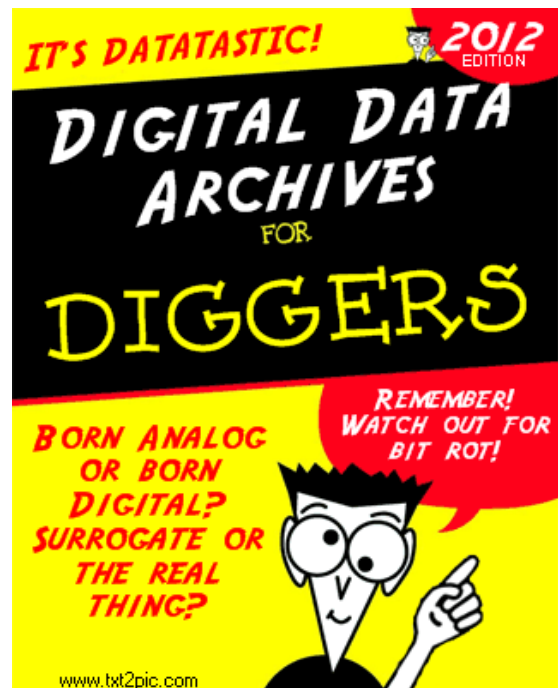


A Digital Afterlife

written by Jeremy Huggett | 02/09/2016

Solutions to the crisis in archaeological archives in an environment of shrinking resources often involve selection and discard of the physical material and an increased reliance on the digital. For instance, several presentations to a recent day conference on *Selection, De-selection and Rationalisation* organised by the Archaeological Archives Group implicitly or explicitly refer to the effective replacement of physical items with data records, where either deselected items were removed from the archive or else material was never selected for inclusion in the first place because of its perceived 'low research potential'. Indeed, Historic England are currently tendering for research into what they call the 'rationalisation' of museum archaeology collections



“... which ensures that those archives that are transferred to museums contain only material that has value, mainly in the potential to inform future research.” (Historic England 2016, 2)

Historic England anticipate that these procedures may also be applied retrospectively to existing collections. It remains too early to say, but it seems more than likely a key approach to the mitigation of such rationalisation will be the use of digital records. In this way, atoms are quite literally converted into bits (to borrow from Nicholas Negroponte) and the digital remains become the sole surrogate for material that, for whatever reason, was not considered worthy of physical preservation. What are the implications of the digital coming to the rescue of the physical archive in this way?

Selection and deselection of material and data have long been hotly contested. After all, how can we know whether something might be of future value to researchers? How can we know whether what we fail to record either through lack of resource or lack of recognition might prove vital to future study? There has always been a tension between what Martin Carver characterised as 'analytical destiny' (1985, 50) in which material/data have a known value and should be collected for an identified purpose, versus the importance of future, as yet unforeseen, value which would otherwise be lost but where there is a cost to its capture or retention. For example, material retrieved from topsoil is frequently seen as having little archival potential and hence capable of disposal, yet some sites exist primarily in topsoil with few closed contexts. What future research potential do we lose in machining off 'down to the archaeology'? Similarly, there are sites where the sheer quantity of material recovered meant only the rims and bases of vessels together with

decorated or stamped body sherds were retained, for example. Nor is this limited to decisions taken in the field: it applies equally to archived material collected many years ago. Could we have predicted that DNA could be extracted from ancient specimens for instance?

Several assumptions seem to underlie the perception that the digital can rescue the physical: that the lost or deselected physical can survive in a digital afterlife. Leaving aside the inherent fragility of digital data, these include:

1. A digital surrogate can adequately stand in place of a physical object

This assumes that the captured record can support the same analyses (and those as yet unforeseen) as the physical object. This is an unrealistic expectation, not least because those physical items being deselected or disposed of are by definition likely to be the least prepossessing and consequently risk a lesser level of detail being captured. How much effort would actually be exerted in recording the detail of a fragment of stone rubble from the infill of a wall, for instance? Or a fragment of clay daub with no impressions? 3D scanned data might enable the subsequent reconstruction of the object, but it would still lack many of the physical and chemical characteristics of the original, and in any case we have a long way to go for this level of data capture to be feasible in terms of time and resources. And the size of these kinds of datasets leads onto the next issue ...

2. Physical storage issues do not apply to the digital

This assumes that, unlike traditional archive stores, digital storage is essentially limitless – whether we simply add more drives to the server cluster or move into the cloud, there is no physical headroom as such. Indeed, it can seem as if space is not a problem – it's difficult to put numbers on such things, but, for instance, a report by the Federation of Archaeological Managers and Employers estimated there were some 2.2 gigabytes of as-yet undeposited digital material consisting of some 1.25 million files in the hands of archaeological contractors in England alone (Smith and Tindall 2012). Although the implication is that this is a lot of unarchived digital data, in a world of 'Big Data', this is pretty small scale! Even with the increased use of SFM imagery and 3D scanning since 2012, overall data size is not that large in the relative scheme of things. But this is naïve – it may be that digital storage is theoretically infinitely expandable, but this doesn't come without cost.

