Joint RLG and NPO Preservation Conference

Guidelines for Digital Imaging

Guidance for selecting materials for digitisation

Paul Ayris
Director of Library Services
University College London

1: The context

It is timely for the Information profession to consider the managerial implications of the digitisation process. Within existing academic libraries, which is my own area of interest and expertise, well-developed mechanisms are to be found for selecting print on paper materials. For digitisation to take its place in the Information world as a standard tool in the armoury of professionals, similar arrangements must exist in the electronic arena.

A number of published studies set the need for the selection of materials in the context of the whole digitisation process. A recent JISC/NPO study into long-term archiving concluded that, for preservation, responsibility for drawing up selection guidelines could be assigned to a number of different agencies depending on the type of material under consideration. The basis for selection should be the permanent value of the data or product. The authors maintained that the same criteria used for selecting print materials could be used for electronic publications. The responsibility for developing detailed acquisition policies should lie with the legal deposit libraries or their agents.[(1)](#) In a more recent study, Beagrie and Greenstein also see the selection

process as being firmly embedded in the policy framework regarding digitisation. (2) A recent investigation by the Data Archive at the University of Essex looked at selection in the context of unpublished research materials. (3) The prevalent view of selection to emerge from the Data Archive's survey can be summarised as a series of difficulties:

- How to predict what will be useful in the future
- How to know when data can be acquired for preservation and access if researchers are still working on it
- How to ensure the integrity of, and responsibility for, data which needs constant updating

A thorough review of selection in the digitisation process has recently appeared in the United States, a study which rightly sees selection of materials as deeply embedded in the whole digitisation process. The conclusion deserves to be quoted here:

The process of deciding what to digitize anticipates all the major stages of project implementation. Digital resources depend on the nature and importance of the original source materials, but also on the nature and quality of the digitizing process itself - on how well relevant information is captured from the original, and then on how the digital data are organized, indexed, delivered to users, and maintained over time. Disciplined efforts to address the themes and questions outlined in this essay will help ensure that new digitizing projects fulfill the expectations of libraries, students, and scholars. (4)

At an operational level, a number of institutions have developed local policies for the selection of material for digitisation. Such policies are hard to find, since few are published. In a search of the World-Wide Web, the following policies have come to light. The National Digital Library Program at the Library of Congress has a project planning checklist, which divides the selection process into a series of six interlinked steps. (5) The University of California has well-developed selection criteria for digitisation, which are divided into two main processes containing a series of twenty steps. (6) Columbia University has also developed a set of selection criteria for digital imaging projects, which are divided into three categories: (7)

- Collection development
- Handling and use
- Added value

It is a significant, and probably an accurate, sign of a general lack of activity that no guidelines have been discovered on institutional web servers at universities in the UK. Harvard has adopted an interesting approach by developing a decision-making matrix for selecting materials for digitisation. This matrix arises naturally out of the conditions which obtain in that university. There is a cluster of nine questions with further elaboration of some of the issues involved. The questions can be summarised as:

- Does the material have sufficient intrinsic value to ensure interest in digitisation?

- Will digitisation significantly enhance access or increase use by an identifiable constituency?
- What goals will be met by digitisation?
- Does a product exist that meets identified needs?
- Are rights and permissions for electronic distribution securable?
- Does current technology yield images of sufficient quality to meet stated goals?
- Does technology allow digital capture from a photo intermediate?
- Are costs supportable? Does an institution have sufficient expertise in project management?
- Is the local organisational and technical infrastructure adequate?
- Can the project be re-defined to recast objectives? Can infrastructure needs be addressed?

