

The Impact of Electronic Publishing on the Academic Community

Session 5: Digital libraries and archiving of electronic information

Preservation of research materials: a domain crossing national boundaries

Yola de Lusenet

Royal Netherlands Academy of Arts and Sciences, P.O.Box 19121, 1000 GC Amsterdam, The Netherlands. yola.de.lusenet@bureau.knaw.nl

©Portland Press Ltd., 1997.

[Copyright Information](#)

Thinking about preservation is thinking about access. Preservation may carry connotations of passivity, of storing things safely, putting them out of reach to protect them against wear and tear. For objects in museums and for monuments there may be some truth in such descriptions, but for research materials it is more appropriate to think of preservation and access as one concept with two poles. The organization I represent, the European Commission on Preservation and Access, was established out of concern for the fate of research collections in archives and libraries that are threatened in various ways. Preservation of these collections is the ultimate goal of the Commission's activities, but the name of the Commission reflects the conviction that this should be done with future use firmly in mind. Libraries and archives preserve books and documents so that they may be used by scholars and scientists. To preserve materials that are not used is easy, but pointless. On the other hand, researchers should realize that continued use of research materials is not a matter of course but depends on extensive efforts for their preservation.

Preservation is still all too often understood solely in terms of protection of our cultural heritage. What needs to be preserved is a mass of publications and documentation, the accumulation of past research which forms the basis for present-day inquiries into many different subjects. And it goes beyond that, into finding ways and means to make sure that what is published today can also still be used tomorrow. Preservation is an essential activity for the continuation of all academic research into the future, not just a concern for old and beautiful remnants of the past that are of little actual use to anyone, however valuable they may be from an aesthetic or cultural point of view. Collections in libraries and archives constitute an important part of our cultural

heritage, but they are also the resource materials for all those researchers who need to work with them. Once we recognize the full range of preservation activities and their relationship to access and use, it is easy to see the relevance of preservation for today's output and for all information carriers, from manuscript to optical disk.

Preservation of research materials has always been a central task of libraries and archives. Over the last decades there has been growing concern that much of the printed material stored in the course of time may be lost to future use because of the rapid deterioration of paper that affects a considerable percentage of the publications of the last 150 years. This decay has endogenic causes, in that paper is subject to chemical processes that affect its structure: these are speeded up by environmental factors, such as humidity, temperature and pollution, but ironically it is use that harms paper materials most of all. If paper could be kept out of the hands of users, it would survive much longer, and one of the factors that puts printed research materials at risk is the exponential increase in use over the past 50 years.

It is revealing for the ambiguous position of the user that in the preservation world there exists a concept of 'benign neglect' [1]. This refers to cases where books and manuscripts were left to themselves on the shelves of libraries or archives, half forgotten, not consulted. If left untouched under normal conditions, such documents often turned out to be in remarkably good shape when discovered again after long periods of such 'benign neglect': they survived precisely because they were not used. And although good library management would certainly not want to deny readers access to materials, non-use of paper materials can be very beneficial from a preservation point of view.

The balance between preservation and access is therefore a problematic issue in safeguarding paper materials. With millions of books and documents at risk and limited resources for preservation measures, it is clear that not everything can be kept for posterity and choices will have to be made about what to save first of all. Use can be a guiding principle in selecting materials for active preservation: books that are consulted often should get priority, both because they suffer most and because they are most needed.

The efforts to make heavily used materials available to readers even if their condition makes it preferable to keep them under lock and key are widespread. Digitization of such materials allows institutions to restrict circulation of the originals, thereby extending their life, whereas at the same time maximum access is guaranteed as readers can consult the digitized version. Digitization in this way becomes a tool for indirect preservation: the book in question is kept safely on the shelf and will still be there for future generations. Microfilming can serve the same purpose of providing information in a different format when the original can no longer be used. All such measures are directed at extending the useful life of books and documents in an effort to reconcile the needs of users and the requirements for preservation.

