The University of Manchester Library

Digital Preservation Strategy 2012

Contents

1. Background ...................................................................................................................... 2
2. Digital Preservation Strategy ...................................................................................... 3
3. Context and dependencies ........................................................................................... 4
4. UML digital collections ............................................................................................... 5
5. Digital Preservation Steering Group: Terms of Reference............................................. 7
1. Background

1.1 The rise of digital collections held by libraries in the last decade has meant that the preservation of these materials has become an ever increasing topic of concern and debate.

1.2 The University of Manchester Library (UML) recognised the importance of its digital collections and a requirement to develop appropriate mechanisms to preserve these in its 2009 Strategy statement, “New Directions, Key Result Area 4.3.4. We will develop a strategy for the curation and long-term preservation of born digital archive material and other digital assets based on appropriate international standards.”

1.3 The UML Digital Preservation Strategy Group was convened to discuss issues around digital preservation and develop an appropriate strategy for UML. The Group prepared an interim report of its work in December 2009 with the following recommendations, all of which were accepted and actioned.
   a. UML’s work on digital preservation initially focuses on content it directly supports
   b. UML undertakes small-scale ‘pilot’ studies of content representative of the Library’s digital collections
   c. UML establishes a ‘Digital Preservation Steering Group’
   d. The proposed digital preservation strategy is reviewed in consultation with an external body

1.4 In the past digital preservation has been shrouded in mystery, misunderstandings and technical confusion. The rate with which information technology changes, the growth in amount of digital material stored, the growth in storage capacity and diversity of digital materials now being created all act to compound these perceptions. However, in reality significant research has been conducted in this area. This research has resulted in a range of tools, support materials, use cases and best practices being developed which help libraries to resolve digital preservation issues.

1.5 UK based organisations (like the DPC, [http://www.dpconline.org](http://www.dpconline.org)) and projects (particularly those funded by JISC) as well as international organisations and consortia have made much progress in the development of digital preservation tools and practices. Noteworthy examples of these are,
   b. DCC curation lifecycle model ([http://www.dcc.ac.uk/resources/curation-lifecycle-model](http://www.dcc.ac.uk/resources/curation-lifecycle-model)): This “provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt”.
   c. Data Asset Framework (DAF, [http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework](http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework)): This “provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets”.
   d. Digital Repository Audit Method Based On Risk Assessment (DRAMBORA, [http://www.dcc.ac.uk/resources/tools-and-applications/drambora](http://www.dcc.ac.uk/resources/tools-and-applications/drambora)): This gives “a paper-based methodology for helping repository managers to develop a documented understanding of the risks they face, expressed in terms of probability and potential impact”.
   f. Life Cycle Information for E-Literature (LIFE, [http://www.life.ac.uk/](http://www.life.ac.uk/)): A methodology that allows organisations to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 20 years.
g. The PARADIGM online workbook (http://www.paradigm.ac.uk/workbook/): This was the outcome of work carried out at UML and The Bodleian Library which explored issues involved in the long-term preservation of born-digital private papers in the context of hybrid archives.


1.6 In summary, tools and advice now exist that help organisations adopt digital preservation practices. These are widely and freely available to Higher Education institutions and cover most aspects of digital preservation work. As a consequence, it is now the case that in broad terms it is only the will and effort required to apply these tools that prevents organisations from addressing their digital preservation concerns.

2. Digital Preservation Strategy

“*The University of Manchester Library aims to ensure the perpetual access to its digital materials in accordance with our content and collection management policies. We aim to minimise the risks of digital obsolescence on our collections for the foreseeable future. Our digital preservation mechanisms will monitor and implement preservation actions in a co-ordinated manner.*”

To achieve the strategic aims, we have adopted the following objectives:

a. The *UML Digital Preservation Steering Group* will,
   - provide policy, guidance and technology to enable effective digital preservation
   - implement and monitor the digital preservation strategy
   - consider not just the short and medium term, but also the genuine long term implications of digital preservation for UML
   - analyse the risks of digital obsolescence to UML’s collections and commission appropriate preservation actions
   - utilise life cycle modelling to cost and plan UML’s digital preservation activities
   - align the digital preservation strategy with UML’s other strategic goals and related policies
   - Be pragmatic, relevant and a point of reference and expertise on digital preservation for the Library and wider University

b. **Develop a robust and sustainable framework for preserving UML’s digital collections** - towards this objective we will,
   - pursue a range of digital preservation approaches that are flexible, cost effective, fit for purpose, and mitigate the long-term failure of our preservation strategy
   - ensure effort is not wasted by, for example, using existing standard tools and approaches developed and utilised by the wider Digital Preservation community
   - implement preservation activities with appropriate timeliness and planning
   - ensure that access is provided for digitally archived objects using appropriate metadata standards (within recognised IP and data protection restrictions)
   - recognise and address issues relating to the wide range of digital resources within UML
execute a communications plan to gain buy in to digital preservation across UML and the wider University
embed knowledge and experience of digital preservation issues within policy and practice across UML
work towards reducing the frequency and complexity of our digital preservation activities by encouraging and assisting content creators to undertake digital preservation actions
work with the University's IT Services Division and related services to ensure that our digital storage infrastructure is sustainable and best fit for long-term digital preservation