The approach is interesting, since at any stage in the process the digitiser can answer 'No' and so halt the work. (8)

2: The task

Most libraries have collection development policies for traditional print on paper materials. Such documents form the core of a library's collection management strategy. In a digital world, however, such documents will not in themselves give sufficient help to those who wish to select materials for digitisation. What is needed is some guidance, or a set of guidelines, to further this process. In truth, it can only be guidance rather than guidelines because relatively little is stable in a digital environment. Where, for example, are the costing models which help us study the economic issues in the digitisation process? What is given here, therefore, is guidance in the form of a decision-making matrix which can underpin the selection process for materials which are candidates for digitisation. There also follows a case study in my own institution, where the effects of implementing this matrix are evaluated.

3: Categories

The sorts of questions which need to be addressed in formulating such guidance can be grouped under the following heads:

- Assessment
- Gains
- Standards
- Administrative issues

Assessment

1. What level of support is there amongst target user groups?
2. Is digitisation consistent with local collection development policies?
3. Would digitisation make a contribution to local or national electronic resources?
4. Is there another product which meets these needs?
5. Is this digitisation for preservation or digitisation to enhance access?

Assessment is really the prelude to all other activities in the selection process and should naturally come first. Let us study one or two of these questions in greater depth. There is no point in selecting materials for digitisation if there is no support for using the resource amongst target user groups (no. 1). This tenet is identical to guidance in a conventional collection development policy. No paper material would be purchased by a library if potential use could not be identified amongst library users. Again, take the point about existing products (no. 4). There is no point undertaking digitisation activity locally if there is a product elsewhere which fits the bill. This idea is also to be found in conventional collection development policies, where the needless purchase of duplicated material is to be avoided.

Gains

1. Does digitisation significantly reduce the handling of fragile originals?
2. Material which has been digitised should enhance the academic use of the work by:

   - the creation of finding aids
   - links to bibliographic resources
   - links to online records
   - creation of training materials

3. Navigation should be easy
4. Where collections are split amongst different sites, the result should be a virtual collection which unites disparate material
5. Where originals are damaged, text and images should enrich the academic use of the collections

For digitisation to be a success, there have to be palpable gains in undertaking such work in the first place and this is the theme of the points in this group. Take just one example, namely the creation of training material (no. 2). No-one would expect a user, who could not read Chinese characters, to cope with material in that language and script. The same is true of digital material, where the necessary training materials should accompany any digital image. Use of the text, or images, and links to embedded resources should be covered as well as use of standard software tools, which can be used to study the digitised resources.

Standards

1. Do the standards being used meet national/international standards, yielding images of suitable quality?
2. Will the resources thus digitised be available from the variety of hardware platforms supported by your institution?
3. Is the software used to deliver the materials readily available and easy to use?
4. Does the metadata conform to agreed international standards, e.g. Dublin Core?
5. What are the requirements for archiving in terms of hardware, software and data migration?

Archiving (no. 5) is a big issue, particularly in academic research libraries. Such libraries acquire print-on-paper, secure in the knowledge that such materials will

continue to be available in fifty years time. No such security of thought can be present in the digitisation process. Technology changes too fast, with hardware and software becoming obsolescent very quickly. In the UK at least, libraries are only just beginning to grapple with the problem of electronic archiving and few university institutions currently have systems in place to cope with it.

Administrative issues

1. Is there sufficient finance to meet the costs of digitisation, and does the outcome of the digitisation process meet the requirements of the funding body?
2. Have copyright permissions be obtained and rights issues addressed?
3. Does the institution have sufficient expertise to carry through the project?
4. Does the digitised resource enable you to create a partnership with a commercial provider?
5. Do the benefits of digitisation justify the costs of doing it?

Many of these issues are so obvious as to require no further comment here. The proposed cost-benefit analysis (no. 5) is an interesting area. For it to work, all costs should be taken into account and there should be a comparison with the cost-benefits of acquiring and storing print-on-paper. It is unlikely that many academic libraries have the figures immediately at hand to undertake such work.

4: UK-based case study

What would be the effect today of using this guidance in a UK academic library? For the purposes of this study, I will take my own institution at University College London. During the last ten years, College has doubled in size due to a complex series of mergers with other institutions in London. There are now approximately 15,000 students and 5,000 staff.