If preservation is understood as primarily serving the demands of users, one can describe the tension between access and preservation perhaps more accurately as a conflict between the needs of present users and those of the future. As with many other resources on our planet, durability becomes a key issue, and careful management of research collections is built on the recognition of the rights of future generations to use them after us, even though we, at this point in time,

cannot presume to know what they will need them for exactly. Present use is a guiding principle in preservation, but it can never be the only motivation for institutional policies. We have to think further than that, for what we do not want today must be kept, just in case someone wants it later. And for archives and libraries, 'later' really means much later, for they are planning in terms of decades or even centuries rather than years.

With the move from the paper to the digital environment, all these relationships have shifted. 'Benign neglect' has become a meaningless concept: in the digital world, neglect can never be benign. The rapid developments in software and hardware ruthlessly bar access to information stored on a disk or tape that is left 'safely' on a shelf for 10 or 15 years. Whether the information is actually still there or has faded over time is not even relevant; predictions about the lifespan of optical and magnetic carriers vary, but the essential problem lies elsewhere, in the incompatibility of carrier and software with the new environment that has emerged meanwhile. All discussions about long-term storage of digital materials emphasize the need for periodical migration of data and stress that policies have to be developed that ensure that institutions undertake this systematically.

When we consider the huge amounts of data involved, the inherent risk in storing data on CD-ROMs, disks and tapes that can actually be put on shelves is obvious. The book that is not consulted for a number of years is still there, its information intact, waiting for an interested reader. The publications that sit on shelves for long periods of time are numerous. A similar fate could easily befall many electronic publications, but in the meantime extensive efforts to preserve them will have to be made all the same; left entirely to themselves, they would become useless very quickly. In spite of the many projects now being undertaken to develop procedures for timely migration, it is very doubtful that procedures can be developed that ensure that all digital materials which are not used will indeed be preserved. The problem is primarily not a technical, but a managerial one, and of such scale that one cannot be optimistic about its outcome.

With non-use being so detrimental to continued access, the key to preservation of digital materials, then, is use. As long as data are accessed at regular intervals, any problems in compatibility will be detected quickly and it will be easier to integrate timely migration in the organizational structures of the institution. In this view, a convincing case can be made for making information accessible online, rather than storing it on off-line carriers. At the moment, online access is the preferred choice for frequently used materials, but it has been suggested that in the same environment electronic publications should be included that are not so often consulted [1]. The argument runs that migration of frequently used materials will take place more or less automatically, as they are in constant demand. In the process of migrating these data to meet new software and hardware requirements, the more esoteric materials integrated into the same system would travel along with the rest, whereas left by themselves on some off-line medium they would not be given priority and might be left until it is too late. This kind of 'symbiosis' could save specialist materials for a limited audience as well as materials for which, for one reason or another, there is temporarily less demand.

In the report for the European Commission by Mackenzie Owen and Van der Walle [2] it is proposed to extend the deposit laws that apply in many European countries to include electronic

materials, so that deposit libraries can keep them safe for the future. Deposit libraries have traditionally had the task of preserving the national production of printed materials for posterity, and by law copies of all publications produced in a country must therefore be deposited there by the producers. Deposit libraries are created specifically for long-term preservation and they have the infrastructure in place to fulfil this task. It is therefore logical to think of them as the best place for long-term storage of digital materials as well.

However, deposit of printed materials is not primarily aimed at making publications accessible: it is a last resort, a kind of paper 'back up' of the nation's publications. A deposit collection is not in all cases accessible to users, and this points to a first problem in relying on the deposit system for preservation of digital materials. If long-term accessibility of digital data requires a dynamic system in which materials are regularly used and in which preservation and access have in fact become synonymous, how would this fit into a deposit system first of all geared at long-term storage only? In a digital context, we should rethink the concept of 'deposit' as meaning use, but it is precisely access and use that are restricted for deposited information --- and at the moment more so for digital than for the paper publications.