For example, David Rosenthal highlighted two reasons why we have tended to ignore cost in the past:

- There was the assumption that **Kryder's Law** (the equivalent of Moore's Law for storage) would continue for ever so that "if you could afford to store the data for a few years, the cost of storing it for the rest of time could be ignored" (Rosenthal 2014). In fact, as Rosenthal shows, storage is nowhere near as cheap as was anticipated (he estimates that it will be 100-300 times more expensive in 2020 than was predicted in 2010).
- It was assumed that the cost of access to the data could be ignored since "as the data got older, access to it was expected to become less frequent" (Rosenthal 2014). However, the cost of access has been underestimated by following an essentially material library-based model – if a book isn't used it can be put in the stack, and ultimately disposed of based on its access statistics – whereas, as we've seen, it is

difficult to predict the future value of data, so what is currently under/unused may yet become significant.

And all this is before we consider the cost of ingest, which Rosenthal argues is going to increase partly because most of the 'easy' content has already been incorporated leaving behind the difficult, increasingly dynamic content (Rosenthal 2014). Whether this is yet precisely the stage reached yet in archaeology isn't clear, but the warning is there. And we certainly know how difficult and time-consuming the generation of adequate metadata is, not least because what may be considered adequate now will likely not be in future.

3. The digital archive is not subject to the same selection/retention issues as the physical

This follows from the preceding assumption: that selection and retention is a problem for the physical archive and not a digital issue in an infinite (and cheap!) digital archive. And yet:

- As the ADS/Digital Antiquity *Guide to Good Practice* emphasises, there may be several potential Preservation Intervention Points requiring decisions about whether data are retained or discarded, recognising that future tools may make available presently unforeseen analyses of data. Consequently, it recommends that intermediate datasets should be retained as well as the final processed result, with corresponding implications for ingest, storage, etc..
- In addition to decisions surrounding Preservation Intervention Points there are decisions to be made about accession in the first place. For example, the ADS *Collection Strategy* (2014) provides several criteria for selection, including an assessment of their intellectual content and the level of potential interest in their reuse, and the extent to which they can be viably preserved and distributed in the future.

The review process isn't entirely clear, but is presumably primarily conducted by the curatorial staff, and, by implication at least, some data collections may be rejected. Furthermore, although the presumption is in favour of retention once archived, it remains possible for data to be subsequently disposed of (see ADS 2014, section 2.17). Arrangements at tDAR are less transparent from their website, but OpenContext uses editors to check consistency, integrity of identifiers, data cleansing, annotation with linked data standards, etc. as well as a process of peer review. In short, selection and retention issues are just as much a part of the digital world as they are the physical.

So the idea that the digital can rescue the physical archive is flawed in many ways – the digital is equally susceptible to the challenges faced by physical material, more so in some respects, and adds significant challenges of its own. Problems of physical selection and retention are not resolved by moving material into the digital realm; it isn't even the case that the shift to digital is adequate mitigation given the loss of information entailed in the process. But, resource implications aside, at least we have a choice of some sort with the physical material regardless of whether we create a digital surrogate: non-selection of born-digital data for preservation will almost certainly mean its loss, whether or not it is subsequently realised to be of value (see for example, Digital Preservation Coalition 2015).

References

Archaeology Data Service/The Digital Archaeology Record 2009 *Guides to Good Practice*.
<http://guides.archaeologydataservice.ac.uk/g2gp/Main>

Archaeology Data Service 2014 *Collections Policy* (6th Edn).
<http://archaeologydataservice.ac.uk/advice/collectionsPolicy>

Carver, M.O.H. 1985 'The friendly user'. In M.A. Cooper and J.D. Richards (eds.) *Current Issues in Archaeological Computing*, BAR International Series 271, (Oxford), pp. 47- 61.

Digital Preservation Coalition 2015 *Digital Preservation Handbook* (2nd Edn).
<http://handbook.dpconline.org/>

Historic England 2016 *Scoping Studies and Guidance for the Rationalisation of Museum Archaeology Collections* (August 2016).
<https://content.historicengland.org.uk/content/docs/grants/7359-rationalisation-museum-archaeology-collections-call-for-proposals.pdf>

Rosenthal, D. 2014 'Costs: Why Do We Care?', DSHR's Blog, November 18th 2014. <http://blog.dshr.org/2014/11/talk-costs-why-do-we-care.html>

Smith, R. and Tindall, A. 2012 *A Survey of Archaeological Archives Held by Archaeological Practices in England, Scotland and Wales*, Federation of Archaeological Managers and Employers (November 2012). <http://www.famearchaeology.co.uk/wp-content/uploads/2012/12/FAME-Report-on-Archaeological-Archives-Final.pdf>