Digitised teaching material?

During the same period, transactions at the Issue Desks have increased fivefold. Clearly, the library service is far busier than it was a decade ago. One of the possible solutions to delivering services in this environment is to digitise more teaching material and make it available over the campus-wide network. Well, there is a three-tiered library committee structure in College:

- College Library Committee
- 8 Faculty Library Committees
- Departmental Library Committees

This committee structure is relatively new, but is already working well. As Director of Library Services, I would have to consult all these committees to find out:

1. If electronic resources are embedded in the curriculum
2. Whether the electronic delivery of material to support taught-course provision would be acceptable

Co-ordination of work within Library Services

If the guidance on action outlined above were accepted, there would certainly be a problem of co-ordination since many of the tasks embedded in the digitisation process cut across traditional library boundaries. The following categories of staff would be affected:

- 23 Subject Librarians
- Staff in the Central Cataloguing Unit and Periodicals Department
- Staff in Planning & Resources
- Staff in IT Services
- Staff in the Subject Support Unit

What would this mean in practice? Subject Librarians are responsible for academic liaison and would clearly be involved in the process of selecting material, liaising with academics and for providing training in the use of the final resource. The Subject Support Unit in UCL is directly responsible for taught-course support and would be responsible for dealing with the necessary rights issues. Cataloguing and Periodicals would be involved in metadata and licence issues. Planning & Resources would be involved in financial matters, ensuring that all costs could be met. IT Services would be involved in technical issues, ensuring that the resource could be networked centrally and that it could be made available on all necessary hardware platforms. How can all this work be co-ordinated and who is responsible for ensuring that all steps in the chain of actions have been taken?

Value-added features

One of the gains to be made from digitising original materials is the creation of value-added features, which enhance access to damaged originals. An instance of this can be found in Cambridge amongst the digitised fragments of the Taylor-Schechter collection. The Taylor-Schechter Genizah comprises tens of thousands of fragments, dating from the early Middle Ages, which were found in Cairo. There are religious texts, but also an enormous quantity of other material which sheds shafts of penetrating light on the Mediterranean world of Judaism at this time. Many of the fragments are damaged and digitisation has given the Taylor-Schechter Unit the ability to move fragments around on the screen, avoiding damaging contact with the originals, in order to try to re-create the original form of the manuscripts. The work of the Taylor-Schechter Unit can be found at http://www.lib.cam.ac.uk/Taylor-Schechter/.

Long-term archiving

One of the issues discussed above is the need to ensure long-term archiving of the digitised materials. When libraries buy books and periodicals in conventional format, they do so in the expectation that the originals will still be accessible in fifty years time. The same model needs to apply to electronic libraries, at least as far as research libraries are concerned. In many universities, however, libraries have no control over the technical process of digital archiving. This falls under the remit of the central Computing Service, who will take an institution-wide strategic decision on how best to meet this need. In addition, many institutions in the UK are only just beginning to

grapple with this problem. CURL, the Consortium of University Research Libraries, is undertaking a study of digital archiving as it affects research-based libraries. They are looking at CD-ROMs, websites, dynamic e-mail discussion lists and other sorts of resources to try and identify models for archiving each type of resource.

