

'In safe hands? Guaranteeing our collections for future generations'
RLUK/BLPAC seminar, British Library, 3 February 2012

To what degree can you use digital surrogates as a preservation strategy now?

Andrew Green

Llyfrgell Genedlaethol Cymru / National Library of Wales

My title takes the form of a highly specific question. It's a question that seems to expect a precise answer - say 78.5% - or at least to suggest that there might be a definitive answer one could give that would suit all circumstances.

But all that I can do is to offer a list of factors that I believe you would need to consider when finding your own answer.

I should make two things clear to begin with: that I hold no specialist technical knowledge in this area, and that in any case, and not only for that reason, I've no intention of addressing the technical challenges of digital preservation, such as coping with bit-stream degradation, migration versus emulation, or the obsolescence of hardware and software. That's not because technical considerations like these are irrelevant, far from it, but because there are other, non-technical factors that are just as significant, if not more so. I'll illustrate some of these from our own experience in the National Library of Wales.

Personally I find it helpful to substitute for the term 'preservation' the phrase 'sustained use', or 'continuing user access into the indefinite future'. There are several advantages to this way of seeing things. Politically, it helps in the often difficult arguments with governors and funders (in our case in the National Library, governments) about the value of preservation as an activity: 'access' and 'use' are in most contexts easier and more obviously appealing notions to convey than 'preservation'. Much more important, it encourages us to move our own thinking away from preservation-for-its-own-sake to preservation-for-use, and to get to grips with what is often termed the 'preservation intent' – the library's intention in preserving what it decides to preserve and choosing the precise methods to do so.

A good way of appreciating the difference is to take the case of microfilm. As a stratagem for long-term preservation of print or archive material, microfilm might be judged in the abstract as a success: adherence to standards and good storage conditions can guarantee the longevity of material, and microfilm is relatively cheap to create and store. But as a medium for *use*, sustained or otherwise, any reasonable person must regard it as a near complete failure: users detest microfilm as a medium with a rare passion, and these days resent having to travel to seek it out. Digitised copies, on the other hand, especially when

networked and freely available, will receive far more use than microfilm, and usually more than the original.

This is relevant to David Robey's recent useful distinction between technical sustainability and preservation. A particular technical solution may enable *potential usability*, but it may not be sufficient in itself to deliver *actual use*.¹

Factor 1: Preserving digital is inescapable anyway

The fact is that we're already obliged to attempt to use the digital as a long-term preservation medium. Most research libraries are collecting, as opposed to just renting, large and increasing quantities of born-digital material that they need to preserve for the future: electronic books, periodicals, archives, websites and many other materials. In future, some of us – and I'm referring here to the three national libraries – may soon face a sudden and vast increase in born-digital material to be preserved, in the form of electronic legal deposit publications². (The three libraries have already taken the decision to shift from collecting print to collecting digital in the case of parallel publishing.) There is no practical alternative to attempting to preserve all this material in something like its original digital form. It simply won't do to say that digital preservation is somehow too problematic to be other than a contradiction in terms: rather, it's an imperative to be faced.

Among our archivist colleagues scepticism on this point is still a regular position. To quote Gillian Oliver in a publication of 2012, 'the challenges and uncertainties of digital preservation are well established, and it is clear that the complexity required to preserve digital objects far exceeds the known requirements for the preservation of analogue materials'³. Librarians may feel they're able to take a different view.

My argument, therefore, is that if sustained use has to be guaranteed for this born-digital category, then there's no reason why we should shy away from considering digital surrogates as suitable for existing analogue objects.

Factor 2: Digitisation is mainstream, digitisation is expensive

By now digitisation is no longer an experimental or ancillary or peripheral activity. It's an integrated and substantial part of the everyday work of research libraries. The National Library of Wales has mainstreamed digitisation funding and strategy for over 10 years. We hold several million digitised items that we regard as worthy of preservation. We expect the

¹ David Robey, 'Improving sustainability of publicly funded digital resources'. in Lorna M. Hughes (ed.), *Evaluating and measuring the value, use and impact of digital collections*, London: Facet, 2012, p.149.

² Once a Regulation to implement the electronic provisions of the Legal Deposit Libraries Act 2003 passes through the UK Parliament.