3. Learn from, implement and contribute to digital preservation best practices - we will do this by,

- applying knowledge learnt from our experience with the preservation of older materials to ensure our preservation of current materials is effective
- applying best practice from the archive and library world where it is applicable for the preservation of digital materials
- sharing our experience with and learning from other organisations through selective national and international collaborations
- working towards a position where the Library is viewed as a focus of expertise and support in digital preservation for the University of Manchester

3. Context and dependencies

3.1 The Digital Preservation Strategy Group identified a number of policies which were either enacted or under development that related to the formation of a digital preservation strategy for UML. The Group believed a sustainable digital preservation strategy was more likely to succeed if it aligned well with these existing policies. Related policies are listed below.

3.2 Overall UML’s existing policies, including its Content development and Collection Management Policies (see 4.6), Digitisation Policy (see 4.7) and Manchester eScholar Preservation Policy (see 5.8) are well aligned with the proposed Digital Preservation Strategy.

3.3 Further consideration is necessary to determine how best to integrate the Digital Preservation Strategy with the preservation of non-digital materials and what principles of the Preservation Advisory Centre (PAC) and International Federation of Library Associations (IFLA) apply (as adopted by UML’s Collection Management Policy, see 4.6).

3.4 Additionally, advice is needed on what impact the University’s Information Governance Policy (http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=609) has on the proposed digital preservation strategy. This replaces the University’s previous Data Protection, Freedom of Information and Records Management policies. As a consequence, the Information Governance Policy will impact on the preservation of materials which contain personal details and the preservation of the University’s corporate digital records and systems.

3.5 The University Information Security Policy (http://documents.manchester.ac.uk/display.aspx?DocID=6525) will be relevant to the Library’s digital preservation activity.

3.6 UML’s Content Development and Collection Management Policies (http://www.library.manchester.ac.uk/aboutus/policies/) state,

- “The Collection Management division is responsible for … preservation and conservation of Library materials” (Collection Management Policy, Section 1)
- “The Library adheres to the principles for the preservation and conservation of library materials of the Preservation Advisory Centre (PAC) and International Federation of Library Associations (IFLA). Material in original format will be preserved for as long as possible and access will be provided to
original and surrogate formats as appropriate. Material with intrinsic value will be preserved in original format in perpetuity” (Collection Management Policy, Section 5a)

- “The Library applies preservation standards for materials and procedures where they already exist and actively supports research and development of new national and international standards. The Library supports scientific and technical research to address problems encountered in the preservation of library and archive material.” (Collection Management Policy, Section 5g)

- “The Library believes that ongoing and routine collection maintenance will reduce the likelihood of serious damage to its collections. …We are working towards a digital preservation strategy to underpin the long-term retention of digital material” (Collection Management Policy, Section 5f)

3.7 UML’s Digitisation Strategy states,

- “Where necessary the preservation needs of the collections will be paramount in all digitisation processes, to ensure that material is preserved and not put at risk” (Section 3.1.3)

- “We will follow a preservation strategy for our digital collections as set out in the separate digital preservation strategy document” (see Section 3.1.3)

- “We will aim to digitise analogue sound and film recordings within archive collections where this would add value and where it is consistent with the Digital Preservation Strategy” (Section 3.1.3)

3.8 Manchester eScholar’s Preservation Policy states,

- “Items will be retained indefinitely under normal operating circumstances”

- “We will try to ensure continued readability and accessibility of content”

- “The original bit stream is retained for all items, in addition to any upgraded formats”

4. UML digital collections

4.1 The UML’s digital collections cover five broad areas:

- born digital records acquired from external sources (e.g. Carcanet Press email archive)

- Digital surrogates/products of digitisation activities (e.g. CHICC image collection)

- Institutional archival records (e.g. documents relating to the merger of UMIST and The Victoria University of Manchester)

- records held in Manchester eScholar (e.g. RAE2008 submission).

- licensed content (e.g. JStor journals)

4.2 In summary, UML hosts a significant range of digital content which needs preserving. From our simple analysis it was clear that this diversity means that each collection should be evaluated for digital preservation on its own merits and case by case. Furthermore, no single preservation approach will satisfy all content.

4.3 However, on the whole the DPC assessment criteria and DCC life cycle model we used were effective at teasing out features which would help decide on what digital preservation approach is best fit for the content and digital preservation actions are highest priority. The following bullet points outline some instances of scoping work undertaken in these areas to identify particular preservation requirements.

