Microfilm and Digitization as Choices in Preservation

by YOLA DE LUSENET

The first conference the European Commission on Preservation and Access (ECPA) organized, 'Choosing to preserve', took place in Leipzig in 1996. For the keynote lecture we had invited a scholar, Professor Bernhard Fabian. This was a very deliberate choice, as the ECPA subscribed to the view that academic researchers, as users of the resources kept in libraries and archives, have to be involved in discussions about their preservation. The conference had attracted quite a crowd, around 150 people who were of course all there for the opening lecture. We invited Professor Fabian because he was a very good speaker, who could really present a convincing case to an audience, which he did also on that occasion. He devoted a large part of his presentation of 45 minutes to the horrors of microfilm, showing us pictures of illegible film, with bits missing, that was impossible to use. He explained how distressing it was for scholars to be forced to use surrogates that don't do justice to the originals that they need to study.

Recently we at the ECPA heard from a preservation manager in a large library in Europe, that those in charge in his institution were on the point of dismantling the microfilming department and planning to use only digitisation from now on. This somewhat drastic decision was taken because microfilming was considered to be outdated and no longer necessary.

In between these two incidents lie seven years in which the scene has radically changed. At the time of the Leipzig conference, we basically had the choice between keeping originals and making microfilm, but now digitisation has entered the stage as the tall, dark and handsome stranger in a play. At the moment the question is whether this attractive hero will indeed carry off the innocent girl and, fickle and restless as he is, abandon her after having squandered her fortune. Or will he prove to be a reliable guy after all and lead her into a solid marriage - and to a happy end of the story?

In the more mundane reality of preservation, the choice seems to be between sticking with microfilm as a trusted and proven technology or moving to new technology that seems to offer so many opportunities -and great risks. Graham Jefcoate has described the new complexity in which the approach of 'digitising for access, microfilming for preservation', has started to break down. The problem with the new situation is that as digitisation is essentially not preservation-driven, preservation of originals threatens to become a subsidiary concern. What is digitised will probably be preserved, but what needs to be preserved is selected for digitisation only if there are strong arguments from a perspective of use. We can safely assume that only a fraction of what is available in

112

libraries and archives will be digitised in the foreseeable future. Low-use, specialist materials will not be among the first candidates. And if they happen to be at risk, other measures will still have to be taken for their survival

Basically no one likes using microfilm. Over the past weeks I discussed the theme of this conference with various people, and every time I heard how they hate microfilm. And, as a librarian once pointed out to me, to make matters worse, users of microfilm are often treated as outcasts and relegated to a bare, unpleasant corner of the library where they cannot disturb the other readers in their struggle with noisy machines. Obviously, most users nowadays would probably prefer or even expect a digital copy. That users have high expectations cannot just be brushed aside. For if users are demanding, they can never be wrong in asking.

Advantages for use are a strong argument for favouring digitisation over microfilming. However, what is worrying in the discussion about the advantages of digital images is that use is so often understood as a quantitative concept only. But if we define use in numbers, they may obscure the complexity that lies behind the straightforward lists of figures that can be plotted in a rising graph. Digitisation is a great favourite with politicians and decision makers because by opening up collections more people will be able to use them. So far so good. But there is use and use. Among the visitors of a site there are bound to be many casual users, window-shoppers. What does it mean if 10,000 people visit a site? How do we interpret user statistics to assess the impact of digitisation? If all we can measure with certainty is *frequency* of use, there is a risk that *increased* use becomes the measure of success, irrespective of what this use consists in and who the users are. This favours selection of things that are of some interest to a great many people, over things that are of extreme importance to only a few - scholars, for instance, who are *dependent* on these materials for their work. A sophisticated concept of use and a thorough understanding of target groups is needed to overcome the pressure of numbers and the risk of promoting trivial use at the expense of serious use.

What also confuses the discussion is that digitisation is often presented as a preservation measure because as a surrogating strategy it relieves stress on originals. This is of course true, but in many cases materials are selected for digitisation that are not directly at risk, or that have been microfilmed previously so that preservation is not an issue. The decision to offer a digital surrogate here is in fact not primarily motivated by preservation concerns, but by the desire to channel use. Arguments to digitise are for instance: providing access to treasures from a collection; distant access; a high frequency of use; the possibility to search, annotate or copy materials for research and educational purposes, or to contextualize materials and integrate them into a larger, distributed context. All of these are valid arguments, but none are preservation arguments in themselves. I would regard digitisation of valuable manuscripts as a preservation measure only if these materials are deteriorating because of frequent handling and no

other satisfactory surrogates are available to users. The distinction is important, because the aim of a digitisation project should determine requirements for quality and maintenance, and requirements for digital preservation masters that can act as substitutes may be different from requirements for user copies. The requirements for digital preservation masters may not be met if the real motivation for digitisation is facilitating use, and it may in fact be an uneconomical decision to create digital preservation masters, to be maintained over time, for originals that are not at risk or that have been microfilmed already. In trying to kill two birds with one stone, one may easily miss one's target altogether.

