

Digital Preservation at Oxford and Cambridge

A collaborative research project to evaluate and provide sustainable recommendations for our digital preservation programmes

Digital Preservation futurology

Posted on [8 September, 2017](#) by [Dave Gerrard](#)

I fancy attempting futurology, so here's a list of things I believe could happen to 'digital preservation systems' over the next decade. I've mostly pinched these ideas from folks like Dave Thompson, Neil Jefferies, and my fellow Fellows. But if you see one of your ideas, please claim it using the handy commenting mechanism. And because it's futurology, it doesn't have to be accurate, so kindly contradict me!

Ingest becomes a relationship, not a one-off event

Many of the core concepts underpinning how computers are *perceived* to work are crude, paper-based metaphors – e.g. 'files', 'folders', 'desktops', 'wastebaskets' etc – that don't relate to what your computer's actually doing. (The early players in office computing were typewriter and photocopier manufacturers, after all...) These metaphors have succeeded at getting everyone to use computers, but they've also suppressed various opportunities to work smarter, too.

The concept of ingesting (oxymoronic) 'digital papers' is obviously heavily influenced by this paper paradigm. Maybe the 'paper paradigm' has misled the archival community about computers a bit, too, given that they were experts at handling 'papers' before computers arrived?

As an example of what I mean: in the olden days (25 whole years ago!), Professor Plum would amass piles of important papers until the day he retired / died, and then, and only then, could these personal papers be donated and archived. Computers, of course, make it possible for the Prof both to keep his 'papers' where he needs them, *and donate them at the same time*, but the 'ingest event' at the centre of current digital preservation systems still seems to be underpinned by a core concept of 'piles of stuff needing to be dealt with as a one-off task'. In future, the 'ingest' of a 'donation' will actually become a regular, repeated set of occurrences based upon ongoing relationships between donors and collectors, and forged initially when Profs are but lowly postgrads. Personal Digital Archiving and Research Data Management will become key; and ripping digital ephemera from dying hard disks will become less necessary as they become so.

The above depends heavily upon...

Object versioning / dependency management

Of course, if Dr. Damson regularly donates materials from her postgrad days onwards, some of these may be updates to things donated previously. Some of them might have mutated so much since the original donation that they can be considered 'child' objects, which may have 'siblings' with 'common ancestors' already extant in the archive. Hence preservation systems need to manage multiple versions of 'digital objects', and the relationships between them.

Some of the preservation systems we've looked at claim to 'do versioning' but it's a bit clunky – just side-by-side copies of immutable 'digital objects', not records of the changes from one version to the next, and with no concept of branching siblings from a common parent. Complex structures of interdependent objects are generally problematic for current systems. The wider computing world has been pushing at the limits of the 'paper-paradigm' immutable object for a while now (think Git, Blockchain, various version control and dependency management platforms, etc). Digital preservation systems will soon catch up.

Further blurring of the object / metadata boundary

What's more important, the object or the metadata? The 'paper-paradigm' has skewed thinking towards the former (the sacrosanct 'digital object', comparable to the 'original bit of paper'), but after you've digitised your rare book collection, what are Humanities

scholars going to text-mine? It won't be images of pages – it'll be the *transcripts* of those (i.e. the 'descriptive metadata')*. Also, when seminal papers about these text mining efforts are published, how is this history of the engagement with your collection going to be recorded? Using a series of PREMIS Events (that future scholars can mine in turn), perhaps?

The above talk of text mining and contextual linking of secondary resources raises two more points...

* While I'm here, can I take issue with the term 'descriptive metadata'? All metadata is descriptive. It's tautological; like saying 'uptight Englishman'. Can we think of a better name?

Ability to analyse metadata at scale

'Delivery' no longer just means 'giving users a viewer to look at things one-by-one with' – it now also means 'letting people push their Natural Language or image processing algorithms to where the data sits, and then coping with vast streams of output data'.

Storage / retention informed by well-understood usage patterns

The fact that everything's digital, and hence easier to disseminate and link together than physical objects, also means better understanding how people use our material. This doesn't just mean 'wiring things up to Google Analytics' – advances in bibliometrics that add social / mainstream media analysis, and so forth, to everyday citation counts present opportunities to judge the impact of our 'stuff' on the world like never before. Smart digital archives will inform their storage management and retention decisions with this sort of usage information, potentially in fully or semi-automated ways.

Ability to get data out, cleanly – all systems are only ever temporary!

Finally – it's clear that there are no 'long-term' preservation system options. The system you procure today will merely be 'custodian' of your materials for the next ten or twenty years (if you're lucky). This may mean moving heaps of content around in future, but perhaps it's more pragmatic to think of future preservation systems as more like 'lenses' that are laid on top of more stable data stores to enable as-yet-undreamt-of functionality for future audiences?

(OK – that's enough for now...)

SHARE THIS:



Share

This entry was posted in [digital preservation](#), [Information Systems](#), [research data management](#), [strategy](#), [technology](#), [tools](#), [Uncategorized](#) and tagged [tool](#) by [Dave Gerrard](#).

Bookmark the [permalink](#)

[\[http://www.dpoc.ac.uk/2017/09/08/digital-preservation-futurology/\]](http://www.dpoc.ac.uk/2017/09/08/digital-preservation-futurology/) .

This site uses Akismet to reduce spam. [Learn how your comment data is processed](#).