4.4 Our work on born digital content, identified the following specific issues for further consideration,

- Relatively speaking the quantity of records (~100’s) and sizes of files are small (kilobytes per file). However, the media in which these are provided can be diverse e.g. 3.5 inch discs, USB sticks, email attachments. There is little control over the media on which content is supplied and extraction of content from this can be problematic.
Digital content exists in a range of file formats. There is little control over what formats are supplied, and preservation needs and requirements are likely to vary greatly from one format to another. We may be able to preserve some formats in full, while for others all we may be able to do is preserve the bit stream.

Although, digital preservation is an important consideration, of the material fits with the UML content development policies, its acquisition is the priority. Digital material is likely to be accepted as part of a larger hybrid collection even if its preservation is deemed too costly. Under such circumstances we would need to inform the record creator that we may not be able to ensure full preservation of the digital material in its entirety.

In part or as a whole, the collection may be subject to copyright restrictions. Copyright clearance to perform preservation actions may be difficult to obtain.

Content will contain personal records which may contain sensitive information and where appropriate should be preserved in a secure environment.

Metadata would normally be created at JRL as part of the archiving process.

In order to fulfil obligations to existing donors and depositors, and to avoid compromising the ability of the Library to acquire recent and contemporary archives, formulating and implementing an appropriate digital preservation action for this content is high priority.

4.1 Our work on the Codex image collection, as an example of digital surrogates/products of digitisation activities, indicated the following,

- The quantity of records is reasonably high (many 1,000's) and the file sizes are large (multiple 10's of megabytes per file). The availability of a long-term sustainable storage solution needs to be resolved.
- File formats and the media on which the content exists are fully controlled. Network speeds between JRL and servers hosted in IT Services are slow for the transfer of large numbers of files.
- Materials are digitised primarily to enhance access and reduce the damage to the source material. Digital preservation is required to offset the cost of subsequent re-digitising the source material.
- Metadata requirements are minimal.
- Licensing agreements with the content originator would need to be updated to incorporate information on digital preservation.
- Digital preservation is at least for the time being less of a necessity with this type of content than with the born digital content considered above (see 6.5) and hence is a medium priority.

4.2 Documents recording the formation of The University in 2004 are examples of institutional records which need to be preserved. Our work in this area identified the following key features,

- In the sample file the quantity of records is low (100’s of files) and file sizes are small (kilobytes per file).
- File formats are semi-controlled and mostly Microsoft Office documents.
- The integrity of transfer of digital records needs to be transparent and recorded. Files are subject to the Data Protection Act and need to be preserved within a secure environment.
- Providing a secure and sustainable archival storage solution for these files would most likely satisfy the immediate digital preservation requirements for this content. As a consequence implementing an appropriate digital preservation action is low-to-medium priority.

4.3 The RAE2008 submission provides an example set of digital records hosted in Manchester eScholar. Key features of these records are,

- Relatively speaking the number files in this collection is high (many 1,000’s) and the files are a medium size (multiple megabytes per file).
- File formats are controlled. Files are stored on sustainable media.
- Content exists in a modern digital format (PDF) and is preserved to sustain access.
• Good metadata exists for all records.
• As content has been acquired in the last 12 months and is accessible for the immediate future, implementing an immediate digital preservation action is low priority.

4.4 The preservation and continued access to Licensed content is an ongoing issue for the academic community and two principal services ensure access can be maintained to journal subscriptions:

• Portico: “Portico is among the largest community-supported digital archives in the world. Working with libraries, publishers, and funders, we preserve e-journals, e-books, and other electronic scholarly content to ensure researchers and students will have access to it in the future.” (www.portico.org)
• UK LOCKSS Alliance: “a co-operative organisation whose goal is to ensure continuing access to scholarly work in ways that are sustainable over the long term” (www.lockssalliance.ac.uk)

5. Digital Preservation Steering Group: Terms of Reference

The Digital Preservation Steering Group is chaired by a member of the Library Leadership Team and has representation from across the Library. It will,

• Own and review as necessary the Library’s Digital Preservation Strategy as described in this document.
• Implement and review the Library’s policies and guidance materials around the preservation of its digital collections
• Establish and maintain a catalogue of and risk log for each of the Library’s digital collections
• Meet regularly to assess and review the state of the library’s digital collections
• Review requests for and initiate digital preservation actions
• Prioritise digital preservation actions
• Assess access requirements
• Advise on digital preservation requirements for new digital acquisitions
• Establish and monitor a digital preservation communication plan within the Library
• Raise awareness of digital preservation requirements and activities within the Library
• Establish and monitor a digital preservation training plan within the Library
• Report to the Library Leadership Team

March 2010, revised July 2012 (following an external review of the strategy by Charles Beagrie Ltd)