So far, the impediment to wide access of electronic publications has mainly been the position publishers have taken to protect their copyright; they have objected to free access to deposited publications, for instance over networks, for the obvious reason that the market for electronic publications would shrink to a minimum if distant access to deposited versions was made possible. At present, agreements differ per country, but basically publishers that agree to deposit materials do so on condition that access is restricted, for instance that only on-site use is allowed. As a deposit collection is basically meant to be an archive of all the publications produced in a country, and publishers are usually obliged to deposit copies free of charge, there is certainly some logic to this viewpoint. However, in the case of electronic publications such separation of storage and use unfortunately may seriously complicate long-term preservation.

Another obstacle to maximum use can be that the national deposit library is not by definition the place where the most complete collection in a specific discipline is held. With one library collecting in a certain field and another acting as deposit library, a situation can be created in which electronic publications are held by the one, fitting into a specialized collection that serves a community of users, whereas in the other the same electronic publications are kept as part of the deposit system, as archival copies. Both institutions will have to migrate the data to keep them accessible. The deposit library will have to do this for materials that are consulted first of all in another library, the one which has the best collection in the field. Is such a duplication of efforts really useful, one wonders. Would it not be much more efficient to preserve information where it is most used?

This may, unfortunately, create another problem when the specialized library does not commit itself to keep the data when they are not used so often any more. Libraries for which the task of 'memory institution' is subordinate to the task of 'information provider' --- not so uncommon, especially in science and technology libraries --- may be little inclined to invest scarce resources into migrating data that are no longer consulted regularly. Perhaps practice and experience in the world of archives can suggest a workable approach here. In public administration records are created and used by various institutions for a certain period of time before being turned over to

another institution, assigned the task of archiving them for the long term. Archives have developed elaborate procedures for appraisal and selection of records at this moment of transferral from one institution to another, which may take place dozens of years after the records were first created, to restrict the heavy burden of long-term preservation to essential data that are most likely to be required in the future too.

A model like this has certain problems of its own, one of the most serious being that at some point materials in all formats are transferred to a central archive that then has to find ways to store them in its own system and keep them accessible. This is why archives are now studying the possibilities for setting standards for the creation of records that should make it easier to transfer and preserve them later. This development means that whereas in the past considerations of preservation played a role at a later stage, they now have to be taken into account right from the start, at the moment of production of information. The fact that a central archiving institution is dependent for the adequate fulfilment of its tasks on the ability of many different institutions to adhere to certain procedures and standards is certainly not without risks and can be a strong argument against decentralization, but on the other hand in the library world, institutions now face similar problems with the materials they receive directly from publishers. It remains to be seen whether these problems would be indeed exacerbated under a system in which a collection of information was transferred from a specialized institution to an archiving institution at some later point in time.

Such a two-step approach, with publications being kept by an institution where they are frequently accessed until such a time that the low level of use makes it preferable to turn them over to an archiving institution, has two important advantages: migration efforts fit in more easily with the management of the special research collections in which they are kept at first and where they are most in demand; and the archiving institution can at a later stage restrict preservation efforts to a selection of materials that has proved to be of lasting value.

In the face of enormous amounts of information being produced and the limited resources available for their preservation, a selection mechanism will no doubt have to be developed anyway. It has long come to be recognized that the idea that everything can be kept is illusory and that selection will have to make it easier concentrate efforts on what is really worth keeping, however difficult it may be to establish criteria for deciding what will have lasting value. In Europe, at a very basic level selection has become institutionalized, as the accepted policy is that every country should preserve its own national production (i.e. books and journals published in the country, by national authors or in the national language), and the system of deposit libraries does exactly that. This policy is rooted in ideas about preservation of a national cultural heritage and is eminently suited to guarantee safekeeping of general publications, literature and books in the national language or about subjects of national importance; in short, all publications that document the cultural developments of a nation. It is a much less logical approach for scholarly and scientific publications which do not primarily have national relevance but are important for academic disciplines that by nature transgress national boundaries. In academic research the organizational principle is by domain, not by country, and it is therefore questionable whether preservation of research materials should be managed at a national level by institutions like deposit libraries without further specialization by subject area.