There are cases where the *primary* or only concern is to keep the information content of materials, without expectations of extensive use in the near future. This applies not only to the specialist materials mentioned above, but also to unique items, archival materials, or other materials an institution is legally bound to keep for the future, which may even include materials to which access is restricted. When the primary goal is preservation of information for materials that are not frequently used, the advantages of microfilming, which have always been staunchly defended by the preservation community, are still very real.

Although I agree wholeheartedly with Graham Jefcoate that because of the widespread introduction of digital born information, solutions to manage digital materials over time must and will be found, they still lack the essential stability that microfilm has. It *does* make a difference when things are tangible, can be stored safely and left for years or centuries and won't rot or fade or get corrupted. Microfilm doesn't require complex equipment to read and it doesn't become inaccessible of its own accord if it is stored properly. These are undeniable advantages over digital material, which is a nightmare for preservation –not because it *cannot* be kept accessible, but because keeping it accessible requires a continuing and intensive process of refreshing and converting, with a risk of something going wrong every step of the way. The carriers on which it is stored degrade rapidly, and the machines and the programmes to access the information become outdated within years. There is no process of slow fading or getting frayed at the edges, no, one day you find it is simply all gone, or the information may still be there, locked in its code, and you can't get at it – like a hungry castaway on a desert island with a tin of beans but no tin-opener.

There is another strong argument in favour of microfilming from a preservation perspective that is not heard much in library circles: *authenticity*. A microfilm image as a direct image of an original doesn't allow the information to be changed or otherwise tampered with as is possible with digital materials. Especially in the archival world the discussion on digital preservation centers around the issue of continued access to authentic digital records, because it is felt to be so problematic to keep the information as it was once created.

In short, although digital masters in practice will be created and will be preserved, from a preservation point of view digitising is still a high-risk approach. After all, you solve one preservation problem, saving the information from a deteriorating original, by replacing it with another, more complex one.

The research in the so-called hybrid approach was based on the assumption that the strengths of both technologies could be combined. The study at Cornell (Kenney, 1997) evaluated the use of high-resolution bitonal imaging to produce computer output microfilm (COM) that meets preservation standards; the complementary study at Yale (Conway, 1996) studied the production of digital images from microfilm. Both projects relied on high-quality microfilm as the preservation master and produced high-resolution digital images (600 dpi, bitonal). The conclusion of these studies was that in terms of quality and costs it is feasible to produce preservation microfilm from digital images as well as the other way around. The 1999 report by Chapman, Conway and Kenney presents a decision tree for determining whether to scan first or film first. Earlier this year RLG published guidelines on how microfilming can be optimised with a view to subsequent digitisation (Dale, 2003). I will leave the technical issues these reports discuss for other speakers who will have more to say about technology and quality control.

However feasible and safe in theory, the obvious problem with the hybrid approach is that it adds to the cost, not only because one has to do two things instead of one, but also because one has to deal with extra maintenance costs. This issue is not dealt with in the reports, as they focus on production, quality, workflow, and metadata.

At the moment there is not enough practical experience in long-term maintenance of digital materials to be sure about anything and certainly not about cost. By way of illustration, I would like to refer to a discussion of relative costs of storage from an article by Steve Chapman. He compares the cost of storing the same amount of information in analog format in the central Harvard Depository and in digital format in the OCLC Digital Archive. I cannot do justice to the discussion in the scope of this paper, and I would highly recommend everyone to read the complete article in order to come to a proper understanding of what the figures in this article reflect exactly. Chapman has made every effort to come to a well-defined comparison, so I think it is justified to mention some of his findings here, but to understand exactly what is covered by these costs and how they have been calculated, the full article should be studied.

Chapman presents figures for the relative costs of storing 729,000 pages of text, or 2,202 volumes, in book format, as microfilm - in the vault or in a standard environment - in the Harvard Depository, and as 1-bit page images and ASCII in the OCLC Digital Archive. Harvard charges per square foot, at a standard rate and at a higher rate if the material is kept in the film vault. OCLC charges per gigabyte, at three rates, depending on the total

amount of data deposited per account: the more is deposited by the account holder, the lower the rate per gigabyte.

Storage in ASCII is the cheapest option, because ASCII is an extremely compact format. It is, however, a format that would only be used if the appearance of the text, the page layout, was no issue and if there were no illustrations in a text. Storage on microfilm comes second (with the costs doubling when the films are kept in the vault). Third comes storage in book format, then storage as 1-bit page images. Even at the lowest rate (for an account of over 1,000 Gb) storage of digital images is 1,5 times as expensive as storing books, and 2,5 times as storing microfilm in the vault. It is assumed that lossless compression was used for the digital files; uncompressed, the files would become 22 times larger and storing them would become 54 times as expensive as storing the same number of pages on microfilm in the vault.

