UNIVERSITY OF WESTMINSTER

RDSS Digital Preservation – Records and Archives Management Pilot
A Report for Jisc by University of Westminster, July 2018

1.0 Executive Summary

This report documents the activities undertaken by the University of Westminster Records and Archives Service in fulfilment of a grant award by Jisc as part of the Research Data Shared Service (RDSS) Digital Preservation – Records and Archives Management Pilot Programme, 1 August 2017-31 July 2018.

The report draws the following conclusions:

- There is a general lack of confidence within the professional community of archivists and records managers in dealing with digital records;
- Core archival principles can still be applied to digital records, and comparisons can be drawn with long-established archival practices;
- More sharing about digital records experiences is needed within the community, including the sharing of failures and mistakes;
- In order to successfully advocate for the custody of digital records, archivists need a stronger understanding of computer science;
- There is potential for the development of an interim solution, akin to the traditional records centre, as an interface between the live records system and the archive.

2.0 Context

2.1 University of Westminster Records and Archives

The University of Westminster Records and Archives Service manages the institutional records of the University and its predecessor bodies dating back to 1838; together with several deposited collections of teaching and research interest including architecture, town planning, Chinese visual culture and the Westminster Menswear Archive. Our archival holdings amount to over 1000 linear metres (including 85 linear metres of hanging garments), 500 Gb of born-digital records and 300 Gb of digitised images to date. They are managed by a professionally-qualified team of 5, (1 head of service, 2 archivists, 1 records manager and 1 curator), supplemented by project-based archive assistants and student interns.

Since 2006 the University has operated an active records management programme for semi- and non-current institutional records. The Records and Archives team centrally manages third party secure off-site storage for hard copy records[1], with 95% of the University's Professional Services departments actively engaged in use of the service. As a result of the active implementation of records retention schedules and secure records destruction, hard copy holdings remain steady at approximately 3500 boxes held off-site per annum.

We began to actively deal with electronic records in June 2015 with the acquisition of dedicated internal server space to store digitised images and born-digital material. This was followed by the procurement of Arkivum 100 storage in June 2016 and the implementation of the Arkivum-Perpetua product which integrates the Arkivum storage with a process application called Archivematica.[2]

[1] We have used the term 'hard copy records' throughout this report to describe all non-digital records, encompassing paper and analogue formats.
[2] See https://arkivum.com/perpetua/ and https://www.archivematica.org/en/ [Accessed 26/7/2018].

Archivematica is a web- and standards- based open-source application which allows the preservation of long-term access to trustworthy, authentic and reliable digital content. In August 2017, following successful testing of the Arkivum-Perpetua product, a business case was approved for its ongoing development as part of the University's digital preservation strategy for institutional records. The business case also included a move to a new archival software solution, AToM (Access To Memory) - an open source, web-based, cataloguing application - from Axiell CALM, in order to provide public access and to increase the discoverability of the University's archive collections.[3] The University's new public catalogue is due to go live in Autumn 2018.

As we began to develop policies and processes for dealing with digital recordkeeping and digital preservation, we started to reflect on how we deal with hard copy records and the extent to which these policies and processes were transferable. The different characteristics of managing digital records - especially the minimal time lag between use and transfer to the archive - led us to revisit and interrogate many of our long-standing processes. Increasingly we began to question the applicability of traditional archival theory and practice to digital records, whilst also acknowledging the need for a reliable theoretical base to enable us to deal with the growing volume of born-digital records in a practicable way.

We believe that we are increasingly moving away from the tradition lifecycle model of archives and records, with its distinct phases, to the continuum model where we are identifying at the point of creation, or even prior to it, the records' enduring value for the institution. The recent implementation of Microsoft SharePoint across the University of Westminster as its electronic document management system was an added impetus. For us, the main challenge was how we might embed a digital preservation workflow into the whole records creation and management process.

2.2 Literature review

There is a growing volume of literature on digital preservation available. Some of the material we found particularly interesting or useful includes:

Archives First: Digital preservation project, Viv Cothey and Cassandra Pickavance (2017) https://www.gloucestershire.gov.uk/archives/digital-curation/digital-curation-projects/archives-first-digital-preservation-project/

Overview of a born-digital archives access workshop held at Wellcome Collection, Victoria Sloyan, Simon Demissie, Alexandra Eveleigh and James Baker (2018) https://doi.org/10.6084/m9.figshare.6087194.v1

The National Archives Digital Strategy (March 2017) https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf

The Theory and Craft of Digital Preservation, Trevor Owens (LIS Scholarship Archive Preprints, created 15 July 2017) https://osf.io/preprints/lissa/5cpjt

These were supplemented by regular reading of several blogs:

Bentley Historical Library, University of Michigan
http://archival-integration.blogspot.com/2016/11/the-end-is-just-new-beginning.html

[3] See https://www.accesstomemory.org/en/ [Accessed 26/7/2018].

