Digitisation in UK libraries and archives: is a national strategy needed?

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- This pre-print has been submitted to the journal, D-Lib Magazine [© Corporation for National Research Initiatives].

Metadata Record: [https://dspace.lboro.ac.uk/2134/471](https://dspace.lboro.ac.uk/2134/471)

Version: Submitted for publication

Please cite the published version.
Digitisation in UK Research Libraries and Archives: is a national strategy needed?

Barbara Bültmann
Rachel Hardy
Adrienne Muir
Clara Wictor

1 Introduction

In August 2004, the Joint Information Systems Committee and the Consortium of Research Libraries Digital Content Creation & Curation Task Force issued an invitation to tender for a study of the current provision of digitised collections for researchers in the UK higher education sector. The objectives of this study included identifying what digitised content exists, user needs and demand, UK research library digitisation activities and plans, funding sources and models and standards. The work also included consideration of the need for a national strategy for digitisation and recommendations for action.

JISC and CURL commissioned a team of researchers from the Department of Information Science at Loughborough University to carry out this survey between 1 November 2004 and 7 March 2005.

2 Study Methods

The study involved desk research, a Web-based questionnaire of research libraries and interviews with key informants. A Web-based survey of UK research libraries and archives was developed in collaboration with JISC and CURL. This was deemed the most appropriate given the time and resource
constraints of the study. The questionnaire included questions on past, current and future digitisation projects, reasons for and against digitisation and experience of in-house and outsourced digitisation and collaborative efforts. Fifty-one replies were received from 47 institutions, including the major research institutions. Unfortunately not all respondents were able to provide detailed lists of material to be digitised, probably because they did not have this information readily to hand.

Thirty-six in-depth interviews were conducted with representatives of different research disciplines and stakeholder groups. Two other organisations responded to emailed questions. These included:

- Members of JISC and CURL
- Representatives of institutions with digitisation experience, including the three UK national libraries
- The National Archives of the UK and Scotland
- Publishers
- Support services
- Funders

A core set of questions was developed for each stakeholder group, but interview schedules were tailored according to the roles, experience and expertise of interviewees. The main aim of the interviews was to learn from the knowledge, experience and expertise of the interviewees and build on the data gathered through the desk research and survey.

This article summarises the main findings and recommendations of the study.

3 Main Findings

The study found an impressive amount of digitised material in the sector and there has been considerable expenditure of UK public funds in the creation of
digital content, amounting to some £130 million in the last ten years. However, funding of the creation of digital content in the UK has been piecemeal (as in some other countries) and completely uncoordinated. In the UK, the National Audit Office has highlighted issues at national level, including risk of duplication, use of diverse standards and the importance of (and opportunities for) collaboration. This study has also found that there is no national oversight for digitisation. It is clear from discussions with various players that there is a need for coordination, but no agreement on how this should be implemented.

3.1 Digitised Content in UK Research Libraries and Archives

Manuscripts and images are the most frequently digitised type of material in libraries and archives, although other types of material, including artefacts, have also been digitised. Publishers are producing some substantial resources by digitising journal back files and monographs. The bulk of material digitised is most relevant to the arts and humanities and social science fields. However, medical image material is also an area where there has been much activity. Many UK digitisation projects have involved selective creation of resources focusing on particular themes or specialist topics, for example prominent individuals, historic events or special “treasures”. There have also been some more comprehensive projects. Some of these have focused on particular genres or collections of material, for example medieval manuscripts, newspapers, official publications or census data. These activities have produced, or will produce, substantial resources that potentially will have large audiences. There is also some evidence of collaboration between libraries, archives and museums in digitisation. Some of this collaboration has been by geographic region, for example in Scotland or across Wales. There has also been UK participation in international activities to bring together dispersed collections, such as the International Dunhuang Project.
3.2 Reasons for digitising

There are various reasons for digitisation, including improving access generally, and particularly to unique and rare material. Selection criteria identified by this study include relevance to institutional mission, uniqueness or rarity and existence of coherent collections. Demand is another selection criterion. Respondents from libraries and archives mentioned that they had digitised collections to which access is requested more frequently than others and would therefore benefit from being in digital format. Publishers mentioned the market research they carry out for their digitisation projects. Although a small number of respondents said they responded to the requirements of funders, this selection criterion was not a frequent as expected.

3.3 Researcher Needs and Demand for Digitised Resources

While demand is a criterion for selection for digitisation, no overview of user needs and demands for digitised content in the UK exists. There is little literature on existing researcher needs and demand for digitised resources. In the recent past, the Research Support Libraries Group carried out a study that with some relevance to the subject discussed here. The British Academy has just completed a major study in the arts and humanities and social sciences, that should add to the picture of what is needed. However, there is still a gap in the sciences and other areas that needs to be filled. The interviews carried out for this study were not able to provide a comprehensive overview in this area. However, there was a suggestion that the most important resources for scientists are journals and datasets, while other sorts of material are little used. While much current material of interest is already available in digital form, there is much activity in the digitisation of journal back runs. Publishers, or bodies such as JSTOR, carry out a lot of this activity. Some of this activity is subsidised by funding bodies: funding bodies either pay for the cost of digitisation or for publishers to make material they have digitised available for free or at a reduced cost.
Some libraries and archives have prioritised material they want to digitise, but few were able to provide detailed prioritised listings. Much of the information given by survey respondents related to manuscript and image material of particular relevance to the arts, humanities and social science fields. Medieval manuscript collections were mentioned, but more modern archival material was also nominated. More than one respondent also listed medical images. The collections suggested for digitisation are considered by their owners to be rare, vulnerable or valuable in some way. The nature of these collections needs to be investigated further by bodies such as CURL, JISC and the newly established UK Research Library Network.

