ARE WE WINNING?

Other measurables for Digital Preservation

Tim N.L. Evans

Archaeology Data Service

United Kingdom

timevans@york.ac.uk

0000-0002-2725-6730

This short paper reflects on the past 25 years of the Archaeology Data Service (ADS), a digital repository for UK heritage data. The paper focusses on the benefits of the Digital Preservation Coalition (DPC) Rapid Assessment Model (RAM). Experience of using the RAM has complimented the strengths of accreditation and has done much to focus on capacity and capability. It has also helped highlight successes that may otherwise be overlooked. For a smaller organization, these successes may not be intrinsically measurable in terms of bytes and processes, or the capacity to build a complex infrastructure, but simply demonstrating the reuse of the data so forensically preserved and the individuals that ensure that it matters.

Keywords – maturity modelling, designated community, data reuse, capability

Conference Topics – Building the Capacity & Capability.

I. INTRODUCTION

The ADS is an accredited digital repository for heritage data in the United Kingdom [1]. Established in 1996, the ADS preserve over 2 million unique digital objects encompassing heterogeneous data produced from fieldwork and scientific research. From simple beginnings, the ADS archive continues to expand at an ever-increasing rate as the importance of digital preservation permeates our community of depositors [2].

Over 25 years our core principles of digital preservation and access have remained. We have built a procedural framework conforming to the ISO 14721:2012 specification of a reference model for an Open Archival Information System (OAIS); our preservation by migration strategy works; we have successfully gained CoreTrustSeal (CTS) [3]. However, as we enter our mid-life phase, and prompted by both CTS review and the uncertainty of COVID, attention has focussed on what it has meant to spend a quarter of a century 'doing' digital preservation. As a relatively small organisation that (at the time of writing) receives no core funding it's often tempting to feel as if we don't do enough, especially compared to the impressive initiatives, technical applications and tools built by others. For example, some technical elements, such as full implementation of PREMIS [4] are not as advanced as we would like, and some procedural and policy areas such as copyright and ethics still have room for improvement. Outside of a timeframe and the goals of achieving accreditation, efforts to improve are often hindered by the realities of day-to-day activities and the constraints of the operational capabilities of a relatively small organisation. At times efforts to improve seem Sisyphean, and that the true objective measure of digital preservation success remains out of reach.

Being a long-standing member of the DPC we are well aware that the digital preservation community is many and varied, and despite our internal frustrations at the rate of progress we are in a relatively fortunate position. Historic core funding and partnerships has enabled the development of a repository system, including ingest. As an organisation whose only modus operandi is digital preservation, having to persuade colleagues of its merit has never been an issue. Thus the aim of this paper is not to reproduce a litany of woes and imperfections, but rather to highlight some of the more

17th International Conference on Digital Preservation

iPRES 2021, Beijing,China.

DOI: 10.1145/nnnnnnn.nnnnnnn

nuanced and often hidden challenges that reflect our current capacity and capability, and the successes that can be gained from tackling them in small ways, and that we hope would have relevance to organizations of all scales across our community.

II. The benefits of Maturity Modelling

In 2019, into this abyss of self-reflection emerged a new initiative, the DPC RAM. After the recent completion of the third ADS RAM involving senior management and all digital preservation related staff, it was clear that this continues to be a valuable tool not only for reflection and self-improvement but celebrating what we had actually done well. The basics were of course re-confirmed and that it was worth reminding ourselves that after 25 years all our data is retrievable, and nothing has been lost [5]. Most data have metadata to aid its technical and thematic reuse. Most data are in formats that we are confident we can migrate when required, and migrations have already taken place albeit with some lessons to be learned [6]. The successful accreditation via Data Seal of Approval (DSA) and then CTS tell us our processes are trusted, and that in general we are implementing a strategy that works.

Conversely, and as we've always known, there is always much to improve. However, it is important to note that in nearly each case where we felt more work was required to improve on an element, the overriding requirement was for those most precious of commodities: time and money. Time for staff to discuss and roadmap a desired result, and then the resource to implement this technically. In the particular case of the ADS and its home-grown repository system, this is normally additional resource for a full-stack developer.

As previously noted, the ADS receive no core funding. Depositors pay a one-off fee at the point where data has been successfully accessioned, and any systems-wide developments are normally covered by relatively small research grants. Improvements in our capability to do 'digital preservation' are often limited to what can be achieved, and fundamentally *should* be achieved subject to the capability within our organisation. Although our repository systems have the basic tools are in place to do routine tasks, they still need staff in place to run them, and assess results. Problems always need to be fixed. With an increasingly busy schedule of ingest and preservation this can leave less time for archivists and managers to spend on implementing change as fast as desired. Thus, one may look enviously at another organisation implementing a new preservation system, and too often do developments within the wider community appear (and sometimes) disappear without the time to understand potential cost benefits.

Our most recent appraisal of the DPC RAM often puts us tantalisingly out of reach of the top scores we *think* we need. However, this approach has done much to show what we have achieved beyond what we'd even realised and often beyond our capacity. Being able to undertake such an exercise - not tied to an ostensible marker of success such as CTS - is a success in itself. Indeed, the conversations that arise from the most recent exercise have done much to refine our notions of success away from purely just the technical, but back to ourselves and the reason we do digital preservation, our users.

