

THE LONG AND WINDING ROAD

Implementing an Open-Source Workflow

Karyn Williamson

abr dn plc archive
Scotland
Karyn.Williamson@abr dn.c
om
0000-0003-0426-2406

This paper will discuss the recent work of the abr dn archive to create and put in place a fully open-source workflow, the barriers faced, and what was learnt from the experience. abr dn plc, an Edinburgh based investment company, is the first financial institution to implement a fully open-source digital preservation workflow. The following text provides an analysis of the work undertaken by the archive thus far, and argues that the least important aspect of implementing an open-source workflow is the software. Our journey has definitely been a long and winding road with many wrong turns, misleading directions, and occasional problems with running out of fuel, but the outcome has been positive and archive colleagues are keen to find out where the road will lead them next.

Keywords - open-source, workflow, software, analysis

Conference Topics - We're all in this together; From theory to practice.

I. INTRODUCTION

abr dn plc, is an Edinburgh based investment company, which currently manages over 550 billion in assets with over 800 investment managers spread over 30 locations globally. The company was formed when Standard Life plc merged with Aberdeen Asset Management plc in 2017 to form Standard Life Aberdeen. A rebrand and name change to abr dn took place in 2021. The abr dn archive is responsible for the records of all these companies and their subsidiaries. The archive contains over 163 cubic meters of physical records and approximately 2tb of

digital material, including both born digital and digitized.

For the last 3 years the archive has worked with a range of internal stakeholders to put in place a robust, scalable fully open source digital preservation workflow which would link in with the procedures already established for physical collections. This paper will document and analyze the steps taken on this journey, the outcome as it currently stands and plans for the future. This paper will serve as a case study for those interesting in implementing an open-source solution within their own organizations', and will highlight the fact that with careful and focused planning, any institution can put an open-source workflow in place to protect their digital holdings.

II. CONTEXT OF OPERATION

The context in which the archive operates is essential to understanding the workflow put in place and the reasons why the decisions described hereafter were made.

Since abr dn was formed, it has been working through a landscape changing company transformation, including a full rebrand project which included a change of name. This has led to the work of the archive taking lower priority in terms of overall company strategy. This has had a direct impact on the budget and resource available to the archive for digital preservation work; both of which were already in short supply.

Although abrdn is a global company, there is no global archive provision and colleague resource has been limited to no more than 2 full time archivists at any one time since the archive function was formed. Less than 1/3 of this resource is allocated to digital preservation work, which had resulted in a sporadic at best approach to digital preservation. The approach at this time was to collect the digital records documented in our collecting policy, but not to process them in any way. This led to a backlog of digital processing work, for which the archive team had limited resource available and little knowledge that could be used to processing these records. A business case for digital preservation system funding was put forward which detailed the gold (preservica), silver (archivemata) and bronze (fully open source) routes to implementing a digital preservation workflow. The business case levels were defined based on the resource required to implement each option, ease of use for staff, the training and knowledge available about each option and the number of records that could be processed per day using each option. Ultimately, the decision was made to implement the bronze option with no other resource approved other than that needed for membership of the Digital Preservation Coalition as. It was recognized that the current skillset of the archive was not in a place to carry out this work without guidance and help. This only became more apparent as the project progressed, particularly in terms of security restrictions.

The cyber security restrictions in place to keep abrdn data safe are among the highest in the world and until this work began, the company had actually issued a blanket ban on open-source software being used on their systems. These restrictions combined with the financial regulations and compliance needs of the company were the backdrop to an already complex problem and it was clear that extensive research and planning would be needed at the outset in order to put a successful workflow in place.