A good illustration of what national deposit rules lead to is the situation in The Netherlands, a country that happens to house some of the largest multinationals in academic publishing within its borders. The consequence of the practice of national deposit is that all (English-language!) publications of Elsevier Science, the thousands of journals in areas like physics and biochemistry, should be preserved by the national library of the Netherlands, which apart from a deposit library is a library collecting specifically in the humanities. In a paper environment, where preservation is first of all (though certainly not exclusively) a matter of adequate storage and the deposit collection is regarded primarily as an archive, this combination of tasks is less problematic than in an electronic world where access and preservation are so inextricably intertwined as to become almost identical. It seems more efficient to preserve electronic publications that are relevant for an international community of scientists and that could moreover be accessible over networks disregarding national borders, in archival institutions with a responsibility for a specific domain rather than in national archives. The biochemical publications of Elsevier Science are in no way specifically Dutch in nature, only so through their place of publication, which is more of a coincidence than a real characteristic. The information contained in them belongs to the international community of biochemists and should ideally be kept accessible together with similar publications --- whatever their origin --- in a way that best serves the interests of that community. A reason for keeping them in The Netherlands could only be the contribution they make to documentation of cultural developments in the country, to illustrate the role and impact of the publishing industry for instance, but to preserve thousands of publications from a multinational of this size for that reason alone seems a bit too much of a good thing.

The model proposed for long-term preservation of digital materials in the report by Waters and Garrett [3] puts the responsibility on the producers of information. This proposal has been criticized from various sides as being, simply said, too idealistic, in that one cannot rely on publishers keeping their products once it has served its use. This is certainly a valid criticism, but on the other hand the model foresees as a second step the creation of archives with the responsibility of actively acquiring and preserving publications in specific subject areas of which producers are no longer able or willing to take care. What this may amount to is that one does not waste efforts on preserving materials as long as they are preserved by the producers anyway and thereby avoids the conflict between preservation and access that is inherent in the deposit system, in which publishers deposit information whose use is subsequently restricted. The model recognizes that as long as publishers can market and sell their publications they will take good care of them, but once preservation becomes unprofitable and therefore unattractive from their point of view, special archival institutions need to step in to make sure that valuable information is not lost.

An important advantage of this approach is that preservation is organized through a network of institutions committed to keeping materials in a specific subject area, rather than all information that meets a formal criterion, such as being produced in one country. It is an organizational model that is much closer to the realities of academic research as an international, discipline-based activity and that moreover makes it easier to take into account the differences between disciplines and different needs of users. This is a central issue for effective preservation, especially when it is so dependent on use as with digital data, of which the importance is sometimes underestimated in proposals for general approaches. Under the heading of academic

publications very diverse and essentially different kinds of research output are in fact lumped together, and generalizations about requirements for access and use which should motivate preservation efforts are very hard to make. Optimistic scenarios about the digital library of the future often seem to stem from a simplified view of the user community they should serve, as if publications and requirements for use were more or less the same over the various academic disciplines.

It was therefore encouraging that in the opening chapter of this volume, Arnoud de Kemp started off by emphasizing the many differences that exist in the processes for creating publications, the formats used, the kinds of publications, and the requirements of users in the various scholarly and scientific disciplines. As he explained, the wide variety in programs used to produce text, images and data, the many publishers that are involved, ranging from multinationals to small university presses specializing in esoteric work, and linguistic differences, all contribute to a stunning heterogeneity in the kind of material produced. While in some of the sciences researchers have moved to publishing brief articles that are electronically stored as entries in huge databases, at the other end of the spectrum scholars write lengthy monographs in much the same way as a hundred years ago. This variety exists not because some disciplines are more advanced than others, but because all disciplines are different and academics working in them all have different requirements. Traditions survive for good reasons, just as developments occur because they serve a purpose: they are determined by the needs of researchers, and any policy developed for collecting electronic information and keeping it accessible over time should take account of the reality of this variety.

