



Cambridge University Libraries Digital Preservation Strategy 2019 – 2024

Filename: STR_CUL_digitalpreservationstrategy_v1_20190415.pdf

Location: <https://doi.org/10.17863/CAM.33770>

Version: 1.0

Last Amendment: N/A

Strategy Owner/Sponsor: Deputy Director, Digital Initiatives

Strategy Contact: Director, Digital Transformation Programme

Prepared By: Digital Preservation Specialist – Policy and Planning (Polonsky Fellow)

Approved By: Senior Leadership Team

Date Approved: 31 January 2019

Next Review: 1 February 2020

Availability: Public

1. Summary

The Digital Preservation Strategy (the Strategy) guides digital preservation efforts over a six-year period, aligning primarily with CUL's organisational priorities around the digital environment. The implementation of technical infrastructure, organisational change and transitioning to more sustainable resourcing for digital preservation functions are key. If achieved, this ambitious Strategy will facilitate the initial step towards ensuring long-term access to digital content, by first ensuring its preservation.

The goals laid out this strategy are essential for meeting the needs of philanthropic and research funders, staff, students and researchers. Achieving them will also set CUL on a path towards the future requirement of Trusted Digital Repository (TDR) certification.

2. Introduction

Digital preservation is fundamental to Cambridge University Libraries' (CUL) stewardship of digital content. Digital preservation is currently considered one of CUL's highest risks. CUL is committed to working collaboratively, holistically and cross-departmentally in order to address digital preservation concerns over the coming six years. The Strategy applies to the main University Library as well as affiliate and dependent libraries.

In order for CUL to be recognised as a twenty-first century research library, to support world-class digital collections, and facilitate the education, learning and research needs of the University of Cambridge (the University), transitioning to Business-As-Usual (BAU) digital preservation functions are fundamental. For the past twenty years, CUL has taken a project-by-project approach to digital preservation, however this is no longer suitable. The Strategy is essential for driving forward CUL's capability to support the acquisition, management, preservation and provision of access to CUL's born-digital and digitised content.

In order to progress towards BAU digital preservation, a programme of work is necessary. This will set in place foundations for managing digital content from the point of acquisition, and implementing processes to carry out digital preservation actions on the digital content. These activities are expected of a library in the digital age. Under the banner of the Digital Transformation Programme (DTP), a portfolio of work is planned, which includes:

- Archives and Digital Preservation Foundation Programme (ADFPF);
- Born-Digital Archiving and Digital Preservation Lab project;
- Jisc Research Data Shared Services (RDSS) project;
- Cambridge University Digital Library (CUDL) Enhancement Programme.

While CUL's active digital preservation capabilities are yet to be established, this Strategy intends to incrementally shift CUL towards achieving TDR certification for CUL's whole 'digital preservation environment'.

This Strategy aligns with key CUL policy frameworks and strategies including:

- *Cambridge University Libraries Strategy 2019 – 2024*
- *Cambridge University Libraries Digital Preservation Policy*

The Strategy and other digital preservation planning will be governed by CUL's newly developing governance structures, with work led by Digital Initiatives and Strategy (DIS) as part of the DTP, in consultation with University Information Services (UIS) and stakeholders representing various CUL departments and libraries.

3. Purpose

This Strategy provides the framework for how CUL will guide its focus of digital preservation activities over the coming six years to establish infrastructure and manage support for at risk digital content. Aligning with CUL's and the University's strategic directions, this Strategy outlines the key drivers and priorities, defining specific goals that will be brought to fruition.

4. Vision, mission and goals

4.1. Vision statement

To ensure CUL's digital collections are preserved for the long-term, encouraging the use of CUL's digital collections in the creation of intellectual knowledge and research within the University, the international research community, and by wider society.

4.2. Mission statement

To support and facilitate education, learning, and research needs through sustained long-term access to digital content at the University. To position CUL in growing to become a leader of research libraries in the United Kingdom (UK), for the acquisition, management and preservation of digital content.

4.3. Goals

This Strategy sets ambitious goals for CUL. By 2024, CUL will have:

- Established a digital preservation project team for the ADPPF, to implement the Digital Preservation System (DPS);
- Transitioned to and/or established a BAU digital preservation team to support ongoing DPS activities;
- Transitioned to BAU for digital preservation, and be actively undertaking digital preservation activities;
- Developed and implemented end-to-end workflows and processes for managing all classes of digital content;
- Implemented fixity management for all digital content in both working space and preservation stores;
- Ingested the first phase of digital content from its collections (that is considered in scope for digital preservation) into the DPS;
- Developed a plan for the ingest of 'backlog' digital content, not included in the first phase of ingest;
- Improved digital preservation capability and maturity, so CUL can achieve baseline rankings across digital preservation maturity models, selected for CUL's purposes;
- Developed a strategy for audiovisual collection items, encompassing transfer from carriers and their preservation, and/or repatriation/transfer into other institutions' custody;
- Undertaken wide-ranging advocacy activities across the University, influencing and ensuring awareness of digital preservation responsibilities and needs throughout the University including governance, staff, students and researchers.

5. Drivers for change

Digital preservation has been identified as the second-highest risk for CUL at present. CUL is committed to increasing its digital preservation capabilities and establishing organisational and technical digital preservation infrastructure, in order to address a range of pressing drivers, both internal and external.

5.1. Internal drivers

Current internal drivers include:

- Commitment to improving the library experience for students, staff and researchers;
- Supporting research excellence and innovation including in digital and data-driven scholarship, including in the digital humanities;
- Supporting CUL's acquisition of collections that contribute to a new generation of knowledge, including the need to acquire born-digital collections and the University's own born-digital institutional records;
- Meeting CUL's and the University's custodial responsibilities to manage and make available CUL's world-renowned collections;
- Mitigating against corruption or loss for digital content yet to arrive in CUL's custody, by way of earlier intervention and acquisition;
- Risk of partial corruption or complete loss of digital content once in the custody of CUL (including content currently held on fragile carriers);
- Enhancing the security surrounding CUL's digital content;
- Highly likely that born-digital content could not be re-acquired or recreated;
- The cost of re-digitising original items and recreating associated metadata would be unacceptably high;
- Reputational risk to CUL and the University of corruption or loss of digital content occurs;
- Organisational knowledge loss regarding specialist digital preservation skills due to project-based resourcing;
- Failure to achieve key components of *Cambridge University Libraries Strategy 2019 – 2024*, should this Strategy not be implemented.

5.2. External drivers

Current external drivers include:

- Meeting current (and future) philanthropic and researcher funder requirements in relation to archiving and preserving digital content;
- Meeting data protection and General Data Protection Regulation (GDPR) compliance requirements;
- Achieving full accreditation through the UK's national Archive Service Accreditation;
- Ensuring CUL and the University will be able to attract research partners in digital scholarship and other disciplines;
- Meeting international standards and good practice in the management and preservation of digital content;
- Building foundations in order to pursue TDR certification (such as the CoreTrustSeal) in the longer-term;
- Meeting expectations of a world-renowned university and research library that holds world-class collections.