Digital Archiving at the University of York (Jenny Mitcham, Digital Archivist)
http://digital-archiving.blogspot.com/

Digital Preservation @ University of Glasgow
https://universityofglasgowdigitalpreservation.wordpress.com/

Jisc Research Data and related topics
https://researchdata.jiscinvolve.org/wp/tag/research-data-shared-service/

2.3 Jisc RDSS Digital Preservation – Records and Archives Management Pilot Programme

In July 2017 we were approached by Jisc and asked to participate in an extension to their RDSS Digital Preservation Pilot Programme[4], specifically to undertake a 12 month project (August 2017-July 2018) exploring the archival and records management aspects of digital preservation.

We proposed to investigate the theoretical and process issues around the selection and appraisal of born-digital and digitised records. We also wanted to explore how issues around the legal and financial implications of the early acquisition of digital records into an archive can impact relationships with IT departments and record creators; as well as exploring how trust is created in digital recordkeeping.

This report details the activities we undertook in fulfilment of the grant award; together with conclusions and recommendations for potential further work.

3.0 Webinar

After starting to use the Perpetua service, the University Records and Archives team had been having a lot of face-to-face meetings with other archivists to explain our digital preservation processes. The webinar was designed to freely share this experience with a wider audience, in line with the open-source nature of the tools we are using.

Invitations were circulated via the listservs and twitter, with attendees asked to register so that we would have a better understanding of the audience. In total, 89 people registered from the UK, Europe and Canada. Of these, approximately 60% were archivists and 25% were records managers, with the rest carrying out a variety of roles across the heritage sector. Respondents came from a broad range of sectors, although by far the largest group were from Further/Higher Education (30%). In terms of their understanding of digital preservation, 74% described themselves as understanding the theory but not yet putting it into practice. 10% of respondents were using either Archivematica or Preservica, and 20% were using a variety of open-source tools, but the majority had no software in place.

We used YouTube's Live functionality as this enabled us to present a slideshow with audio, with a chatroom running during the live webinar, and to upload the recording of the webinar immediately afterwards. We practised both the technology and the content in advance with colleagues at another institution, which was helpful and led to a number of re-writes. Even though we had only been engaged in digital preservation for a short time, we discovered that we were already taking a lot of knowledge about the subject for granted.

[4] https://www.jisc.ac.uk/rd/projects/research-data-shared-service# [Accessed 26/7/2018].

Titled *Work in Progress: reflections on our first year of digital preservation*, the webinar took place on 4 December 2017 and ran for just under an hour. We had 32 people present for the live webinar. A link to the recording was circulated the next day and it has had 355 views.[5] The webinar is available on our YouTube channel at https://www.youtube.com/watch?v=k-SVO6lQEZM

3.1 Lessons learnt

The URA team have continued to receive questions and positive feedback in response to the webinar recording. When we receive requests for face-to-face meetings now, our policy is to direct enquirers to the recording in the first instance, and then we are happy to answer any follow-up queries they may have. This has freed up staff time considerably, while still enabling us to share our experience with our professional colleagues.

Writing and recording this webinar allowed us the opportunity to reflect on the decisions we had made and consider whether we wanted to continue with the same approach. However it also pushed us out of our comfort zone because it required us to see ourselves as digital preservation practitioners. This a theme that will come up again in relation to the workshops. Like many archivists, we feel uncertain in this relatively new terrain and did not want to be presenting ourselves as experts. Nonetheless, our peers were regularly asking about our experience in this area and it was clear that we had something to share. The title of the webinar was designed to reflect some of this ambivalence and it also fed into our approach to developing the workshops.

4.0 Workshops

Following on from the webinar, the workshops were conceived as a forum for organisations already engaged in digital preservation. We were keen to make them software agnostic so that they wouldn't replicate existing user forums but would offer an alternative space for practitioners to discuss the application of archival theory to the digital arena.

