 Title

Use the filename given by the creator/depositor of each file as the title.

Where the title lacks context or structure, you should use the Scope and Content field to add extra information about the file (see 3.3.1).

3.1.3 Date(s)

Ideally you should use the ‘last modified’ date as the date of creation in the catalogue. Where this has been overwritten by the transfer process you should use the content creation date or similar (this should be accessible in the embedded metadata of the file).

The note field should be used to state the origin of, and rationale for, the date recorded (see 3.6.1).

3.1.4 Level of description

Contextual catalogue descriptions will be produced down to series level. Below this digital objects (the record and its metadata) will be catalogued at file level (equivalent to the lowest reproducible format). There should be no folder structure below series level. The only exception for this may be for hybrid collections, where it may be necessary to add an additional sub-series level to give logical order to the collection (e.g. to ensure unique reference codes exist for physical and digital components).

Where digital objects are deposited with a multi-level folder structure we will record the original folder structure in the arrangement field, rather than replicate it digitally in the catalogue structure (see 3.3.4).

3.1.5 Extent and medium of the unit of description (quantity, bulk, or size)

At series level - identify the number and type of digital files in the series. For example – rather than just noting 150 digital files, we should say 100 documents, 25 image files and 25 audio files.

At file level - include the file format and size of the file (e.g. 1 digital file - Microsoft Word Document (23 KB)).

3.2 CONTEXT AREA

3.2.3 Archival history/Custodial history

Record the date of the migration and any preservation activity that has taken place.

3.3 CONTENT AND STRUCTURE AREA

3.3.1 Scope and content

File level – use to provide more information about the record to support information already given in the title (filename) and arrangement (place in the structure) where necessary.

3.3.4 System of arrangement

Mandatory for all digital files. Record the place of the digital file within its original folder structure.

For example:

This born digital record was arranged in the following file structure:

- Dame Stephanie/Speeches/1997-2000/1997

3.4 CONDITIONS OF ACCESS AND USE AREA

3.4.1 Conditions governing access

Note any access restrictions here. Digital objects with access restrictions will only be made available on the forensic PC in the reading room.

3.4.4 Physical characteristics and technical requirements

Use to record information about whether a record or collection of records contains corrupted or partially corrupted files, or if specific software is required to access the content.

3.6 NOTES AREA

3.6.1 Note

Use this field to record the type of date recorded and the rationale for this, and any other supplemental information not covered in other fields.

LOCALLY APPLIED FIELDS

Special Collections & Archives use the following local fields when cataloguing digital records:

Digital filename (user wrapped 5)

Use to record the name given to the master copy of the file. This should be the reference number of the file record followed by an ‘M’ (e.g. MER-002-001-D-M.doc).

Digital File location (user text 3)

Use to record the digital store location of the master file. Include the whole of the file path.

Fixity (user text 4)

Use to record the checksum has value of the file.

Access filename (user text 6)

Use to record the name given to the access copy of the file. This should be the reference number of the file record followed by an ‘A’ (e.g. MER-002-001-D-A.doc).