5.3. Approach

CUL needs to ensure support for all digital content in its collections. CUL's digital preservation capability is currently low and requires significant improvements to be made in both organisational and technical infrastructure, as well as resourcing. Considerable improvements across all areas must be made to provide the necessary processes, infrastructure and resourcing to meet current and future digital preservation demands. Having assessed CULs digital preservation needs against international good practice, the mechanism settled on for achieving incremental and sustained change (while ensuring a holistic approach to managing and preserving digital content), is to aim for TDR certification. Progressing towards the attainment of TDR certification requires major organisational and infrastructure changes and developments. It also requires smaller modifications of the day-to-day activities of existing staff.

5.4. How this will be achieved

Digital preservation will be a rapidly developing area for CUL over the next six years. To ensure CUL is 'travelling in the right direction', annual reviews of this Strategy (and the associated Digital Preservation Policy) will occur. Annual reviews will include digital preservation maturity modelling, using selected maturity models – the Digital Preservation Capability Maturity Model (DPCMM) and the NDSA Levels of Digital Preservation – chosen to best meet CUL's needs. This is coupled with

assessing against the CoreTrustSeal, which has been identified as the TDR certification scheme that CUL should be aiming for in the longer-term. Taking a maturity modelling and TDR assessment approach on an annual basis is necessary.

Practical steps forward, include undertaking the ADPFP, which is one necessary component of increasing CUL's digital preservation capability. The main focus of this project is on major digital preservation infrastructure implementation, along with some organisational change. It is intended that as an outcome of the ADPFP, CUL should be able to achieve Level 2 of the NDSA Levels of Digital Preservation.

In order to further increase CUL's digital preservation capabilities, other more specific processes, documentation, skills development and the resourcing and digital preservation activities must occur, focussing on known gaps (organisational, resourcing, skills and technical infrastructure). To ensure progress is achieved, a high-level annual implementation plan has been included in this Strategy, which is expected to be reviewed annually. The process of reviewing the implementation plan each year will also produce an annual digital preservation work plan, addressing known gaps. The goals outlined in the high-level implementation plan are necessary for achieving TDR certification in the future. This is due to TDR certification being dependent on meeting certain maturity levels across a holistic range of requirements, including resourcing as well as organisational and technical infrastructure.

Digital preservation awareness education across CUL governance, senior leadership and management, as well as ADPFP project managers is also seen as essential over the early years of this Strategy. Digital literacy and digital preservation literacy amongst CUL staff must also be developed over the duration of this Strategy. It is acknowledged that due to the skill level of existing CUL staff, developing requisite digital and digital preservation literacy of existing staff will be a longer process than can be accomplished over the timeframe of this Strategy.

6. Scope

All digital content in-scope of CUL's Digital Preservation Policy is covered by this Strategy.

7. Context

7.1. Content

All six classes of digital content in CUL's collections are considered in scope of this Strategy, regardless of format or carrier. Outlined fully in the Digital Preservation Policy, these classes include:

1. Born-digital personal and corporate records;
2. Born-digital university records;
3. Research outputs;
4. Published born-digital content;
5. Digitised content;
6. In-house created content.

Preservation masters, co-masters, born-digital originals and associated materials are considered in scope. Both digital and analogue audiovisual content is considered in scope of this Strategy, until such times that a dedicated audiovisual digitisation and preservation strategy is developed.

7.2. Challenges

Digital preservation is much more than a technical challenge, and must be addressed holistically. Using an approach based on the *Three-Legged Stool Model for Digital Preservation*,¹ illustrates where CUL faces challenges, which at present are on all fronts.

¹ Kenney, A. & McGovern, N. (2013) *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Available from: <http://dpworkshop.org/dpm-eng/conclusion.html> [Accessed 1st November 2018]

7.2.1. Organisational infrastructure

CUL’s digital preservation activities currently operate on a ‘project by project’ basis, which means that organisational and specialist knowledge of digital collections plus digital preservation activities and processes are lost, when each project ends. Currently there are no workflows or practices exist that are considered suitable for preservation purposes. Sustainable resourcing and operational processes must be established over the coming six years.

A lack of high-level and specialised policies, standards, procedures and guidelines (PSPG) also hampers CUL current ability to progress digital preservation efforts.

7.2.2. Resources framework

Over the past two decades, funding for digital preservation has been achieved through grants or philanthropic means. Continuing in this manner is a considerable risk to CUL, as ongoing core funding, as well as contingency funding are required for both digital preservation activities and TDR certification. Over the coming six years, CUL will develop a long-term strategic plan to fund BAU digital preservation activities, exploring funding models that may include chest funding, cost-recovery and/or further philanthropic sources.

7.2.3. Technical infrastructure

While digital preservation can be mistakenly viewed as a wholly technical challenge, CUL chooses to take a holistic view to ensure that technical solutions are only one part of a wider digital preservation remit.

A certain amount of technical infrastructure is required, in order for stewardship of digital content to be possible. CUL does not have a DPS, specialised tools, equipment, preservation storage or end-to-end workflows, necessary for digital preservation activities to occur. Over the coming six years, CUL will increase its digital preservation technical capability by first implementing several key components to support end-to-end workflows for digital content, including a DPS and preservation storage.

7.3. Stakeholders

CUL considers its high-level designated community to be the staff, students and researchers of the University, as well as other users that include the international research community and wider society.

7.4. Strategic fit

This Strategy aligns with CUL’s goals and vision, and supports the University’s mission and core values.

8. Implementation

As CUL does not have any BAU digital preservation capability at present, these activities will commence as project-based tasks, with the intention to transition to BAU over the course of the duration of this Strategy. The following goals are planned as part of the six-year Strategy:

| Year | Focus | Goals |
|------|--|--|
| 2019 | Commence the digital preservation component of ADPPF | <ul style="list-style-type: none"> • Source funding for a DPS • Procure and commence implementation of a DPS • Procure preservation storage • Implement data security for preservation storage • Annual review of this Strategy (including undertaking digital preservation maturity modelling) • Develop annual digital preservation work plan for 2019 and 2020 • Annual review of the Digital Preservation Policy • Define and document designated communities for all classes of digital content • Develop digital access and metadata policies, and metadata standards • Develop data preservation storage, backup and recovery policies and procedures |

| | | |
|------|----------------------------------|---|
| | | <ul style="list-style-type: none"> • Establish Communities of Practice (as defined by DPOC project) • Deliver digital preservation awareness training to governance, senior leadership and management and ADPPF project managers • Establish digital literacy training for CUL staff |
| 2020 | Support for digitised content | <ul style="list-style-type: none"> • Continue implementation of the DPS to support digitised content (including configuration and establishing workflows for digitised content inside the system) • Provide requirements for digitised content and feedback to the vendor for future enhancements and versions of the DPS tool • Implement and configure an end-to-end workflow management tool for digitised content • Document end-to-end workflow procedure(s) and diagram(s) for digitised content • Develop and implement end-to-end workflows for digitised content (for processes preceding the DPS) • Integrate DPS with CUDL delivery system, AMS, Alma and iDiscover for digitised content • Develop content models for digitised content • Develop migration plan for digitised content from DAMS and other identified locations • Migrate digitised content from Digital Asset Management System (DAMS) and preservation network stores into the DPS (as Bagit Submission Information Packages (SIPs)) • Acquire equipment and tools for a born-digital archiving and digital preservation lab • Develop fixity check policy and procedures • Develop migration policy • Develop digitisation policy, strategy, plan and procedures • Transfer digitisation protocols into standardised guidelines and standards documents • Develop and document quality control procedures for digitised content • Annual review of this Strategy (including undertaking digital preservation maturity modelling) • Develop annual digital preservation work plan for 2021 • Annual review of the Digital Preservation Policy • Train Digital Content Unit (DCU) and Digital Library Unit (DLU) staff on digitised content workflows and the DPS system • Establish plan for backlog of digitised content (not held in DAMS or preservation network stores) to be ingested into the DPS (as Bagit SIPs) |
| 2021 | Support for born-digital content | <ul style="list-style-type: none"> • Continue implementation of the DPS to support born-digital content (including configuration and establishing workflows for born-digital content inside the system) • Provide born-digital requirements and feedback to the vendor for future enhancements and versions of the DPS tool • Implement and configure an end-to-end workflow management tool for born-digital content • Document end-to-end workflow procedure(s) and diagram(s) for born-digital content • Develop and implement end-to-end workflows for born-digital content (for processes preceding the DPS) |