III. Success is other people

This last point has been in-part stimulated by the COVID lockdown(s) in the UK. Over these periods, there has been a noted increase in use of ADS digital objects compared to annual average. Prompted by a mixture of ennui and curiosity, our designated community seems to be using our data more. Perhaps even, people from outside our designated community are now using the data for reasons we know nothing about! This of course leads to multiple questions that consistently arise over any discussion of designated communities in OAIS, namely who and why [7,8], but also, and regardless of how we classify our users, are they able to do anything with the data? This known unknown occasionally becomes less opaque when we are able to capture a reuse event [9], but the majority of the time remains as an assumption that a download of data equals a successful outcome.

These thoughts chime with a recent paper by Abrams [10] which opines that the attention of digital preservation specialists is largely on the trustworthiness of attempts to maintain digital objects, and less on the evaluative quality of success, and by association the experiences of the designated community that uses the data. Thus, while we can point at usage statistics as proof of success, we're never entirely sure that is what the community wants or needs without the capacity to engage the target audience and encourage them to feedback.

From previous experience at the ADS this is difficult, user surveys and pleas on social media often yield fairly low response rates, and conflicting opinions. For example, some users are happy with a PDF download that they can print, others desire embedded tools for visualisation, others request capacity for downloads across datasets to feed the next generation of data heavy analytics. However, at least trying to understand that 'communicative experience', understanding if our users are readers or data users, and capturing the elusive positive experiences is a clear marker to/of success in itself. In our case study, being able to identify - through technical or more human means - at least a small number

of case studies and then highlighting them back to our designated community, for example through a series of guest blogs [11], helps to publicly and internally close the circle of OAIS. In other words, this is why we do it.

IV. IS OUR DATA USEFUL?

Associated with this measure is, in our opinion, a need for a greater understanding of reuse value. All data is useful, but within a designated community some data may have more use than others. In our case, over half of our digital objects are raster images, most of which fulfill a need of an archaeologist to record, but not to understand or convey any meaning. Across the spectrum of our data an archive of a PDF comic created so that children may understand the impact of aggregates dredging on marine archaeology [12], sits alongside an extensive database of Roman amphora types [13]. This is not to say one type is more important than the other, all collections we archive no matter the scale or content need to be preserved and accessible, but that within our remit should be a commitment to helping users understand what it is they are looking at.

This somewhat inevitably comes back to CTS, but also FAIR. Much has recently been written in recent years about FAIR, Digital Repositories, Accreditation, and the overlap therein [14]. Metadata is, as ever, key, in ensuring that the data objects deposited are reused and understandable to the intended user community, and that all metadata deemed necessary for this purpose are of sufficient quality. However, in our opinion, there is still work that can be done in enhancing the level of true understanding for our designated community beyond simple classification. Conversely, this is a tricky ethical and practical situation. Should a repository be 'grading' or reviewing work? In truth we already do via Ingest Operations and appraisal by ADS Archivists of the formats and metadata supplied by the depositor, if it's not good enough it is normally returned for correction.

Thus, a logical extension would be able to convey a sense of relative FAIR-ness in the archive itself. For example, that because of the quality of metadata, formats used, variety of formats (i.e. tabular data in format such as CSV rather than embedded in a PDF) some data has a higher potential for innovative reuse. Arguably this sort of reflection is more common in the discourse of Archives, as opposed to the Digital Preservation community [15] and being able to implement some sort of consistent Policy subject to a multitude of concerns about how this is implemented is still some way off. However, in our opinion, there is however a need for this to at least be acknowledged and discussed by the repository. Perhaps due to the nature of collections and objects there is no need, conversely perhaps because of the spectrum of data more work focussed on how users are directed to data is valid. Again, at least discussing this and putting *an* emphasis back on access has been a small success and helped re-energise all staff on the needs of the designated community.

V. BUILDING AND MAINTAINING CAPABILITY

Amongst all of this still sits the issue of capacity and capability – what we can achieve with what we have but also finding new ways of improving without a large budget. Historically, within a small organisation, building a team of staff that can simultaneously straddle all the myriad facets of Digital Preservation is difficult. In both the RAM and CTS accreditation there is frequent mention of an area having a defined 'role'. In the case of the ADS, many members of staff have multiple roles – from oversight of integrity checks, security, IT refresh to the ideas of user engagement and the idea of what it is to be an archive or a repository (and which are we?). Most archivists have a diverse range of technical skills encompassing software expertise, programming and scripting languages, metadata standards and interoperability.

Even with an increasing number of Higher Education courses in the UK incorporating data management and Digital Preservation modules, these practical skills do not come fully developed off the shelf. In our particular case, since 2019 there have been three new members of staff inducted into the team of designated digital archivists. All have been deliberately aimed at entry-level to the profession, with onsite training dealing with internal procedures and external training - mainly through DPC events - covering the introductory elements of 'why' and 'how' digital preservation is undertaken.