III. WORKFLOW PLANNING

Initial approaches to working out an open-source workflow focused mainly on software, and it quickly became clear that this was the wrong place to start. Focusing on how tasks would be done meant we very

quickly lost sight of why we were doing the task in the first place. Despite being completely new to the world of digital preservation, we had tried to start at one of the most technical points in the process and our lack of technical knowledge led to us becoming overwhelmed and confused. Taking a step back and beginning with a holistic view of what was required allowed us to start at the beginning and work through what would be required at each phase of the workflow. This step-by-step approach also allowed for an ongoing period of research which help build up the skillsets and technical knowledge that would be required later in the project. Working in this way also made it clear that despite being at the beginning of our digital preservation journey, some of the processes and systems already in place across the company were compatible with our requirements. Although purely coincidence, this gave team morale a boost and helped us move forward with the process with more confidence and purpose. Our confidence was increased further when we started using the resources provided by the Digital Preservation Coalition.

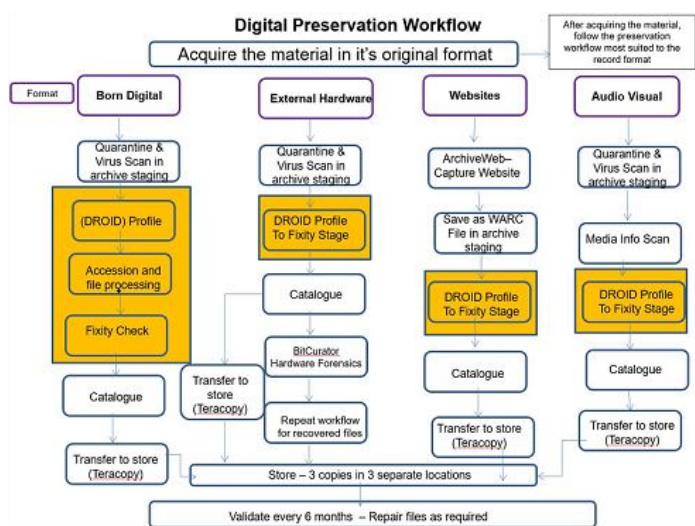
Many of the resources freely provided by the Digital Preservation Coalition are incredibly useful for beginners to the sector who are looking to implement a system in their organization.¹ The DPC RAM assessment helped us to map out our current position, and what would be needed to get us to where we wanted to go, and the DPC handbook outlined the steps we would need to take to get there. Consulting with the company IT department was also beneficial as they helped us understand where our workflow could merge with current company procedure and where a new process would need to be developed. Having access to this range of technical knowledge combined with input from the digital preservation community was essential to the success of the program. To build up the knowledge of our current position further, we combined the results of the DPC RAM assessment with the results of an NDSA levels of Digital Preservation assessment.² Completion of the NDSA levels matrix gave us a basic overview of where we were technically and how the requirements of the workflow being implemented might impact on the wider company. Having this knowledge would be

¹ Mentioned resources plus many more available here, [Digital Preservation – Digital Preservation Coalition \(dpconline.org\)](https://www.dpconline.org)

² NDSA guidance and matrix available here, [Levels of Digital Preservation \(ndsa.org\)](https://www.ndsa.org)

essential when we reached the next phase and began looking at software options.

Once the planning phase was complete, Fig.1 below is the workflow we chose to implement. The orange section highlights the repetition of the full workflow detailed in the born digital flow in the smaller orange section in the other flow lines.



At this point, no software was included on the workflow and so work began on finding the programs that were best suited to the tasks we wanted to perform.

As mentioned previously, software can be one of the most confusing and intimidating parts of the digital preservation process. The sheer volume of programs available is daunting in itself. In addition, technical experience varies greatly from one digital preservation practitioner to the next and not every open-source tool available is beginner friendly. The main digital preservation practitioner on this project trained firstly as an archivist working with physical papers and working with digital records was a learning on the job process which slowed things down significantly. The archive entered a prolonged period of testing to work out what programs best suited each role and finally settled on the software documented in Fig.1. The security restrictions and regulations the archive operates under made this section of the work plan particularly problematic. Although the company had approved the use of open-source software by this stage, it was under the caveats that any program being put on our systems had to be approved by the software approval board (SAB) and they would only approve programs which

had been robustly tested and were being used as part of the final workflow. The SAB would only approve software with a graphic user interface as security restrictions meant no access to programs running via the command line. To get around this I was given a laptop with no connection to the company servers with various data sets loaded onto it that I could use to test any program I wanted to and then confirm intended use with the SAB so it could go through the approval process.