| | | |
|------|---------------------------|--|
| | | <ul style="list-style-type: none"> • Develop requirements for delivery systems for non-image born-digital content • Integrate DPS with delivery system(s), AMS, Alma and iDiscover for born-digital content • Undertake born-digital acquisition(s) and process born-digital content through the lab environment • Develop content models for born-digital content • Develop plan for processing born-digital collections to prepare them for ingest • Ingest born-digital content into the DPS (as Bagit SIPs) • Set up equipment and tools to develop a functioning born-digital archiving and digital preservation lab • Develop digital acquisitions policy and strategy • Develop digital acquisitions guidelines for donors and staff • Develop pre-ingest processes and documentation including digital processing and preconditioning technical policy and procedures • Develop and document quality control procedures for born-digital content • Annual review of this Strategy (including undertaking digital preservation maturity modelling) • Develop annual digital preservation work plan for 2022 • Annual review of the Digital Preservation Policy • Train CUL Manuscripts and University Archives staff on born-digital content workflows and the DPS system • Review JISC RDSS project with possible view to migrate research data to CUL DPS • Continue in the background, backlog ingest of digitised content (not held in DAMS or preservation network stores) into the DPS (as Bagit SIPs) |
| 2022 | Support for research data | <ul style="list-style-type: none"> • Based on review of JISC RDSS project, implement the DPS to support research data (including configuration and establishing workflows for research data inside the system) • Based on review of JISC RDSS project, provide research data requirements and feedback to the vendor for future enhancements and versions of the DPS tool • Implement and configure an end-to-end workflow management tool for research data • Document end-to-end workflow procedure(s) and diagram(s) for research data • Develop and implement end-to-end workflows for research data (for processes preceding the DPS) • Develop requirements for delivery systems for research data • Integrate DPS with delivery system(s), Symplectic Elements, Dimensions, iDiscover and other CUL systems for research data • Undertake large-scale research data acquisition(s) and process research data through the lab environment • Develop content models for research data • Based on review of JISC RDSS project, develop migration plan for research data from Apollo and other identified locations • Based on review of JISC RDSS project, migrate research data from Apollo into the DPS (as Bagit SIPs) • Develop and document quality control procedures for research data |

| | | |
|------|---------------------------------|--|
| | | <ul style="list-style-type: none"> • Annual review of this Strategy (including undertaking digital preservation maturity modelling) • Develop annual digital preservation work plan for 2023 • Annual review of the Digital Preservation Policy • Train Office of Scholarly Communications (OSC) and Research Data Management (RDM) staff on research data workflows and the DPS system • Develop a strategy for transfer of data from digital carriers and transfer devices • Develop audiovisual digitisation strategy and audiovisual preservation strategy and guidelines • Source funding for audiovisual digitisation and preservation • Continue in the background, backlog ingest of digitised content (not held in DAMS or preservation network stores) into the DPS (as Bagit SIPs) |
| 2023 | Support for content on carriers | <ul style="list-style-type: none"> • Based on resourcing and staff capabilities, undertake a programme of work to transfer data from digital carriers and transfer devices • Based on resourcing and staff capabilities, implement the DPS to support disk imaged content (including configuration and establishing workflows for disk image content inside the system) • Based on resourcing and staff capabilities, provide disk image content requirements and feedback to the vendor for future enhancements and versions of the DPS tool • Based on availability of funding for audiovisual preservation, implement the DPS to support audiovisual content (including configuration and establishing workflows for audiovisual content inside the system) • Based on availability of funding for audiovisual preservation, provide audiovisual requirements and feedback to the vendor for future enhancements and versions of the DPS tool • Implement and configure an end-to-end workflow management tool for audiovisual content • Document end-to-end workflow procedure(s) and diagram(s) for audiovisual content • Develop and implement end-to-end workflows for audiovisual content (for processes preceding the DPS) • Develop requirements for delivery systems for audiovisual content • Integrate DPS with delivery system(s), AMS, Alma, Symplectic Elements, Dimensions, iDiscover and other CUL systems for audiovisual content • Undertake transfer of digitised and born-digital audiovisual content acquisitions and process audiovisual content through the lab environment • Develop content models for audiovisual content and disk imaged content • Based on availability of funding for audiovisual preservation, develop plan for processing digitised audiovisual collections to prepare them for ingest • Based on availability of funding for audiovisual preservation, ingest audiovisual content into the DPS (as Bagit SIPs) • Based on availability of funding for audiovisual preservation, commence migration of audiovisual content into the DPS (as Bagit SIPs) • Develop physical carriers disposal policy |

| | | |
|------|--|--|
| | | <ul style="list-style-type: none"> • Update audiovisual digitisation strategy and audiovisual preservation strategy and guidelines • Develop and document quality control procedures for audiovisual content • Annual review of this Strategy (including undertaking digital preservation maturity modelling) • Develop annual digital preservation work plan for 2024 • Annual review of the Digital Preservation Policy • Train DCU, DLU, OSC, RDM and Manuscripts and University Archives staff on audiovisual workflows and the DPS system • Continue in the background, backlog ingest of digitised content (not held in DAMS or preservation network stores) into the DPS (as Bagit SIPs) |
| 2024 | Transfer at risk collections from carriers and prepare for TDR certification | <ul style="list-style-type: none"> • Based on resourcing and staff capabilities, continue programme of work to transfer data from digital carriers and transfer devices • Based on availability of funding for audiovisual preservation, continue programme of work for the digitisation and preservation of audiovisual content • Continue in the background, backlog ingest of digitised content (not held in DAMS or preservation network stores) into the DPS (as Bagit SIPs) • Undertake trials of preservation actions • Develop preservation actions technical policy and guidelines • Develop other PSPG, strategy and associated documentation required for TDR certification • Develop the <i>Digital Preservation Strategy 2025 – 2030</i> |

9. Strategic priorities

CUL's strategic priorities over the coming six years are categorised as five themes, representing CUL's drivers for change. They are:

1. Working together to provide an unrivalled education and student experience;
2. The library as a partner in research;
3. Sharing our collections and knowledge openly with the world;
4. An integrated digital and physical environment;
5. Strong foundations for the future.

CUL's key priorities for each theme are found in the *Cambridge University Libraries Strategy 2019–2024*. Specific digital preservation priorities are outlined in this Strategy, aligning with CUL's themes and key strategic priorities.

