

CREATING CONTINUITY FOR DIGITAL PRESERVATION PROJECTS

Steps for archiving project outputs when Business-As-Usual has not yet been achieved

Edith Halvarsson

*Bodleian Libraries
University of Oxford United Kingdom*
edith.halvarsson@bodleian.ox.ac.uk
orcid.org/0000-0003-4634-3140

Sarah Mason

*Bodleian Libraries University of Oxford
United Kingdom*
sarah.mason@bodleian.ox.ac.uk
orcid.org/0000-0002-3156-1586

Abstract – This poster abstract summarises how the Digital Preservation at Oxford and Cambridge Project self-archived its research outputs, with the aim of extending the impact of digital preservation activities at the end of the project. **Keywords** – digital preservation, research data management, digital archiving, project outputs

Conference Topics – Designing and Delivering Sustainable Digital Preservation; Building Capacity, Capability and Community

I. INTRODUCTION

Digital preservation requires ongoing commitment to providing stewardship for digital collections over time [1, p. 128]. However, digital preservation is still not business-as-usual (BAU) for many organisations. One-off project funds is a common model for supporting digital preservation activities, but one which comes with risks to continuity. Digital preservation research outputs can easily become orphaned as project governance boards are dissolved and fixed-term staff move on to new roles [2, p. 242]. In this way, digital preservation projects are exposed to exactly the same risks as other academic projects. Questions of research output sustainability often does not come to the forefront until project funding streams end [3, p. 62].

This is an issue which should be taken seriously by organisations running digital preservation projects, as they are not immune. Issues such as web technology obsolescence, linkrot, as well as institutional memory loss has affected past digital preservation

projects.^[1] The Internet Archive is in some cases the only method of accessing these resources when there is no remaining digital preservation staff at an organisation.

The Digital Preservation at Oxford and Cambridge (DPOC) project ran for 2.5 years (from July 2016 to December 2018). DPOC aimed to break the project cycle for digital preservation at Bodleian Libraries (Oxford) and Cambridge University Library and establish BAU programmes [5]. Despite successfully acquiring further funding for digital preservation through the development of local business cases, the funding still sits within a project based framework. Acknowledging the risks which the project model poses to the continued stewardship of research outputs, the DPOC project undertook a self-archiving component in October-December 2018

II. PURPOSE

The purpose of the DPOC self-archiving exercise was to improve the usability, findability, and citability of digital preservation research and knowledge gathered during the project. The intent was to make future

[1] Examples of this type of loss includes the CASPAR project and DigCurV project. CASPAR's website <http://www.casparpreserves.eu/> has been inaccessible since 2015. DigCurV reports were published via <http://www.digcur-education.org>, which now redirects to a simplified WordPress site [4]. Some of the reports are now not available on the new site or from the Internet Archive's snapshots.

stewardship and decision making about project resources easier for staff at the libraries by making content 'ready for preservation'. By undertaking the exercise, the libraries can also use it as a practical applied example when speaking to academics about digital preservation and end-of-life scenarios for research projects.

This poster outlines the steps which the DPOC project took in order to self-archive project outputs. It is the hope of the authors that presenting this process will spark discussion around how the digital preservation field can manage continued access to its intellectual outputs in an environment where digital preservation BAU is still not the norm for many organisations.

A. *Step 1: appraisal of internal project outputs*

In the 2.5 years that DPOC ran it created a large body of internal documentation as an outcome of its research activities. The appraisal of research outputs was recognised as a key deliverable of the self-archiving component to ensure that staff who are not familiar with the DPOC project can easily navigate and make use of project outputs. The basis of all appraisal decisions was the project's Communication Plan developed at the start of DPOC. The Communication Plan outlined the intended use of all communication platforms (including email, instant messaging applications, task management tools, and SharePoint). It also outlined backup schedules, file naming, and versioning methods.

This was the most time consuming aspect of the self-archiving component as each document and tool was individually assessed. However, as file naming and versioning rules had been largely adhered to the preparation of content was made simple.

B. *Step 2: preparing content*

The second step was to create additional metadata and prepare content for the Libraries' RMS or institutional repositories – Cambridge University's Apollo and the Oxford University Research Archive (ORA).

Apollo and ORA were chosen as the natural homes of any externally facing project outputs as each individual dataset could be assigned a persistent identifier and be included in any future digital preservation actions performed on content in the repositories.

The Libraries' internal Wiki was used as the main source of internal information regarding the DPOC project. An about page regarding key deliverables of the project was added to the Wiki, which integrates with the Libraries' SharePoint instance and also links out to any code stored on the Libraries' internal GitLab. Additional information regarding each document was added to SharePoint to ensure that documents remain understandable even if the link between the Wiki and SharePoint breaks.

C. *Step 3: social media and blog*

The DPOC blog (www.dpoc.ac.uk) was the main method of exposing the DPOC research outputs before they were moved into the institutional repositories. www.dpoc.ac.uk is a WordPress site maintained by Bodleian Libraries. While the WordPress platform is among the less complex applications for Bodleian Libraries to maintain, it is still an application based platform which requires ongoing maintenance that may alter the functionality, look and feel of the DPOC blog. It cannot be guaranteed that files uploaded to the blog remain accessible and persistently citable over time.

For this reason, a snapshot was taken of www.dpoc.ac.uk using Archive.it. Once the snapshot is complete a static version of the WordPress will be created and the interactive features (such as blog comment functionality and Twitter integration) will be removed.

A social media dataset was also created in ORA which contains statistics, individual blog downloads, and Twitter identifiers for content covering the period July 2016 to March 2018. If www.dpoc.ac.uk needs to be closed in the future due to obsolescence or security concerns, the libraries can instead direct users to the social media dataset without having to undertake further work.

D. *Step 4: stewardship*

Finally - ownership of documents, repository datasets and the WordPress was formalised and assigned to a senior member of library staff. This role, or a successor to the role, will make curatorial and preservation decisions about any DPOC project outputs (as well as other follow on digital preservation project outputs) going forward.

III. CONCLUSION

Continuity of research outputs is a challenge when digital preservation still operates within a funding environment heavily reliant on project one-off monies [3, p. 62]. Although the goal of digital preservation is ongoing stewardship, applying research data management and self-archiving approaches may be necessary. Applying these methods may help an organisation retain some institutional memory around digital preservation, so that it does not have to start from scratch, even where there are gaps in project funding.

ACKNOWLEDGMENT

The authors would like to thank team members and project managers on the DPOC project, in particular Somaya Langley for her wealth of knowledge on digital preservation projects.

REFERENCES

- [1] E. M. Corrado and H. M. Sandy, *Digital Preservation for Libraries, Archives, and Museums*, 2nd ed. Lanham: Rowman & Littlefield, 2017.
- [2] E. Halvarsson, S. Mason, L. Pretlove, D. Gerrard,
- [3] S. Langley, and J. Mooney, "Could collaborative research between two major libraries help consolidate digital preservation and break the "project cycle"?" In *Digital Preservation in Libraries: Preparing for a Sustainable Future*, J. Myntti and J. Zoom, Eds., Chicago: ALA Editions, 2019, pp. 241–261.
- [4] S. Langley and E. Halvarsson, "Planning for the end from the start: An argument for digital stewardship, long-term thinking and alternative capture approaches," in *Proceedings of DCH 2017*, 2017, pp. 62–63.
- [5] Digcurv: Digital curator vocational education europe. [Online]. Available: <http://schreibman.eu/digcurv/>.
- [6] Digital preservation at oxford and cambridge. [Online]. Available: <https://www.dpoc.ac.uk>.