5: Decision-making matrix

Using the twenty issues/questions outlined in paragraph 3, it is possible to arrive at a basic decision-making matrix to inform the selection process in a digital arena. This matrix is given in full at the end of this article. The matrix is, of course, not in itself a collection development policy. Rather it is a decision-making tool which informs the process of selection, which itself forms part of a statement about collecting strengths. It would be perfectly possible for the decision-making matrix to be applied to the same resource in two different libraries and for different decisions to be made as to whether the resource should be digitised. This would be quite in order, as a similar outcome could be expected when selecting conventional paper materials for purchase. It might also be possible for individual libraries to assign different weights, or values, to each of the questions in the matrix in order to ensure that they reflect local conditions as accurately as possible. What is important, however, is that libraries validate their selection procedures for digitisation with reference to external, internationally-agreed, criteria. Digital libraries are best formed through national and international collaborations, especially since collection development is increasingly a collaborative activity between libraries. In addition, external funding is most likely to be available where proposed digitisation programmes meet agreed criteria in terms of preparation, selection and image capture. It is important that the selection process does not become isolated from national and international debate. The decision-making matrix is offered, therefore, as an aid to decision-making, and as a tool which can be applied in a variety of libraries and archives, but which is sensitive to the context of all.

6: Conclusion

This paper has identified a number of recent studies which have considered the role of selection in the process of digitisation. It is clear that some recent essays have taken the question very seriously and are treating it as a core part of the management process in a digitisation project. A number of operational guidelines also exist on websites, although little such activity is to be found in the UK. This study suggests a decision-making matrix of twenty questions grouped around four issues to aid the selection process:

- Assessment
- Gains
- Standards
- Administrative issues

Methods for implementing the matrix are also suggested. For co-operative digitisation projects to succeed, and for funding bodies to consider substantial financial investment in such projects, it is desirable that all digitisation projects validate their activity by reference to agreed international criteria. A decision-making matrix is not in itself a collection development policy, but it does form a contribution to the

creation of such a policy. A case study is then offered for the proposed implementation of the decision-making matrix in a typical research library in the UK. It can be seen that many of the decisions required by implementing the matrix cut across traditional administrative boundaries in libraries. In addition, the demands of long-term archiving place an almost insupportable burden of responsibility on local institutions. In the UK, few academic institutions are equipped to implement the decision-making matrix without changes to their internal structures and decision-making processes. The future is exciting, but it poses some interesting challenges for anybody wishing to invest seriously in digitisation to deliver materials to their users.

Bibliography

1. D. Haynes, D. Streatfield, T. Jowett and M. Blake, Responsibility for digital archiving and long term access to digital data, Elib study P1 (London, 1997), p. 44.
2. N. Beagrie and D. Greenstein, A strategic policy framework for creating and preserving digital collections, British Library Research and Innovation report no. 107 (London, 1998), section 4.5.
3. The Data Archive, University of Essex, An investigation into the digital preservation needs of universities and research funders: the future of unpublished research materials, British Library Research and Innovation report no. 109 (London 1998), pp. 10-11.
4. D. Hazen, J. Horrell, J. Merrill-Oldham, Selecting research collections for digitization (Washington, 1998), p. 18.
5. NDLP, Project planning checklist, available at http://lcweb2.loc.gov/ammem/prjplan.html
6. University of California, Selection criteria for digitisation, available at http://www.library.ucsb.edu/ucpag/digselec.html
7. Columbia University, Selection criteria for digital imaging projects, available at http://www.columbia.edu/cu/libraries/digital/criteria.htm
8. Hazen, Horrell and Merrill-Oldham, Selecting research collections, p. [19].

Appendix

Decision-making matrix to support selection activities in the digitisation process

	Question 1	Question 2	Question 3	Question 4	Question 5
Assessment	Is there user support?	What are local collection development policies?	Does this form a national or international contribution?	Does a similar product already exist elsewhere?	Is this conservation or preservation?
Gains	Does digitisation reduce wear on the originals or open up access?	Is the intellectual content of the work enhanced?	Is navigation easy?	Are disparate collections unified?	Is use of the damaged original material enriched?
Standards	Have suitable standards been followed?	Are the originals available from a variety of hardware platforms?	Is the software available and easy to use?	Does the metadata conform to agreed standards?	What are the archiving requirements?
Administrative Issues	Do you have enough money?	Have copyright and rights issues been secured?	Does your institution have enough expertise?	Is there a partnership with a commercial provider?	Do the benefits justify the costs?