This stage of the process took over 15 months to complete. The learning process around each program tested was complex and long due to the knowledge base of the archivist at the beginning of the project. Upon reflection, Implementing this workflow made it clear that time is not always a good indicator of the success of a project and that taking the time to get the right result for the archive was more important than having a set-up complete within a set timeframe. This section of the journey was particularly bumpy with a few pit stops taken to re fuel, re group and look at things from a different perspective to find the answers we needed.

IV. WORKFLOW TESTING

With all the chosen software in place, the next step was end to end workflow testing to ensure that each section worked not only independently but also as part of the workflow. Due to the earlier intensive planning and testing, this stage was relatively smooth. It was only at this point when the workflow was being tested and found to be working that the team felt like they were actually 'doing digital preservation'. From a beginner's perspective, it can seem like you need to be doing the technical parts and physically preserving records and using software to be carrying out digital preservation tasks, but this is definitely not the case. The planning and testing work carried out was essential to the success of the workflow being implemented and having worked through this process, it is clear that the earlier preparation work was more important than the final test phase. Very few issues required fixing and the workflow is now in use to archive abrdn records as part of BAU work tasks (business as usual). It took over two and a half years to get to the point where preserving the digital records of abrdn was integrated as part of the day job but looking back, the journey was worth it. We are the first

financial institution to implement a fully open-source workflow (that we are aware of) and the archive as a whole now has far more knowledge about dealing with digital records and the work involved in preserving them.

Although the workflow is in place and working well, we are still on the long and winding road with the end not quite yet in sight. The process outlined in this paper works well for small digital collections, but automation of various tasks is required to enable the archive to process larger collections. There is also work to be done around the file formats identified by DROID that we can't yet open or preserve. This includes a range of files stored on DVD, CD and floppy discs. The next step after this will be to look at our digitization processes and document them in a similar workflow that can link to the current digital preservation workflow.

V. CONCLUSIONS

The story of the journey the abrdn archive has been on is an important one, because it highlights that anyone in any organization can implement a digital preservation workflow that covers all required bases. The planning phase, as this paper documents, is the most important part of the process and shouldn't be rushed. Networking and attending training events held across the digital preservation sector is also invaluable to those at the beginning of their own digital preservation story. The connections made to those with more experience and listening to what others in the field are working on helps to challenge the imposter syndrome that many beginners feel when starting on their digital preservation journey as well as helping them to connect with others working in a similar space who they can learn from and collaborate with. The resources available online, including those from the Digital Preservation Coalition, are also an invaluable step on any beginner expedition into preserving digital records. The main lesson learnt on this journey, is that software selection is a very small part of the overall work involved in creating a digital preservation workflow. In addition, the software testing phase runs far smoother when informed by the planning phase, as the practitioner will know what is required of each program and can quickly discard programs that don't meet this predefined

criteria. To refer back to the title of this paper. The journey from complete beginner with very little knowledge to a fully working, continually developing digital preservation workflow for the abrdn archive has definitely been a long and winding road, and one that still has a lot of twists and turns before its final destination will be reached. But this shouldn't put off other beginners looking to start their own journey. Lau Tzu stated that "The journey of a thousand miles begins with a single step".³ Never has a truer statement been made that describes working in digital preservation. Taking each step on the journey one at a time and making sure you have a strong foothold before moving on, and not being afraid to ask for a helping hand (or a rest!), when needed will help ensure long term success and a fully functioning open-source digital preservation that meets all requirements. The most important step on this journey, is the one that gets you started.

³ Tzu Lao, *Tao Te Ching*, 1933