Chapman also compares storage costs for print and various digital formats. As costs are directly related to the number of gigabytes required for storage, file size is the decisive factor, and file size follows from the quality chosen for reformatting: bitonal, or 1-bit, gray or 8-bit, and colour or 24-bit. The effect on costs of storing masters at higher quality is overwhelmingly clear from his calculations: storing the same pages in 8-bit images instead of 1-bits is about 42 as expensive, as 24-bit images storage costs are about 125 times higher compared to 8-bits.

It should be borne in mind that in many cases the costs of storing film or digital files, and in the hybrid approach of both, will be *additional* to the costs of storing print material, as many institutions keep the originals after filming or scanning. And if, in order to limit risks, several copies of the film or the digital files are kept, this adds to the costs as well.

Storage is only one factor, and there are others to consider as well, but this comparison of storage costs alone makes clear that producing both film and digital, and maintaining both microfilm and digital preservation masters (which would presumably be high-quality files) increases costs very considerably. This almost inevitably leads to the conclusion that the hybrid approach is a viable option only for the materials that are so central and important that money is no issue.

However, research into the hybrid approach has been valuable for demonstrating how film and digital can be combined, if not in a simultaneous approach than perhaps in a phased approach. If preservation is the primary concern, one can choose to film so that the information is kept. The films, or some of the films, can always be digitised later should there be a need for including them in a digital collection. Or one may digitise from the film on demand, if a user at some point in time wants a digital copy. Conversely, if materials are digitised first, film can still be produced at a later stage,

should the need arise, for instance because it is not considered cost-effective to keep digital images which by then may only be consulted infrequently.

If the originals are stable enough to be kept for the long term, preservation is no issue at all, and quality can be defined on requirements for use**.** Lower resolutions or compressed files may be adequate and reduce costs. It is often said that requirements will go up in the future, with high-speed networks and new types of monitors, but perhaps it is cheaper to rescan the originals later than invest in uncertain use in the future now. After all, information may be superseded or use may shift.

What this comes down to is that the choice of technology and quality requirements differ with the goals one tries to achieve. Preservation of information should be distinguished from preservation of the use of that information, and the technologies can be seen as complementary. User requirements and preservation requirements do not always have to be met with the same technology. And an established technology like microfilm still has a role in a policy for collection management that balances different requirements instead of going for a 'one size fits all' solution.

When I read Nicholson Baker's *Double Fold* (Baker, 2001) I was struck by the parallel between the enthusiasm around the large-scale introduction of microfilm several decades ago and the present spread of digitisation. With such high hopes placed in a new technology, don't we run a risk of ignoring what is good about the things that have served us faithfully for years? I like to believe that the old is not simply superseded by the new but that the function of established technology shifts with the arrival of newcomers. Yes, I know, we have lost Morse code and the trusted telegram and punch cards and wax cylinders. And yes, perhaps those in the business of preservation tend to be natural conservatives wishing to keep things as they are – and rightly so! For it does pay to hold on to things that work. Those who bet all their money on the Internet bubble now sit at home and are no doubt happy that there are still books to read!

REFERENCES

1. Baker, N. *Double Fold: Libraries and the Assault on Paper*. Random House, 2001. ISBN**:** 0375504443.
2. Chapman, S., P. Conway & AR. Kenney. *Digital Imaging and Preservation Microfilm: The Future of a Hybrid Approach for the Preservation of Brittle Books*. Washington, DC: Council on Library and Information Resources, 1999. http://www.clir.org/pubs/archives/hybridintro.html
3. Chapman, S.: "Counting the Costs of Digital Preservation: Is Repository Storage Affordable?". *Journal of Digital Information*, May 2003. http://jodi.ecs.soton.ac.uk
4. Conway, P.: *Conversion of Microfilm to Digital Imagery: A Demonstration Project. Performance Report on the Production Conversion Phase of Project Open Book*, Yale University Library, August 1996. See also, Conway, P.: "Yale University Library's Project Open Book: Preliminary Research Findings,' *D-Lib Magazine*, February 1996. http://www.dlib.org/dlib/february96/yale/02conway.html
5. Dale, Robin L. *RLG Guidelines for Microfilming to Support Digitization*. RLG, 2003. http://www.rlg.org/preserv/microsuppl.pdf
6. Kenney, A.R.: "The Cornell Digital to Microfilm Conversion Project: Final Report
7. to NEH". *RLG DigiNews* 1(1997)2. http://www.rlg.org/preserv/diginews/diginews2.html#com

WEB SITES REFERRED TO IN THE TEXT

The European Commission on Preservation and Access (ECPA).
http://www.knaw.nl/ecpa/