CREATING AN END-TO-END PROCESS FOR IMPLEMENTING A DIGITAL ARCHIVING WORKFLOW

How we are putting theory into practice

Leo Konstantelos

University of Glasgow
United Kingdom
Leo.Konstantelos@glasgow.ac.uk

Emma Yan

University of Glasgow
United Kingdom
Emma.Yan@glasgow.ac.uk

Abstract – This paper presents the efforts of Archives and Special Collections (ASC) at the University of Glasgow to produce an end-to-end process for implementing our digital archiving workflow. This will be achieved by conducting a pilot project and in this paper we discuss the project, our methodology and outputs and outcomes.

Keywords – Digital archiving, workflows, case study

Conference Topic – From theory to practice.

I. INTRODUCTION

Archives and Special Collections (ASC) in University of Glasgow Library is responsible for managing, promoting, providing access to and supporting engagement with the Library's unique and distinctive collections. These collections increasingly involve digital materials and in developing our digital archiving service we have begun to put a number of processes in place. In 2022 we used one of our collections, the NVA Archive, as a case study to assess our digital preservation capability [1]. After completing this study, we moved on to a new pilot project which is the focus of this paper.

The purpose of the end-to-end digital archiving pilot project purpose is to produce an end-to-end process, related procedures and methods for implementing the ASC digital archiving workflow [2].

The pilot has been allocated ten months and runs from December 2022 to September 2023, and the only required resource is dedicated staff time to develop and deliver the project.

II. METHODOLOGY

We started by outlining a project plan on what we wished to achieve with the pilot. We evaluated our existing practices against the workflow and identified where we need to extend and develop our services.

Within the ten-month period we intend to employ the following methodologies to develop, test and evaluate the necessary framework for delivering robust digital preservation and digital archiving services.

We intend on achieving this by:

- Developing an archival forensics workflow
- Extending our quarantine methods for storage media
- Creating a digital processing action logging system
- Creating an access procedure for the Archival Forensics Lab
- Identifying our digital media holdings
- Creating a digital holdings prioritization tool

iPRES 2023

- Conducting a collections development policy review
- Exploring appraisal and description with archival forensics
- Exploring transfer of research data at end-of-life
- Updating the risk register
- Reviewing donor agreements
- Implementing an end-to-end digital archiving case study
- Involving the DPC in the review and evaluation of developed procedures

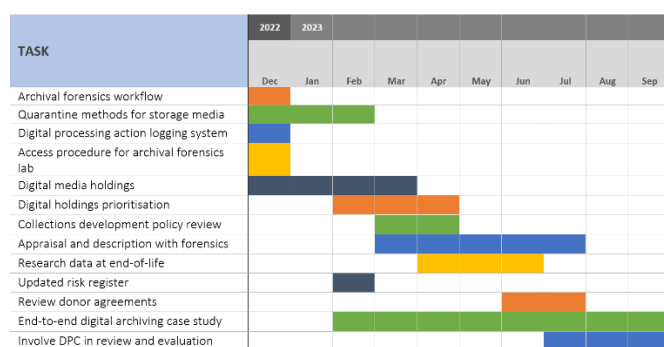The timescale for the work is planned on a Gantt chart, as follows in Fig. 1

Fig. 1: Pilot project timeline

As seen above, we are starting with getting systems and procedures in place and most time will be spent on surveying our digital media holdings, appraising and describing collections using digital forensics software and on the end-to-end digital archiving case study.

Further detail on what we will achieve is as follows.

A. Archival forensics workflow

We have developed an archival forensics workflow for digital storage media [3] that operates as both standalone and as an integration with the digital archiving workflow. This workflow will be tested during the end-to-end digital archiving case study.

B. Quarantine methods for storage media

We will work with the ASC Conservation & Preservation team to extend the current quarantine procedures with actions catering for the specific needs of digital storage media. The intended output is a revised quarantine procedure and guidance.

C. Digital processing action logging system

We are exploring solutions to log digital archiving and preservation actions to records, in a manner that aligns with the archiving and forensics workflows. Unfortunately digital archiving and preservation actions cannot be logged in a meaningful way in our collection management system, so we are creating a separate database to log actions to collections where data can be transferred between the database and the system.

D. Access procedure for archival forensics lab

We are developing a procedure for access to the Archival Forensic Lab that provides conditions of access the Lab. Effective implementation of this policy will minimize unauthorized access to the AFL and further safeguard the collections.

E. Digital media holdings

We are surveying our collections and creating a list of physical storage media held by ASC across our University Archive, Business Archive, Theatre Archive and manuscript collections. Using this list we will identify the risk of loss based on medium type and condition.

F. Digital holdings prioritization

We have devised a methodology and tool for prioritizing the digital archiving and preservation processing of current digital holdings, using community resources and good practice guides.

G. Collections development policy review

In consultation with other members of the ASC team we will review and amend the Collections Development Policy to incorporate aspects of born-digital collections and methods of acquisition for digital records. This will include determining acceptable formats based on our current and projected digital archiving capabilities.