9.1. Priority 1: Working together to provide an unrivalled education and student experience

9.1.1. Defining designated communities

CUL will define its designated communities for each class of digital content in 2019. Reviewing designated community definitions will take place in subsequent years, aligning with the foci for each year of the digital preservation implementation plan.

9.1.2. Knowledge sharing

Digital preservation requires wide-ranging knowledge and expertise. Information sharing is seen as one of the most efficient ways of addressing gaps in CUL's current born-digital archiving, audiovisual and digital preservation specialist knowledge. With current limited resourcing for digital preservation activities, it is crucial that knowledge sharing is established. This may take place within CUL, the University, the UK's Legal Deposit Libraries (LDLs), partnering organisations, the Digital Preservation Coalition (DPC) and its membership, and with other organisations that CUL is a member of, or maintains an association.

Knowledge sharing may take place via CUL's involvement membership of or involvement with the following organisations, including: the Archives and Records Association UK and Ireland (ARA), the British Records Association (BRA), the Chartered Institute of Library and Information Professionals (CILIP), DPC, Jisc, the International Association of Sound and Audiovisual Archives (IASA), the Institute of Conservation (ICON), the International Council on Archives (ICA), the Information and Records Management Society (IRMS) and Research Libraries UK (RLUK) as well as relevant born-digital archiving and digital preservation product user groups etc.

9.2. Priority 2: The library as a partner in research

9.2.1. Collaboration

Where strategic alignments can be made, and efficient information and knowledge sharing can take place, CUL will seek to form collaborations with other organisations.

CUL will continue the brand and the commitment to collaborative digital preservation work that has been established through the Digital Preservation at Oxford and Cambridge (DPOC) project. CUL continues to benefit from the collaboration and partnership it has established with Bodleian Libraries, Oxford, in the digital preservation space.

9.2.2. Digital content creators and donors

CUL will develop processes for engaging with digital content producers and donors, earlier in the digital content lifecycle. Where relevant, content producers will be encouraged to submit Data Management Plans (DMPs). This approach will mitigate against the risk of digital content being lost or corrupted before it has the opportunity to be transferred into CUL's custody.

9.2.3. Digital content

9.2.3.1. Digital content on carriers

CUL will develop a strategy to transfer digital content from fragile carriers to managed storage and preservation storage, supporting Non-Print Legal Deposit (NPLD) digital content and Special Collections needs. Collaborative approaches will be taken, to ensure minimised duplication of effort, including working with LDLs, affiliate and dependent libraries. Transferring at risk digital content from carriers and preserving it, is the first step towards making this digital content available.

Original carriers will not be provided to staff, students and researchers for use. Digital content must be transferred for preservation purposes in the first instance, with an access copy provided for use.

9.2.3.2. Audiovisual content

CUL's audiovisual content exists on both digital and analogue carriers. This content is considered high-risk, due to the fragility of the carriers. International experts advise that the content is highly likely to become irretrievable from these carriers over the next decade, with a 'deadline' for transfer of content set between 2023 and 2029.² CUL will develop a strategy that addresses the specific needs of audiovisual preservation including transfer from these fragile carriers.

Until such times that an audiovisual digitisation and preservation strategy is developed, restricted use of audiovisual content will be adhered to. In order to ensure future preservation of audiovisual content is possible, requests for use of audiovisual content will not be supplied to staff, students and researchers without intervention, if the only copy that exists is the original copy on a carrier. Only

² Casey, M. (2015). Why Media Preservation Can't Wait. *International Association of Sound & Audiovisual Archives Journal*. 44. 14-22. Available from: <https://www.weareavp.com/mike-casey-why-media-preservation-cant-wait-the-gathering-storm/> [Accessed 1st November 2018]

Lacinak, C. & Lyons, B. (2014) The Cost of Inaction: A New Model for Physical Audiovisual Media Holdings. *MAC Newsletter*. 42 (2), 20-23. Available from: <https://lib.dr.iastate.edu/macnewsletter/vol42/iss2/10> [Accessed 1st November 2018]

National Film and Sound Archive of Australia (2016), *Deadline 2025*. Available from: <https://www.nfsa.gov.au/corporate-information/publications/deadline-2025> [Accessed 1st November 2018]

Wright, R. (2013), Going, Going, Gone: prospects for analogue audiovisual content. *PrestoCentre*. Available from: <https://www.prestocentre.org/blog/going-going-gone-prospects-analogue-audiovisual-content> [Accessed 1st November 2018]

access copies will be made available for use (after ensuring an original copy or preservation master exists). Where audiovisual content has been requested for use (and only an original copy on carrier exists, without an access copy being available), arrangements must be made for this content to be transferred or digitised. The resultant access copy will be provided to staff, students and researchers.

9.2.4. Licencing and continuity of access

In order to manage and ensure preservation responsibilities for digital content are known, all records of agreements, contracts, licences and Memorandums of Understanding (MOU) will be centralised in an Electronic Document and Records Management System (EDRMS) over the coming six years. Moving forward, agreements, contracts and licences will be standardised, such as using boilerplate agreements or templates, to ensure ease of compliance and improve workplace efficiencies.

9.3. Priority 3: Sharing our collections and knowledge openly with the world

9.3.1. Access and reuse

In order to support long-term access and reuse of digital content, preservation is critical. CUL aims to better integrate systems, shifting towards centralised access to digital content while supporting a variety of stakeholder needs. The intention is to improve overall efficiencies while complying with funder requirements and relevant legislation.

CUL will ensure that workflows developed do not bypass preservation steps, so that long-term access is made possible. CUL ensures that all activities undertaken to improve digital preservation capabilities increase or maintain existing levels of service, in relation to access. Dedicated policies and guidelines will be developed to supporting access and use, reuse and takedowns.

Facilitating certain types of digital scholarship also requires mechanisms are in place for access to preservation master content, which in future will be stored in a dark archive. This will be managed via appropriate access controls and processes.

9.3.2. Policy and strategy

As digital preservation is an actively developing area for CUL, CUL commits to reviewing this Strategy and the Digital Preservation Policy annually. Annual work plans and roadmaps will be developed to align with this Strategy. These will be created based on annual digital preservation maturity modelling and in response to the developing digital preservation environment, which will be implemented through the ADPPF, the DTP, the University's RDM group, work undertaken as part of LDLs and other collaborations, such as the Jisc RDSS project. CUL will also develop new policies and guidelines (such as to support end-to-end digital stewardship, digital content lifecycles and the digital preservation environment), in order to fill known policy gaps.

The cross-departmental Born-Digital Archiving and Digital Preservation Implementation and Governance Group is responsible for overseeing annual work plan reviews and year-round monitoring achievements of identified goals. This group is also responsible for overseeing the development of new policies, strategies, plans and roadmaps to support born-digital archiving, audiovisual and digital preservation; ensuring activities, goals and resourcing needs are identified reported to relevant governance, management and leadership as required.

9.3.3. Ethics

Digital preservation encompasses many traditional disciplines. CUL will ensure that it executes good practice by following charters, codes of ethics, codes conduct, diversity policies, guidelines, and principles as outlined by the professional bodies that are international leaders in digital preservation and related disciplines. As a member of selected organisations – including the DPC, IASA and the ICA – CUL ensures that it follows the codes, principles and good practice outlined by these organisations.