As with the webinar, invitations to participate were circulated through the professional listservs and via twitter. In all 27 organisations applied to take part. In order to facilitate small group discussions, we limited attendance at each event to 18 people. Attendees were chosen to ensure that a variety of sectors, software solutions and length of experience were represented. Most participants attended both workshops, which enabled us to continue and develop conversations across the two days.

Both workshops followed the same format. In the morning there were two presentations exploring the archival theory behind the day's topic and presenting some case studies. After lunch the participants were divided into three groups for guided discussion. The workshops were held under the Chatham House Rule[6], which was emphasised at the start of each session. We believe that this was an important part of encouraging open conversation at both events.

4.1 Workshop 1: Records Management

Workshop 1 was held on 18 April 2018 and was attended by participants from a range of different public and private sector organisations. The theme of the workshop was current and semi-current digital records held in archives. Digital records often need to be transferred to an archive earlier than hard copy records might usually be transferred in order to ensure their preservation. The aim of the workshop was to explore some of the practical, theoretical and legal issues that can arise from archives holding records that are still considered current, or semi-current, by the institution.

[5] As of 24 July 2018.

[6] When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.

The first presentation in the morning discussed the subject of early records transfer in a holistic, lifecycle context, arguing that the requirement for early action introduces a new layer of problems for archivists and records managers. Three possible approaches were presented: (1) moving records directly from an active network into permanent preservation, (2) abstracting records from an EDRMS and (3) imagining what an ideal digital semi-current storage service might look like. The second presentation focussed on the issue of email preservation, drawing comparisons between US and UK approaches at the national level.

Both papers set the scene well in terms of giving participants a lot to think about and in providing different ideas to explore further in the afternoon workshop activities. Participants were divided into three groups for the activities, ensuring a range of backgrounds and experience in each group.

For the first activity, participants had been asked to bring examples of one success they had had with digital transfer, and one difficulty. These were shared within the group, with one chosen for the group to focus on for further discussion. Each group was asked to consider how the scenario was similar to experiences with hard copy records, how it differed, how best practice could best be disseminated, and how challenges could be met.

Group A chose to focus on 'data dumping', as several participants had identified this as an area of difficulty. Examples included a business department switching to a new IT system, with all previous data put onto a hard drive so the department could start afresh; and a personal archive donated on a USB stick with no organisation of the files. The group agreed that there were similarities to large disorganised accruals of hard copy records (e.g. those that arrived in bin bags), with the same need to quarantine records until preservation needs had been assessed. Differences included the potential for a much larger volume of records to be transferred, given the storage capabilities of portable media; the need to intervene sooner than with hard copy records due to the potential for unseen digital decay; the difficulty in persuading donors to do some kind of basic listing prior to transfer as many believe archivists can just search across the digital files; and the need to acquire information about the technological aspects of file creation in order to properly understand and preserve the context.

Group B chose to focus on a success, that of the transfer of a collection of digital-only records from an internal department. They felt that there were similarities in arranging the transfer of the collection, in determining any preservation issues and ensuring its accessibility, and in considering whether they had the capacity to store it. As with Group A, Group B expressed a need to perform basic preservation actions more quickly than they may do with hard copy records. However they also felt that it was easier to transfer a digital collection in its entirety, and then appraise post-transfer, than it would be with a similar volume of hard copy records, due to the practicalities of finding space for storage and processing.

Group C considered the issues around providing access to digital records, and how this may affect decisions taken during the transfer process and relationships with depositors and donors. While the aspects that need to be considered are the same as in the case of hard copy records, particularly in ensuring compliance with legislation, the group felt they could not be confident that they had done everything in their power to ensure this. For example, with hard copy records it is easier to put a physical barrier in place whereas there was perceived to be an inherent risk with digital records that they might reveal more than was intended. One example given was where the original filepath of a document had contained information that was not supposed to be made public. The group felt that there is currently no consensus on how to provide access to digital records – some repositories make them available in the search-room on a standalone PC, some provide online access with prior registration, and others provide unregistered remote access. There is a need to ensure that depositors understand from the start how access will be provided to their records, but it was also felt that there is a risk that this might lead to increased restrictions around the use of the records, or

potentially even prevent the deposit from happening. Conversely there is also a need, particularly with internal transfers, to ensure that depositors understand that there will not be instant access to the records after transfer, but instead a period of closure to enable appraisal and cataloguing.