This last point was highlighted again through ADS DigiCurv and the RAM, and indeed feeds into the perennial question of 'what is a digital archivist'? At the ADS we are fortunate that the current number of staff classed in this role sits at 5, and fundamentally these roles encompass a significant and varied amount of work. From dealing with user and depositor queries, helping inform and revise Policy and Procedure, and all the technical intricacies of bitstream preservation. That we have been able to induct three members of staff, and have them undertaking these tasks safely

With this has come a renewed interest in DigiCurv, for assessing where staff are in terms of appropriate knowledge and skills, and where gaps in training are. This in turn, and again prompted by the DPC RAM as a *collective* exercise, has led to more reflective assessments of success. For example, this may be as simple as ensuring organizational-wide knowledge that a particular Policy exists, how it is updated and who

updates it. Conversely, this has also highlighted where important knowledge is restricted to a small number of individuals, or in some cases just one. This then has two successes – knowledge that the problem exists, and then identifying who within the cohort can deputise, or else develop a portfolio of skills and knowledge that means they can contribute.

VI. DISCUSSION

The overarching question for this paper has been trying to establish different measures of success beyond the obvious. This is of course not to diminish fundamentals of digital preservation, to to suggest we should focus less on technical elements. Rather, the success of importance to us are being able to understand our capabilities, understand that failure to achieve as much as we want is not in itself a problem (all knowledge is good!), demonstrating reuse and value to our designated community, and having the ability to maintain our capability through staff turnover.

This last point is arguably the most important as the world of digital preservation expands. Organizations do and will vary significantly in their resources, and small numbers of staff may be expected to do the work of larger numbers. The capability of institutional or external repository systems to do everything and with flexibility, will also vary. Perhaps one thing for the community to think about is not only about the next generation of tools and applications, but the next generation of those undertaking the work or operating the system. Fantastic work has already begun in skilling the sector, and we hope the community will continue to combine to develop a range of materials and approaches that continues to build the human capacity to do the job.

ACKNOWLEDGMENT

With thanks to Owen Evans for the title of the piece, and Everton Football Club for the overarching sense of dread.

REFERENCES

[1] Archaeology Data Service: Our Work, 2021. [Online]. Available: https://archaeologydataservice.ac.uk/about/ourWork.xhtml

[2] Archaeology Data Service: Annual Report 2020-2021, 2021. [Online]. Available: https://archaeologydataservice.ac.uk/about/annualReports.xhtml

[3] P. McKinney, "Conformance With PREMIS", in Digital Preservation Metadata for Practitioners, Springer Press, 2016, pp. 247-257.

[4] Archaeology Data Service: Accreditation, 2021. [Online]. Available: https://archaeologydataservice.ac.uk/about/accreditation.xhtml

[5] T.C. Wilson, "Rethinking digital preservation: definitions, models, and requirements", *Digital Library Perspectives* vol. 33, no.2, May 2017, pp. 128-136. https://doi.org/10.1108/DLP-08-2016-0029

[6] K. Green, K. Niven and G. Field, "Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service", *International Journal of Geo-Information,* vol.5, no.44, March 2016. https://doi.org/10.3390/ijgi5040044

[7] L. Talboom and D. Underdown, "Access is What we are Preserving': But for Whom?", 2021. [Online]. Available: https://www.dpconline.org/blog/access-what-we-are-preserving

[8] R.S. Bettivia, "The power of imaginary users: Designated communities in the OAIS reference model", *Proceedings of the Association for Information Science and Technology*, vol. 53, no.1, 2016, pp.1-9. https://doi.org/10.1002/pra2.2016.14505301038

[9] Archaeology Data Service: "Linking the virtuous circles: Citation and Tracking Reuse", 2020. [Online]. Available: https://archaeologydataservice.ac.uk/blog/2020/08/linking-the-virtuous-circles-citation-and-tracking-reuse/.

[10] S. Abrams, "Tacit attitudinal principles for evaluating digital preservation success", *Archival Science*, 2021. https://doi.org/10.1007/s10502-021-09360-5

[11] Archaeology Data Service: "Data Mining with past publications from the ADS: The search for Neolithic crannogs", 2020. [Online]. Available: https://archaeologydataservice.ac.uk/blog/2020/10/data-mining-with-past-publications-from-the-ads-the-search-for-neolithic-crannogs/

[12] Hampshire and Wight Trust for Maritime Archaeology, "Early Years Children's Book: Derek the Dredger [data-set]", 2008. https://doi.org/10.5284/1000110

[13] University of Southampton, "Roman Amphorae: a digital resource [data-set]", 2014. https://doi.org/10.5284/1028192

[14] M. Mokrane and J. Recker, "CoreTrustSealCertified Repositories. Enabling Findable, Accessible, Interoperable, and Reusable (FAIR) Data", in M. Ras, B. Sierman, A. Puggioni (eds), *Proceedings of the 16th International Conference on Digital Preservation*. 2019, pp 92-100.

[15] P. Conway, "Archival quality and long-term preservation: a research framework for validating the usefulness of digital surrogates", *Archival Science*, vol. 11, 2011, pp. 293–309.