9.4. Priority 4: An integrated digital and physical environment

The digital environment is a key priority in *Cambridge University Libraries Strategy 2019 – 2024*, which this Strategy aligns with. Progressing goals aligned with the digital agenda will facilitate establishing infrastructure, workflows and processes fundamental to the supporting the acquisition, digitisation and preservation of digital content. CUL recognises that active digital preservation is

critical to the long-term access to digital content and associated metadata, including current and yet to be thought of future uses.

9.4.1. Digital content management

9.4.1.1. Integrity and authenticity

Fixity will be established and implemented for all classes of digital content, including digital content stored in managed network locations, and in the DPS (once this is implemented). As a minimum, fixity will include checksum generation and/or verification for each file, with each checksum hash accompanied by the associated filename and file path. Regular reports are produced to support fixity activity, as well as event-driven reports, such as checksum verification failures. In order to support newly implemented fixity checking of digital content, a process for addressing fixity issues will also be established.

CUL will establish the foundations for creating and recording provenance metadata, and maintaining links between all components of digital content in its collections, including metadata and associated materials as part of the ADPPF. Versioning of digital content and metadata will also be implemented as part of the ADPPF, either using the capability of the DPS tool or through content models.

9.4.1.2. Data quality and standards

Standards enable efficient and streamlined management and quality control of digital content. Adherence to standards sets the foundations for scaling-up digital preservation activities and future automation.

CUL will implement existing international or national standards (or guidelines), relevant to digital preservation activities, processes and infrastructure. Where standards do not exist, CUL will develop its own internal standards based on existing international or national good practice.

Implementation and development of standards will be effectively governed. A newly-formed CUL Standards Group will oversee the implementation of existing standards, and development of internal standards, including those for use in born-digital archiving and digital preservation. Documentation and testing of internally-developed standards will occur and CUL will share developments with the international digital preservation community.

9.4.1.3. Digital acquisitions and appraisal

Digital acquisition is a new area of work for CUL. Processes will be established to support the acquisition of all classes of digital content. CUL will also establish associated workflows and infrastructure to for acquisition and appraisal of digital content.

Digital content currently arrives at CUL via all departments and teams. Going forward, collaboration on approaches to acquisition and appraisal of digital content must occur, including through formalised governance groups, such as the Born-Digital Archiving and Digital Preservation Implementation and Governance Group. Working collaboratively with content producers, donors, other departments of the University, Colleges and UIS will occur, as necessary.

Standardised forms and guidelines will be developed, so as to ensure documentation of activities is undertaken, as part of digital acquisition and appraisal processes. These will be retained as records and eventually placed in an EDRMS. Where possible, practices, forms and guidelines will be borrowed from the international born-digital archiving and digital preservation communities, and adapted to CUL's needs.

Bare-minimum fixity (such as checksum generation and/or verification) is an essential activity that will be carried out as part of all acquisition and appraisal processes.

9.4.1.4. Metadata

Over the coming six years, CUL will shift towards implementing a more holistic approach to creating, enhancing and preserving metadata associated with its digital collections. Workflow processes for managing digital content and metadata will change from current practices. As a minimum, all digital content (or collection items undergoing digitisation) will have a 'shell' descriptive metadata record created. CUL will work towards a single 'source of truth' for all metadata types.

As part of the ADPPF and wider DTP portfolio of work, CUL will generate and record preservation metadata (including provenance) as well as recording existing administrative, technical and structural and rights metadata for all digital content.

9.4.1.5. Ingest

As part of the ADPPF, all existing digital content that is considered in scope for preservation will be ingested into the DPS. Until such times as the DPS is available, all digitised image content will be ingested into the existing Digital Asset Management System (DAMS) and the University's Apollo service for research data. For digital content that does not have a corresponding management system (such as born-digital content and digitised audiovisual content), its existence will be recorded in a Digital Asset Register.

Once workflows for each class of content have been established, all new digital content acquired or digitised by CUL will be ingested into the DPS as a matter of priority.

9.4.1.6. File formats

CUL will develop guidelines for donors, encouraging the use of preservation formats for content they are depositing. These preservation formats will either be standards-based, or follow international good practice.

Given CUL's low-level of digital preservation maturity and digital literacy, normalisation of file formats is considered out of scope and will not be attempted. This may be revisited after the DPS and digital preservation environment are fully established, if CUL has the requisite digital preservation expertise. Any changes that need to be carried out on digital content will be undertaken on a copy of the preservation master, co-master or born-digital original, with all files versioned.

9.4.1.7. Processes

Digital preservation is about undertaking ongoing processes to ensure digital content is actively managed for the long-term. As CUL does not have any ongoing digital preservation processes established, the intention of this Strategy is that over the course of the six-years a range of processes will be established, tested, implemented, documented and maintained so as to support born-digital archiving and digital preservation functions. In addition to fixity checking and virus checking, other processes will include file format identification, characterisation and/or validation.

Given the amount of CUL's digital content currently at risk, the priority is to ingest all digital content into the DPS (once this has been implemented and is available). Where possible, processes will be run within the DPS, with a minimum of pre-ingest processes taking place only to ensure unwanted digital content is not retained and ingests are successful.

Once in place, all changes to processes will be adequately tested and documented prior to enhancements or changes occurring.

9.4.1.8. Digital collections assessment and prioritisation

Digital content is held in a range of locations, carriers and storage media throughout CUL. Digital content not held in CUL's main 'preservation master store' is considered high-risk. Over the coming six years, this digital content held in 'non-standard' locations will have plans developed to support migration and/or transfer, with the ultimate aim of this digital content being migrated into the 'preservation master store' and ingested into the DPS. This digital content will be prioritised in line with the ADPPF and digital preservation implementation plan, which is revised annually. Prioritisation for migration of content into the preservation master store (and ingest into the DPS) may be influenced by preservation needs assessments on certain classes of material. Inaccessibility, obsolescence, rarity, significance, urgency, value to staff, students, researchers, users and/or stakeholders and demand for use of CUL's digital content are driving factors that may also influence prioritisation.

9.4.2. Technical infrastructure

9.4.2.1. Standards-based technologies

As part of the ADPPF, CUL will implement a well-managed and auditable DPS, that is based on international standards (where applicable). In alignment with the Digital Preservation Policy, where fit-for-purpose, CUL will implement open source systems and tools. As part of implementation, documentation and testing of how technologies meet international, national or CUL standards will take place. Safeguards and escrow agreements will also be established to ensure digital content is not 'stranded' in any system, and that exit strategies are available (and tested) for all digital content and metadata.

9.4.2.2. Workflows and processes

Processes and workflows for handling digital content will be established, to ensure consistent handling. End-to-end workflows for digital content that incorporate digital preservation stages do not currently exist at CUL. Standardised workflows for all digital content classes (that allow for scalability and future automation) will be developed and documented as part of the ADPPF. All workflows will include quality assurance for digital content, that will be undertaken by relevant subject matter experts. It is intended that all workflows will be managed by a workflow management tool.

Prior to any automation taking place, analysis of digital content and context by subject matter experts will occur, to mitigate against unintended preservation actions and results.

9.4.2.3. Technical systems and infrastructure

As part of the ADPPF, CUL will procure and implement a DPS. As part of the wider DTP portfolio of work, to support end-to-end workflow management for all classes of digital content including ingest into the DPS, CUL will procure (and/or develop) and implement a workflow management tool.