With regards to best practice, there was a general feeling by all the groups that there should be more opportunities for sharing of workflows and examples, and of both successes and failures. Several participants believed we should move away from the idea of 'best practice', which carries notions of success, to simply discussing 'practice' more generally. This might help to avoid a 'one size fits all' approach and allow professionals to benchmark themselves against similar-sized organisations, rather than large memory institutions.

The second activity asked the groups to consider what the transfer and use of digital records might look like in 10 years' time. Discussion was inevitably wide-ranging but the following key themes emerged across all the groups:

- The role of increased automation in the transfer of records. The consensus was that a larger volume of records would be transferred to archives in the future, and that archivists would have less time to carry out any appraisal or cataloguing activities, therefore these would probably only occur at series level. However technical metadata will be generated for each individual digital file in a way that could not be done for hard copy records.
- Technical metadata will still require cleaning and interpreting by archivists. In particular, users will need to understand the fallibility of some technical data (e.g. where the creation date of a file reflects the date it was ingested into SharePoint, rather than the date it was actually created by the record's creator).
- Encouraging records creators to generate metadata at the time of records creation and/or transfer, particularly flagging up any potentially sensitive data.
- The difference between internal transfers, and deposits from individuals or external organisations. Even if we could create an ideal world transfer scenario within our own organisations, we cannot mandate the same policies for external deposits but need to be more pragmatic.
- The potential impact on the use of hard copy records is unclear. Records which have not been digitised and have comparatively minimal metadata are likely to be less prominent in search results. However, for some researchers the idea of seeing the physical version may have additional allure.

Following on from this, participants were asked to discuss what we ought to be doing now:

- Greater appreciation of the costs (time and money) of processing digital records. Some organisations have begun to carry out this work for themselves but it would be useful to have a consistent methodology and to see results being shared. Not only will this help with funding, but also in managing expectations as to why digital records are not available immediately after being accepted into the archive.
- Better awareness of IT for archivists, in order to understand the record creation environment – for example, in systems like SharePoint, the file format exists only as metadata. This has implications for understanding the diplomatics of digital records. If we are going to continue to be able to vouch for a record's authenticity and integrity, we need to be able to understand exactly what we are preserving.
- Investigating how users want – and expect – to access digital records and what different groups may require from these records. For example, some groups may be more interested in maintaining the original look and feel of the record, others in the underlying technical data. How do we ensure that all groups are catered for?

4.2 Workshop 2: Appraisal

Workshop 2 was held on 16 May 2018. The range of institutions was broadly the same as the first workshop, although some institutions chose to send different representatives. This workshop received more applications than the first, suggesting that this is an area in which many archivists feel unsure.

The morning began with two presentations. The first was a historical overview of appraisal theory, covering Schellenberg and the 'More Product, Less Process' methodology.[7] As well as a welcome reminder for the archivists in the room, this also proved to be useful for those who had come into the sector through less traditional routes. This was followed by a fascinating case study, looking at small scale appraisal in a large organisation, and large-scale appraisal in a small organisation. For the afternoon workshop activities participants were again divided into three groups, ensuring a range of backgrounds and experience in each group.

The first workshop activity asked participants to breakdown and define the activities involved in appraisal. As a profession we tend to speak about appraisal as if it is a single activity that occurs once in the lifetime of an archive. Discussions of digital appraisal have characterised it as an iterative process[8], due to the need to carry out some processes on the material at the point of receipt and others at a later date. This exercise was therefore designed to identify the various activities carried out as part of the appraisal process, and to highlight any areas which are genuinely different between digital and analogue appraisal.

Appraisal activities identified by participants included:

- Initial description of material by the donor/depositor
- Verifying legal ownership/copyright
- Establishing fit with collecting policy
- Format assessment
- Decision to accept the records
- Transfer of records
- Quarantine
- Disposal of duplicates/other unwanted documents
- Assessment of unstable/fragile formats

Significantly, none of the groups identified any appraisal activities that they carry out on digital records that do not have an analogue equivalent, and vice versa. Comparisons were made with unusual analogue audio-visual formats, hard copy records which had been damaged by water or pests, and analogue formats demonstrating rapid deterioration such as thermal paper.

This suggests that, while archivists may feel unsure around appraising digital records, their training and theory should stand them in good stead. Greater familiarity with carrying out these processes in a digital environment – either personally or through the sharing of professional experience – should help to increase both competence and confidence in this area.