H. Appraisal and description with forensics

We are using our suite of digital forensics tools to forensically appraise digital acquisitions, following the archival forensics workflow. We have a Digital Intelligence FRED Forensic Workstation and FTK software and using these we intend to explore ways to leverage forensic technology to (semi-)automate metadata generation. We will focus on this during the end-to-end case study.

I. Research data at end-of-life

We will explore mechanisms and methods for transferring research data that have reached end-of-life and are considered archival records. This work will involve liaising with the Library's Research Information Management team who manage the digital research data repository.

J. Updated risk register

Work on the risk register has been ongoing since February 2021 and we intend to finalize updates to the risk register and prioritized risks list.

K. Review donor agreements

We will review and amend donor agreements to include clauses for forensic processing of digital acquisitions, especially regarding data carving for recovery of deleted files; decryption and password recovery for protected files.

L. Involve DPC in review and evaluation of developed procedures

We intend to engage with the DPC to help with reviewing and evaluating the developed processes and procedures.

M. End-to-end digital archiving case study

One major piece of work in this pilot is the case study, where we have selected two digital acquisitions to process and document from beginning to end of the digital archiving workflow, covering all steps and actions; and recording time taken to complete each. The case study will allow us to test a number of the newly created procedures discussed above.

Once the study is complete we will produce an end report recording our progress and decision-making, which we intend to publish.

III.   ABOUT THE COLLECTIONS

We chose two collections to test during the pilot. Both are hybrid paper and digital collections and were chosen due to their size, complexity, content and perceived processing time as we want to ensure that we can complete the pilot project within the timeframe.

Our intent is to use these collections to test the workflow and while working though processing the collections, create policies, procedures and identify any sticking points or anything that is not working in the same way that we intended and make the necessary changes.

The first collection is the papers of Professor John Briggs, now an honorary research fellow at the University of Glasgow in the School of Geography and Earth Science, and previously Clerk of Senate and Vice-Principal (2012-2018) and Professor of Geography (1996-2012).

Briggs's research focuses on relationships between the use and management of natural resources and sustainable rural development in low-income countries, the impacts of structural adjustment policies in Africa on peri-urban development in the major cities and understanding the nature of agricultural landscapes in low-income countries. This collection, gifted to us in November 2021, focuses on his teaching materials as Professor of Geography, and the hybrid paper and born digital collection reflects the University's transformation from paper records to digital records, replacing print-outs and handwritten notes with PowerPoint presentations, Word documents and Excel spreadsheets. The digital material in this collection is fairly small in terms of size, amounting to 668MB, and was donated to us on one USB flash drive.

The second collection is the records of Dance House Glasgow, a creative arts organisation involved in supporting Glasgow's professional dance sector and offered community development programmes for over 20 years. In 2018 the company lost its Creative Scotland funding and ceased operating, and the collection was gifted to us via the Business Archives Surveying Officer in 2019.

The collection dates from c.1990 to 2018 and includes governance, financial, staff and project records, along with photographs, press cuttings, and promotional material. The digital material mainly consists of photographs and audio-visual material.

The physical extent of the digital material is 3 HDDs, 7 DVDs, 70 CD-Rs and 3 MiniDV cassette tapes. We know from our digital media holdings survey that one of the hard drives and 14 of the CDs were not working in January 2021, however we hope to interrogate this further using the archival forensic technology now at our disposal.

IV.   OUTCOMES

The pilot is a work in progress and at the date of writing this paper as per our intended project

timeline we have already completed some of the tasks outlined above.

The Archival Forensics Workflow is complete and has been published on COPTR's Community Owned Workflows webpage.

The access procedure for the Archival Forensics Lab is complete and is awaiting sign-off from ASC senior management. We will offer a staff training session at to inform ASC staff of the procedure as well as the appropriate actions to take to request access to the Lab.

The storage media prioritization methodology and tool are complete, as are the updates to the risk register and the database for logging digital preservation actions.

We have started work on the end-to-end case study which has prompted us to start on using digital forensics to interrogate the files and test the archival forensics workflow.

We intend to share our methods and outcomes with the wider digital preservation community, and by the time of the iPres2023 conference in September we will be able to give a thorough account of the end-to-end digital archiving pilot project and how we translated the theory to practice.

V. REFERENCES

[1] L. Konstantelos, C. Paterson, E. Yan, "Evaluating Digital Preservation Capability With Large At-Risk Collections: Lessons Learnt From Preserving the NVA Archive," in iPres2022, Glasgow, 2022, pp.282-286. [Online]. Available: https://www.dpconline.org/docs/miscellaneous/events/2022-events/2791-ipres-2022-proceedings/file

[2] University of Glasgow Archives & Special Collections, Digital Archiving Workflow https://coptr.digipres.org/index.php/Workflow:Digital_archiving_workflow_(high-level)

[3] University of Glasgow Archives & Special Collections, Archival Forensics Workflow https://coptr.digipres.org/index.php/Workflow:Archival_Forensics_workflow_(storage_media_deposit)