³ Gillian Oliver, 'The digital archive', in Lorna M. Hughes (ed.), *Evaluating and measuring the value, use and impact of digital collections*, London: Facet, 2012, p.52.

volume of use of our collections to move increasingly in future from printed or archival analogue material to digitised material; this is a trend that is already under way.

As a result handling, publishing *and storing* digitised material are all familiar functions. Thus, libraries have an interest in ensuring sustained use through developing and adhering to the highest standards.

Digitisation is not only very common, and done on a large scale, it's also very costly. This means that libraries simply can't afford to imperil what is a massive investment by failing to give sufficient attention to sustained use of such material. This is true of 'special collection' digitisation and also mass digitisation of more mundane material. It makes no sense to digitise without having sustained use at the forefront rather than at the back of one's mind. No one wishes to re-scan or re-OCR because they have adopted standards that will not stand the test of time: it's always preferable to use high standards at the beginning. No one wants to fail to take adequate steps to ensure the preservation of digitised material.

Factor 3: Good enough is usually enough

What's the difference, for the user, between an original and its surrogate? There's an obvious distinction, of course, between interest in (and value derived from) the intellectual content of the original item, and interest in the information inherent in the item as a physical object – its materiality, or what archivists term 'intrinsic value' - for example, what can be learned from a detailed inspection of the condition of a binding, or manuscript marginalia in printed volumes.

Does the potential need to cater for both requirements inevitably mean retention of the original, and its production for use when requested? Not necessarily. Some types of material may be significant for their intellectual content but barely if at all for their qualities as objects. One needs to consider the *weight of use* of the digitised version compared with that of the original: a much greater degree of use may outweigh any reasons there may be for relying on the original. Further, many digitisation programmes seek to address 'intrinsic value' by offering the user access to high-resolution scanned images or other electronic surrogates as well as the transformed digital content, so reducing the need to see the original to a very small number of cases (these images are likely to be much more satisfactory than competitor surrogate media such as microfilm). To take another example, this time a non-digital surrogate, some libraries manufacture analogue facsimiles, to all but expert eyes indistinguishable from the original, as a means of preserving the original from overexposure or excessive handling.

Of course, these electronic approximations to intrinsic value can fall well short of users' legitimate expectations: an extreme case would be a digital copy of an old master oil painting, where even such sophisticated digital representations as those in Google's 'Art Project'⁴ are a very poor substitute for close physical examination.

⁴ Google's Art Project <http://www.googleartproject.com/>

Factor 4: What is the surrogate replacing?

It pays to ask, what function associated with the original is the surrogate replacing? What exactly is happening to the original? There is a range of options. One might destroy it, recycle it, or otherwise dispose of the original. One might withdraw it from immediate access and relocate it, maybe in a remote store or in another institution altogether (the National Library of Wales has 'repatriated' original parish registers to local record offices, having digitised and published their digital surrogates.) One might retain the original but adopt the equivalent of a 'do not resuscitate' policy, deliberately failing to prevent natural deterioration leading to eventual self-destruction (an example would be wood-pulp paper newspapers, certain to crumble to dust unless expensive de-acidification programmes are employed). One might retain the original, but not allow access, on the grounds of fragility or adequacy of the surrogate. (I recently had an interesting complaint from a user that he had been denied access to the original on the grounds that the digital surrogate should have sufficed – when in his view it did not.)

There are other possibilities but I've mentioned enough to make the point that literal substitution of the original lies only at one end of a spectrum of replacement and supplementation.

Factor 5: Can the digital surrogate extend the life of the original?

Another obvious question is, can a digital surrogate reduce use of the original item and so extend its life, even if it cannot entirely substitute for the original in every respect?

Or, the opposite: could the additional visibility afforded by digitisation encourage the desire by users to see and handle the original, so accelerating its physical decline? Could the process of digitising itself damage the original? Libraries will need to reach their own conclusions on these questions.

In the National Library our digitisation of the Black Book of Carmarthen, the manuscript volume containing the earliest surviving poetry in the Welsh language, induced an increased curiosity to see the original, not least in the town of Carmarthen, leading in turn to the making of a facsimile of the volume that could be exhibited and lent with no danger to the original.

Factor 6: Can a surrogate be more than a surrogate?