The second activity moved on to considering the appraisal of digital records that do not have a hard copy equivalent. Examples were taken from the Paradigm Project's list of digital objects that might

[7] T R Schellenberg, *Modern Archives: Principles and Techniques* (Melbourne: F W Cheshire, 1956); Mark A Greene and Dennis Meissner, 'More Product, Less Process: Revamping Traditional Archival Processing', *American Archivist*, 68.2 (2005), 208-65 and Mark A Greene, 'MPLP: It's Not Just for Processing Anymore', *American Archivist*, 73 (2010), 175-203.

[8] For example, Victora Sloyan, Simon Demissie, Alexandra Eveleigh and James Baker, *Overview of a born-digital archives access workshop held at Wellcome Collection* (2018); Victoria Sloyan, 'Born-digital archives at the Wellcome Library: appraisal and sensitivity review of two hard drives', *Archives and Records* 37.1 (2016), 20-36; Kate Cumming and Anne Picot, 'Reinventing Appraisal', *Archives and Manuscripts* 42.2 (2014), 133-45.

be more expensive to retain[9] and included undocumented formats, complex or compound objects (such as email) and objects which were subject to digital encryption.

Issues discussed included:

- The difference between preserving digital objects and providing access to them. The former should always be possible, but the latter may be more complex, especially if some kind of emulation is required to get beyond the raw data.
- The importance of provenance and context in appraisal decisions, particularly where access to the digital objects isn't immediately possible. Understanding the potential significance of the digital records can help archivists to decide how much time and money to spend on trying to preserve them.
- A proper understanding of the long-term financial implications of preserving different types of digital records, so that archivists can make an informed decision about whether to accept them.
- The need for good relationships with expert third parties, and the potential for archive repositories to develop specialisms in the types of formats they can accept so that each repository does not have to be able to manage every type of digital format.
- The problems of adopting a data-repository-style approach of specifying in advance which formats the archive will accept. This could create problems where the experimental format forms a significant part of the digital object's research interest, such as early computer music and art.
- The necessity of realistic discussions with depositors around what can be confidently preserved and what will be riskier, as well as the need for migration or emulation in the future.
- A need to balance the wish to accept complex digital objects that a solution may be developed for in 5-10 years' time, with the likelihood of archivists having the time to return to those objects in the future, given the amount of digital data the archive will be dealing with then.
- An increasing practice of re-appraisal of records accepted into the archive and the need for de-accessioning after cataloguing, especially where cataloguing is machine-generated. Participants agreed that these decisions are often taken with little transparency at present but could be presented to researchers more confidently as part of responsible stewardship of the records.

5.0 Conclusions

Our engagement with the UK archive community through the webinar and the workshops we delivered has led us to believe that despite the wide variety of experience in digital records preservation, nonetheless most of the issues and challenges affect us all. The majority of participants agreed that digital records require earlier intervention compared to traditional hard copy records and that this presents a significant resource and expertise challenge for the archive profession. As archives increasingly accept digital material into their collections, providing access to these records is a growing concern with a lack of consistency across the sector as to how this should be addressed or how to manage the expectations of users who assume instant access following a deposit of digital records. Overall, a general lack of confidence within the community in dealing with digital records makes it difficult for us to create trust with our donors and depositors. However, fundamentally, the core archival principles are still applicable, even though they may at first seem to be challenged by digital records.

[9] http://www.paradigm.ac.uk/workbook/appraisal/appraisal-issues.html [Accessed 26/7/2018].

The following recommendations are made:

- More sharing about digital records experiences is needed within the professional community of archivists and records managers, including the sharing of failures and mistakes;
- In order to successfully advocate for the custody of digital records, archivists need a stronger understanding of computer science.[10]
- There is potential for the development of an interim solution, akin to the traditional records centre, as an interface between the live records system and the archive.

We hope the archive community, together with ongoing projects led by Jisc, DPC, and others, will begin to address these themes so that we can be confident digital practitioners and ensure the safeguarding of the future preservation of our society's digital output.

Anna McNally and Elaine Penn, University Records and Archives, University of Westminster
www.westminster.ac.uk/archives
archive@westminster.ac.uk

We would like to thank all the individuals and institutions who participated in the workshops; who attended, or subsequently watched, the webinar; and who provided feedback on the first draft of the webinar.

[10] We note the recent announcement by Liverpool University Centre for Archive Studies (LUCAS) of their consultation with a view to developing a Certificate in Computational Archival Science and look forward to more developments in this area.