All configurations and workflows setup within systems and tools will be documented and retained.

9.4.2.4. Security and storage

Secure storage will be implemented for all preservation storage copies, as part of the ADPPF and associated DTP work. Mechanisms will be established so that only authorised users are able to access relevant digital content, with the aim to integrate all the systems via a centralised security function. CUL will implement fixity checking, regardless of location, including for secure working space storage, where digital content is processed. Virus checking and rechecking will also be established.

As part of the ADPPF and DTP, storage infrastructure to support the required number and geographically separated preservation masters will be established. Mechanisms for monitoring and reporting on the preservation master storage (comprising a continuous testing process) will also be established.

Secure working space storage for processing digital content will be implemented, with access only provided to authorised users. This will be separated from CUL's main network infrastructure, via an 'airgap' or other mechanism.

Carriers containing sensitive digital content will be secured in environmentally controlled conditions, the digital content can be transferred.

9.4.2.5. Born-digital archiving and digital preservation lab

As part of the wider DTP portfolio of work, CUL will establish a physical workspace to enable born-digital archiving activities, including the transfer of digital content off carriers, and will provide equipment for preservation actions requiring access to digital forensics and/or obsolete software, hardware and/or peripherals.

This lab will be developed to contain the minimum essential hardware, software, peripherals, storage (suitable for all types of carriers) and secure working space storage, necessary for processing digital content and creating 'information packages' for ingest into the DPS. The lab will also have or be fitted with adequate network infrastructure (and bandwidth) as well as suitable environmentally controlled conditions.

9.4.2.6. Equipment and infrastructure procurement and refresh schedules

CUL operates on a five-year refresh and/or replacement schedule. All equipment, working space and preservation storage, as well as other relevant infrastructure that comprises CUL's digital preservation environment is subject to these refresh and replacement schedules. It is intended that at least once during this Strategy, a storage replacement or refresh will be tested and carried out. Any replacement and disposal of equipment and technology is in line with DIS and UIS policies, guidelines and procedures.

9.4.3. Documentation and reporting

9.4.3.1. Operational and technical documentation

CUL's digital preservation environment must be sustainable. Ownership of tools and systems must take place collaboratively, with responsibility shared so there is no 'single point of (human) risk' in the operation of any tool, system, workflow or process.

Any new tool or system that is implemented will have associated guidelines and procedures created, detailing configurations, workflows and integrations with other systems. Likewise, in order to support the acquisition and management of digital content, digital organisational records must be produced and retained. All documentation produced will be managed in an EDRMS. To support this need, CUL will investigate procuring and implementing an EDRMS.

9.4.3.2. Technical reporting

CUL will establish regular reporting as defined in the Digital Preservation Policy, to be implemented as part of the ADPPF. This will commence with fixity reporting, specifically checksum verification failures.

Reports on the state of preservation storage will also be implemented and produced annually, or as required.

9.4.4. Planning

9.4.4.1. Towards certification

The next six years are focussed on establishing essential foundations for digital preservation. For CUL to be able to attempt TDR certification in the future (such as the CoreTrustSeal certification), it is crucial for CUL to address digital preservation needs holistically. Digital preservation certification relies on meeting organisational infrastructure, resources framework as well as technical infrastructure components, and cannot be achieved without specific requirements across all three areas having been met. This Strategy is designed to address all required areas of CoreTrustSeal certification.

9.4.4.2. Preservation planning

As digital preservation is growth area for CUL, coupled with the ambitious goals outlined in this strategy, it is essential that annual reviews of this Strategy and annual preservation planning occurs. This will be underpinned by annual digital preservation maturity modelling, while also taking into account other developments achieved through the DTP portfolio of work and priorities identified in CUL's organisations strategy. The results of all annual reviews will feed into the digital preservation annual work plan.

9.5. Priority 5: Strong foundations for the future

9.5.1. Funding

Over the coming six years, CUL will seek out and establish programme and BAU funding to support digital preservation activities. This is a fundamental need of a sustainable digital preservation programme, and is a key requirement for TDR certification.

9.5.2. Organisational infrastructure

CUL will work to develop the necessary organisational infrastructure so that digital preservation is considered a concern of the whole of organisation, and the responsibility of all staff.

9.5.3. Governance

CUL commits to developing leadership and management across the organisation, in order to support the digital preservation environment being established, as part of the DTP. This is necessary for underpinning new activities, workflows, teams and ways of working to support our digital collections.

CUL establishes necessary governance including cross-departmental policy and strategy, standards, and born-digital archiving and digital preservation implementation and governance groups. Cross-departmental groups are essential, so as to iron out existing duplication, standardise processes and create workplace efficiencies. These groups will provide essential support for CUL's staff to develop and manage processes and documentation for digital content and the digital preservation environment. CUL will implement formalised governance and approval processes for the implementation of PSPG, strategy and systems.

As digital preservation is considered CUL's second-highest risk at present, the Born-Digital Archiving and Digital Preservation Implementation and Governance Group is responsible for overseeing the annual review of digital preservation activities (including maturity modelling), in order to drive annual work plans and address organisational capability and capacity building.

9.5.4. Risk mitigation and reporting

CUL will produce project and annual reports, outlining progress in the area of digital preservation risk mitigation and achievements. CUL will also produce annual reports illustrating the value of preserving digital content in its collections. These reports will assist in facilitating funding bids.

9.5.5. Expertise and resourcing

CUL will build, establish and develop a BAU digital preservation team that includes a mixture of broad-ranging specialist skillsets. This is necessary for ensuring that continuous expert guidance is available to support born-digital acquisitions, audiovisual and digital preservation activities, for all classes of digital content. CUL will establish a suitably qualified project team for implementing the DPS, as part of the ADPPF.

CUL engages and maintains connection with the DPC membership and other international digital preservation networks in order to obtain specialist subject matter expertise, particularly where skills gaps exist.

9.5.6. Digital preservation literacy

Over the coming six years, as digital preservation infrastructure is implemented, organisational change is essential. Digital literacy baselines will be increased across the organisation, as this is foundational for digital preservation awareness and activities to occur. Digital preservation awareness (for management and leadership), digital preservation literacy and/or practical 'hands-on' digital preservation training (for practitioner staff) will take place as soon as digital literacy levels have improved. Communities of practice that have been defined by the DPOC project will commence, so skills and knowledge development can be shared cross-departmentally, increasing standardisation and workplace efficiencies across the organisation. Digital and audiovisual preservation awareness training will also commence, drawing on training opportunities offered by organisations that CUL is a member of, as well as other strategic programmes run by national leading audiovisual organisations.

10. Roles and responsibilities

In order for the ambitious goals outlined in this Strategy to be achieved, they must be resourced. Resourcing for the digital preservation activities associated with the ADPPF is contingent on funding being provided by the University. Sourcing ADPPF funding is the responsibility of the Director of the Digital Transformation Programme, Deputy Director, Digital Initiatives and Deputy Director, Research Collections, with contributions from the University Librarian and Senior Leadership Team, as required.

Other digital preservation activities outlined in this Strategy are yet to be resourced. Resourcing and recruiting staff with the requisite digital preservation skillsets, necessary for meeting the goals outlined in this Strategy is the responsibility of the Director of the Digital Transformation Programme and Deputy Director, Digital Initiatives in collaboration with other CUL staff, as required.