When we look at what is happening now in collection development, we see how institutions working in a specific discipline, libraries, research institutions as well as scientific and scholarly organizations, are constructing databases, acquiring electronic publications and creating gateways to information kept elsewhere for the community of academics they serve. They do this on the basis of a familiarity with the field and a knowledge of the needs of researchers working in that field that are only found in specialized institutions. If anything, it has become more difficult, compared to the paper world, to gauge the relevance of publications for a discipline, given the wide variety of publications and the many organizations involved in their production. To build up a good gateway to all the information scattered over the Net may in the end require more expertise and more insight in the field than to build up paper collections through acquisition of journals and books from reputed publishers.

Again, this may prove to be very different for different subject areas: some disciplines are very specialized, small and highly organized, and so are their procedures for publishing research results, in others publication efforts are much more dispersed over small organizations over the world. In all cases, however, a good policy for creating optimal access and effective preservation, based on a true understanding of the needs of users, requires a familiarity with the field that is characteristically found in specialized institutions and that is at the moment already directing efforts to meet the new demands of the electronic environment. An effective system for long-term preservation of digital information should build on the expertise which is being acquired in this way about the possibilities to use the new media to the best advantage for specific disciplines. In the context of provisions for resources for academic research that is not specifically nationally organized, all this would plead for a domain-based approach to archiving

of digital information that is not limited within the borders of one country, rather than an approach that works through national deposit.

However valid such considerations may be, it is doubtful that they can indeed lead to a supra-national approach in the near future. The strength of the system of national deposit libraries is that these are all well-established, large institutions with a special responsibility for long-term measures which have more-than-average resources to explore options for permanent storage and carry out research into the technical problems involved. Second, the reality in the European context is that preservation of library and archive materials is perceived as falling under the heading of preservation of national cultural heritage and, whatever the limitations of such a view may be, is consequently also funded on a national level. A pragmatic approach to the problem needs to take these realities into account and would therefore almost necessarily have to start by building on to the infrastructure that has been developed over the years for deposit of publications.

Meanwhile, it can be explored how to come to the most effective solution that combines requirements of access and preservation in a dynamic system best geared to the needs of users in different academic disciplines. The developments continuing independently in the building of electronic libraries and digital archives will no doubt at various points catch up with the discussion on long-term preservation and perhaps redirect its course. An initiative that could have a great impact on the present debate is the plan to establish a networked European deposit library that is now being developed by several large national libraries in Europe on the basis of a recommendation in the report of Mackenzie Owen and Van der Walle (p. 95) [2]. This might be the first step to a supra-national approach which in the long term could lead to a system of distributed archiving that offers more scope for a domain-based organization on a European level. For the future of academic research it would in any case be a serious improvement if the preservation of scholarly and scientific information that travels freely over global networks was not confined within national borders. For only if the recognition that academic research is an international activity organized in many different disciplines is translated into a concerted, international effort to preserve research materials can optimal long-term access be realized.

References

1. Maggie (1996) Long-term Management Issues in the Preservation of Electronic Information. In *Multimedia Preservation. Capturing the Rainbow*. Proceedings of the Second National Conference of the National Preservation Office, Brisbane, 28--30 November 1995. National Library of Australia, Canberra
2. Mackenzie Owen, J.S. and Van der Walle, J. (1996) Deposit Collections of Electronic Publications. European Commission DG XIII, Luxembourg
3. Waters, D. and Garrett, J. (1996) Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. Commission on Preservation and Access, Washington, DC

©Portland Press Ltd., 1997.

[Copyright Information](#)