Sometimes a digital copy has the capacity to tell you more about the original object than you could glean from the object itself, or the object in isolation. We might call this the 'transcendent' effect of digitisation. If a photograph is scanned to very high resolution, the user may zoom in to light on a detail very difficult to make out in original. Likewise, linking the single digitised copy with large aggregations of related digital copies can provide a context for finding out more about the item: an example might be a series of historic maps of the same location, that one could impose digitally one on another to bring to light changes over time that are difficult to discern so easily by comparing originals.

On the other hand, a digital surrogate of an archive item, even a high quality one, might strip out the archival context, that is it might remove evidence of the complex relationship between the original archival object and the objects with which it was originally associated. And of course a digital surrogate may also lack the fixity of its paper equivalent, given the ease with which digital text can be corrupted or lost, for example in transferring data from one physical medium to another.

Factor 7: Does my digitised copy have to be the one to be preserved?

Assuming the original is not unique or very rare another relevant question is, are there digital copies held in other repositories which are committed to their long-term preservation? If so one could rely on these, without having to produce one's own digital copy or even to ensure the long-term future of one's own original. A good example is the Hathi Trust⁵, a repository set up by US libraries that participated in the Google Libraries Programme in response to anxieties about Google's commitment (or lack of it) to long-term preservation of the digitised collections they were creating.

Looking to the UK and the future, could one imagine a co-operative 'two (or three) last copies' policy being developed in the UK, such that a digital copy would allow un-rare originals to be discarded? The three national libraries are cooperating to share responsibility for preserving born-digital publications, under future legal deposit legislation: it might not be unreasonable for them and others to do something similar for material digitised in the UK. This kind of co-operative scheme may be the key to allowing individual libraries to be more selective and to switch scarce resources to prioritise the preservation of those items that are genuinely valuable.

Factor 8: What does my library's strategy say?

Probably the most important question to ask, though, when considering the appropriateness of digitisation as a sustained, not a current access stratagem, is how does it fit in with one's overall library strategy, and especially with one's digitisation, ICT, digital preservation and collection development strategies?

So, for example, are you committed to digital preservation of born-digital material? Have you guaranteed access, internal or external, to the required skills and storage infrastructure? What is your policy on the storage and accessibility of print or archive originals? What are the expectations of your readers in relation to the nature of their use of originals? Have you established policies on de-accessioning and discard? Do you share responsibility on preservation, of originals and/or copies, with other institutions? Or do you

⁵ <http://www.hathitrust.org/>. The first goal of the Trust is 'to build a reliable and increasingly comprehensive digital archive of library materials converted from print that is co-owned and managed by a number of academic institutions'.

store and preserve digitised material on behalf of other organisations, as a trusted digital repository?⁶

**

These, then, are some of the factors that will influence how you set about answering, for your own institution, the question, 'Can I rely on digitisation as a preservation strategy?'

I can hear you thinking, 'But what does he really believe is the answer to the question?' To satisfy your curiosity, there's no doubt in my own mind that a provisional or conditional 'yes' is increasingly likely to be given to the question today.

In 2000, in response to determined public campaigns on both sides of the Atlantic, whose most influential and vociferous advocate was the American novelist Nicholson Baker, the British Library retreated publicly from its then policy of discarding the originals of US newspapers and depending on microfilm as a sole surrogate. By 2012 most libraries, including the British Library and the National Library of Wales, have now closed their microfilming units, and hence their programmes of transferring print (newspapers) to microfilm as a long-term preservation medium. If there are going to be surrogates for newspapers today, they are, by default, going to be digital surrogates.

As long ago as 2004 the Association of Research Libraries in the US endorsed digitisation as 'an accepted preservation reformatting option for a range of materials', and urged its members to adopt standards, best practices and policies that would ensure the long term availability of digital surrogates⁷.

The clock has moved on much further since then – even if the answer is not yet perhaps 78.5%.

⁶ The National Library of Wales stores digitised theses on behalf of Welsh higher education institutions. A live question now is whether the Library could perform the same role in the case of institutional repositories currently held maintained by individual HEIs.

⁷ Association of Research Libraries, *Recognizing digitization as a preservation reformatting method*, ARL, 2004, p.1. http://www.arl.org/bm~doc/digi_preserv.pdf