All other roles and responsibilities in relation to digital preservation activities are outlined in the Digital Preservation Policy.

11. Definitions

Please refer to Appendix A.

12. Implementation framework

CUL is establishing and implementing digital preservation practices, policies, strategy, guidelines, procedures, standards and skills to support the acquisition, creation, management and preservation of its digital content. CUL is also commencing the DTP. Part of this portfolio of work involves the procurement and implementation of infrastructure, systems and workflows to support digital preservation activities.

13. Related documents

This Strategy should be read alongside the following CUL and University of Cambridge documents.

13.1. Related Cambridge University Libraries guidelines, policies and strategies

- *Cambridge University Libraries Digital Preservation Annual Review Guidelines*
- *Cambridge University Libraries Digital Preservation Policy*
- *Cambridge University Libraries Strategy 2019 – 2024*
- *Special Collections Digital Strategy (2014)*

13.2. Related University of Cambridge strategies

- *Business Systems and Data Integration Strategy (in draft)*
- *Digital Teaching and Learning Strategy (in draft)*
- *University of Cambridge Digital Strategy for Education 2016-2020*
- *Web Strategy (in draft)*

13.3. Related charters and codes

- *Archives and Records Association UK and Ireland (ARA) Code of Ethics*
- *Chartered Institute of Library and Information Professionals (CILIP) Ethical Principles and Code of Professional Practice for Library and Information Professionals*
- *Digital Preservation Coalition – Digital Preservation Community Charter (in draft)*
- *Digital Preservation Coalition – Inclusion and Diversity Policy*
- *International Association of Sound and Audiovisual Archives (IASA) Ethical Principles for Sound and Audiovisual Archives*
- *International Association of Sound and Audiovisual Archives (IASA) Safeguarding the Audio Heritage: Ethics, Principles and Preservation Strategy (IASA-TC 03)*
- *United Nations Educational, Scientific and Cultural Organization (UNESCO) – Charter on the Preservation of Digital Heritage*

14. Approval

This Strategy was approved by CUL's Senior Leadership Team on 31 January 2019 for immediate implementation.

15. Implementation

This Strategy is implemented on 18 April 2019.

16. History

This is the inaugural Digital Preservation Strategy.

17. Prepared by

Digital Preservation Specialist – Policy and Planning (Polonsky Fellow)
07 December 2018

18. Document history and version control

| Version | Date approved | Approved by | Brief description |
|---------|-----------------|------------------------|-------------------|
| 1.0 | 31 January 2019 | Senior Leadership Team | First release |

| | | | |
|--|--|--|--|
| | | | |
|--|--|--|--|

19. Appendix A

Definitions in use by the international digital preservation and digital curation communities are sourced from the Digital Preservation Coalition's [Digital Preservation Handbook Glossary](#), the Digital Curation Centre's [Glossary](#), the National Digital Stewardship Alliance (NDSA) [Glossary](#) and Wikipedia.

Access copy – The version of a file or data that is made available to users. Access copies are distinct from preservation masters or co-masters, however in some cases an access copy file is identical to a preservation master or co-master. Access copies are not intended for long-term preservation. Where preservation masters and/or co-masters have been lost, or can no longer be located, a duplicate of an access copy may be made, and deemed the preservation master.

Archival Management System (AMS) – The tool that is used to create and manage descriptive metadata records for archival collections.

Archival master – See preservation master.

Associated data – See associated materials.

Associated materials – Data or files that provide context, assist in the interpretation of, or are essential in the process of rendering and/or making available digital content. Associated materials may include a wide variety of files such as algorithms, code, sidecar files, scripts, transcripts, Optical Character Recognition (OCR) files etc. They may also be referred to as associated data or supporting documents. Also see sidecar file.

At risk – Digital content that due to obsolescence, fragility of carriers and/or formats, or for another reason, may become corrupted or lost without immediate or near-immediate action.

Authenticity – 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.

Born-digital – Digital materials that are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form. This term has been used to differentiate them from digital materials that have been created as a result of converting analogue originals, and digital materials that may have originated from a digital source but have been printed to paper, for example some electronic records.

Born-digital original – The first-known instance of a file that is transferred from a donor or researcher to a collecting institution or other custodian. Due to the nature of digital content, there may be multiple prior versions of a file, however this is the known and documented version of a file, without any alterations made to the content format or other properties of the file.

Carrier – A physical item on which content is recorded, encoded or fixed. This can be stored as analogue or digital information. Selected examples of physical carriers include magnetic tape (carrying analogue or digital audio and/or video content), motion picture film (carrying optical moving image and/or audio content), disks (zip disks, 3½ inch and 5¼ inch floppy disks, carrying digital data), optical media (such as compact discs, DVDs and Blu-ray discs, carrying digital data, audiovisual or multimedia content) portable hard disk drives or USB flash drives (carrying digital data). May also be referred to as a physical carrier, physical format carrier, physical format digital carrier, portable media or transfer device.

Checksum – A unique alphanumeric signature derived from a file. Used to compare copies. A checksum can be conceptualised as the 'fingerprint' of a digital file.

Class – CUL digital content is broadly grouped into a number of different ‘types’ of digital content. These types are somewhat associated with the CUL’s collecting streams. How digital content is managed in the long-term, may differ slightly, depending on the class to which it belongs.

CoreTrustSeal – A harmonisation of the Data Seal of Approval (DSA) and the ICSU World Data System (WDS). The CoreTrustSeal is a core-level certification. The requirements reflect the core characteristics of a Trusted Digital Repository.

Co-master – A co-master file is derived from a preservation master file. The co-master may have cropping, filtering and/or other similar actions applied. It is typically the source from which access copies are generated, and may itself be considered a high-quality access file.

Data migration – Moving data, such as metadata, from one location to another. Typically, this term refers to moving data (or metadata) held in one system to a new system (such as from one database to another).

Delivery copy – See access copy.

Derivative – See access copy.

Designated community – An identified group of potential consumers who should be able to understand a particular set of information.

Digital asset – See digital content.

Digital Asset Management System (DAMS) – This is a tool used for managing day-to-day use of digital content that is not restricted to high-quality digital assets (master copies).

Digital content – Any arbitrary item created, published or distributed in digital form, including, but not limited to text, data, sound recordings, photographs and images, motion pictures and software. May also be referred to as digital assets, digital materials, digital objects or digital resources. This encompasses both born-digital and digitised content.

Digital content lifecycle – The different stages that digital content passes through. The Digital Curation Centre developed the DCC Curation Lifecycle Model to illustrate the different stages. CUL considers these stages to be: Conceive, Prepare, Create, Evaluate and Negotiate, Appraise, Acquire, Arrange and Describe, Pre-Ingest, Ingest, Store and Manage, Preserve, Deliver and/or Provide Access, Discover, Use and/or Reuse.

Digital imaging – See digitisation.

Digital materials – See digital content.

Digital object – See digital content.

Digital Object Identifier (DOI) – A technical and organisational infrastructure for the registration and use of persistent identifiers widely used in digital publications and for research data. The DOI system was created by the International DOI Foundation and was adopted as International Standard ISO 26324 in 2012.

Digital preservation environment – The whole end-to-end set of systems, tools, workflows and processes to support comprehensive stewardship of digital content. Rather than a single tool or system, it is the entire digital ‘ecosystem’ for managing digital content.

Digital Preservation System (DPS) – The tool used for the preservation and management of digital content, coupled with the underlying storage. The Digital Preservation System also incorporates other ‘bolt on’ tools or adaptors for ingesting content, undertaking preservation actions such as checksum generation and verification, format identification and characterisation etc. and other preservation activities. This is typically a ‘dark archive’.

Digital preservation tool – This is the digital preservation software tool used for managing digital assets. The digital preservation industry may sometimes describe this as a ‘digital preservation repository’. It should be differentiated from the Apollo Research Data Repository which does not have the capacity to undertake preservation functions.

Digital resource – See digital content.

Digital surrogate – Manifestation of a physical collection item in digital form, for example, a digital image file of a manuscript item. These can be made more widely available than the original physical collection item they represent and allow for new modes of discovery and use that may not be possible with the original physical collection item. A digital surrogate does not replace the original physical collection item. This is typically referred to as digitised content.

Digitisation – The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, can then be classed as digital content and is subject to the same broad challenges as born-digital content, involved in preserving and maintaining access.

Digitised content – The term used to refer to digital content that has been created by way of a digitisation process of an original physical collection item. This may also be referred to as a digital surrogate.

Distribution copy – See access copy.

File format – A file format is a standard way that information is encoded for storage in a computer file. It tells the computer how to display, print, process, and save the information. It is dictated by the application program which created the file, and the Operating System under which it was created and stored. Some file formats are designed for very particular types of data, others can act as a container for different types. A particular file format is often indicated by a filename extension containing three or four letters that identify the format.

Filename – A name used to uniquely identify a computer file stored in a file system.

File path – A specified unique location that is formed by the list of directories (or folders) and sub-directories, that forms a path to where a file is located in a file system.

File system – Controls how data is stored and retrieved.

Fixity – Checksum(s), filename and file path that are generated and/or recorded for a specific file. The combination of these properties can be used to identify and verify whether any changes have taken place to a file between two points in time. For information on this process see fixity check.

Fixity check – A method for ensuring the integrity of a file and verifying it has not been altered or corrupted. During transfer, an archive may run a fixity check to ensure a transmitted file has not been altered en route. Within the archive, fixity checking is used to ensure that digital files have not been altered or corrupted. It is most often accomplished by computing checksums such as MD5, SHA1 or SHA256 for a file and comparing them to a stored value.

Format migration – A means of overcoming technological obsolescence by transferring digital content from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of ‘digital objects’ and to retain the ability for users to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the digital content with the new generation of technology. May also be referred to as migration.

Integrity – Data that has remained unchanged. For example, data that has undergone a process (such as transmission or storage and retrieval) and is identical after the process to how it was before the process began.

Information package – A set of files that is made up of the digital content and the associated metadata. As defined by the Open Archival Information System (OAIS), there are three types of information packages: the Submission Information Package (SIP), the Archival Information Package (AIP) and the Dissemination Information Package (DIP).

Legal Deposit Libraries (LDLs) – Cambridge University Libraries is one of six Legal Deposit Libraries in the UK and Ireland. The other Legal Deposit Libraries are the British Library, the Bodleian Libraries, Oxford, the National Library of Scotland, the National Library of Wales and Trinity College, Dublin.

Metadata – Information that describes significant aspects of a ‘digital object’. This encompasses many types of information including descriptive, administrative, technical, structural, rights and preservation metadata. Most discussion to date has tended to emphasise metadata for the purposes of resource discovery. For digital preservation purposes, the emphasis is on what metadata are required to successfully manage and preserve digital materials over time and assists in ensuring essential contextual, historical, and technical information are preserved along with the digital content. The *PREMIS Data Dictionary for Preservation Metadata* has become a key de facto standard in digital preservation.

Migration – The process of transferring digital content from one system to another. This may involve export from one system and ingest into the new system. For migration from one file format to another, see Format Migration. For metadata migration, see data migration.

Normalisation – The process of transferring digital content into another file format, usually a predetermined file format that is considered a suitable preservation file format.

Open Archival Information System (OAIS) – An Open Archival Information System (or OAIS) is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a designated community.

Open source – A philosophy and methodology (often used by software developers) where the source code is made freely available, so that others may continue to develop the software. Open source is the philosophical opposite of proprietary.

Original – A collection item that has physical properties, for example a book, manuscript, painting, sculpture or carrier. May also be referred to as original physical item, physical item, physical original.

Physical carrier – See carrier.

Physical format carrier – See carrier.

Physical original – See original.

Portable media – See carrier.

Preservation action – A specific, defined and measurable task undertaken on a digital file or files for the purposes of stabilising the file(s) and/or making them accessible. Preservation actions can be reversible or irreversible. Each preservation action should be documented.

Preservation master – Digital content targeted for preservation that is considered the ‘master’ version of the intellectual content of any ‘digital object’. Preservation master files are created to high capture standards and may also capture additional information about the original, beyond the content itself. In extremely rare cases, preservation master files may take the place of the original file if the original is destroyed, damaged, or lost. Preservation masters generally do not undergo significant processing or editing (this takes place on co-masters). Preservation masters are often used to make other copies including reproduction and distribution copies. May also be referred to as an archival master.

Proprietary – In reference to hardware technology, software applications and/or file formats, the state of being privately owned and controlled. A proprietary design or technique implies that the company has not divulged specifications that would allow other companies to duplicate the product.

Provenance – The origin, or the source of something, or the history of the ownership or location of an item or object. The term is used in a wide range of professional fields including, art collection, archival management, librarianship, computing and law. In most fields, the primary purpose of provenance is to confirm or gather evidence as to the time, place, and – when appropriate – the person responsible for the creation, production, or discovery of the item or object. Also referred to as chain of custody.

Reproduction copy – See access copy.

Shell record – A brief record, or stub record, that is a placeholder for a full catalogue record and contains a minimum of descriptive metadata.

Sidecar file – Computer files that store data (often metadata), which is not supported by the format of a source file. There may be one or more sidecar files for each source file. They may also take the form of ‘metadata databases’, where one database contains metadata for several source files. In most cases the relationship between the source file and the sidecar file is based on the filename; sidecar files typically have the same base name as the source file, but with a different extension (as they are different file formats). Most computing systems including Operating Systems, file systems, digital preservation systems and tools have no knowledge of these relationships. If the filename or file path of a sidecar file is modified, or if the sidecar file is removed, the relationship to the file is lost. The worst-case scenario is that the file dependent on the sidecar file can no longer be rendered. They are also known as buddy files or connected files.

Supporting documents – See associated materials.

Transfer device – A temporary carrier for digital content. Transfer devices are typically external hard disk drives or USB flash drives. Also see carrier.

Trusted Digital Repository (TDR) – A trusted digital repository has been defined as having ‘a mission to provide reliable, long-term access to managed ‘digital resources’ to its designated community, now and into the future’. The TDR must include the following seven attributes: compliance with the reference model for an Open Archival Information System (OAIS), administrative responsibility, organisational viability, financial sustainability, technological and procedural suitability, system security, and procedural accountability. The concept has been an important one particularly in relation to certification of digital repositories.