However, there is a question of whether material should be digitised just because it is rare or vulnerable, or whether there should be a demonstrable need. While it would make sense for these bodies to take forward digitisation of material held in libraries, this activity could perhaps wait until a clear overview of research needs is available. At this point a more comprehensive gap analysis could be conducted. While librarians and archivists are dedicated to supporting researchers and are often good judges of what content will be useful, taking the long-term into account, there is a need to gather more direct input from researchers.

Another issue suggested through this study is that use of digitised material, particularly in the arts and humanities field, is less than optimal because of a lack of researcher awareness and perhaps even resistance in some cases. Another issue is that there is a need to enhance raw digitised content with enhanced functionality such as text-searching capabilities and facilities for manipulation of digitised content. The Arts and Humanities Research Board (now Council) has shown an interest in not only the creation of digitised resources for research, but also the use of information and communication technologies in the arts and humanities in general through its strategic ICT in Arts and Humanities Research programme. The AHRC ICT Strategy Projects Scheme apparently considers the need for and development of appropriate tools for researchers.
A possible way forward on finding out more about researcher needs could be through the UK research funding councils. The research councils have various structures in place, such as research programme managers and panels that could be consulted or which could conduct consultations with the research community in different fields. The councils with ICT programmes could perhaps use this vehicle. The data gathered could be analysed at the level of individual research councils and possibly aggregated at national level. A more comprehensive survey of the views of subject associations, academies and royal societies than the limited number of interviews carried out for this study could also provide more detailed feedback on researcher needs. An alternative to this approach would be user needs surveys carried out by research libraries, perhaps through the Research Libraries Network. This should be a more focused study than the exercise carried out for the Research Support Libraries Group. Researchers may well be more motivated to respond to research bodies than the RLN, but the library community can help to encourage RLN awareness and thus enhance its impact.

The findings of these studies could inform policies and strategies of the research councils and be shared between JISC and CURL and/or the Research Libraries Network so that the response to the findings can be co-ordinated. The initial exercise will necessarily be large-scale, but knowledge on needs and demand could be updated periodically. This could be linked to trigger events, for example periodic strategic reviews of research and/or content creation programmes.

### 3.4 Barriers to Digitisation

#### 3.4.1 Funding Issues

For some projects/organisations, the selection for digitisation material was linked with funding opportunities, cost and resource requirements. For certain material e.g., fragile, rare or unique, it was easier to apply for, and be
successful in funding applications. Lack of funding, as well as a lack of expertise, is a barrier to digitisation. Digitisation is funded on a project-by-project basis by a number of different funding bodies, large and small. Major funders have included the Andrew W. Mellon Foundation, the New Opportunities Fund (a UK National Lottery fund) and JISC. While the JISC in particular has funded projects, often in cooperation with other bodies, that aim to provide more comprehensive resources with potentially large audiences, there is a lack of an overarching strategy to co-ordinate digitisation activities in the UK and no national oversight for digitisation. With a better overview of researcher priorities, funders could develop strategic programmes and/or issue targeted calls for proposals for projects to meet specific gaps.

There is a hint from the study that organisations planning to digitise have to spend a great deal of time identifying and exploring funding opportunities. It would seem sensible to have a more co-ordinated approach to the identification of funding opportunities. Existing support bodies in the UK, such as the Technical Advisory Service for Images, already identify potential funding bodies, but the possibility of some sort of portal that provides a “one-stop shop” for funding information could be developed and maintained. This should be international in scope and could provide up-to-date information on funding bodies with links to their Web sites and documentation. It could also provide information on new programmes and calls for proposals. This portal could be useful to libraries and archives (in all sectors) and also researchers who wish to start digitisation projects.

### 3.4.2 Collection Management Issues

Although not the most frequent response, some survey respondents did identify copyright as a factor in selection of material to digitisation. Use of legal experts was also mentioned as a source of expertise consulted by would-be digitisers. There is a specialised copyright clearance (and digitisation) service for the sector in existence in the UK (HERON), but our findings suggests it is not well used. It is not clear why this is the case, unless
digitisers are not aware of its existence. There are now also digitisation services available that could be used by digitisers. So digitisation activities do not necessarily need to disrupt collection management processes, nor does digitisation need to be integrated into the collection management function. It can be outsourced, indeed digitisation was outsourced by some of our survey respondents, although selection and quality control may have to remain internal activities. Respondents did not comment on how resource intensive these activities are. Given that much digitisation seems to involve at least some external funding and obtaining this funding requires the preparation of funding bids, this process may be disruptive of library operations.

Respondents to the questionnaire survey also seemed concerned about the long-term management of digitised resources, both in terms of funding and expertise. Research libraries planning digitisation projects need to take this into account and plan for it. Funding bodies are now beginning to expect this from proposals. The big question is how it will be funded and whether it is appropriate for funding bodies to provide for on-going maintenance or whether it is the responsibility of digitisers. Digitisers need guidance on long-term management and preservation. They need to be aware of the many sources of guidance exist and which support services can assist them. The UK Digital Preservation Coalition should continue its work on raising awareness and could consider the provision of more case studies from its members and international contacts. The newly established UK Digital Curation Centre should also be able to help here, although its focus will be born digital material. Funding bodies (if they do not do so already) and recipients of funding should consider the use of existing data archives to facilitate safe storage and preservation of digitised resources when planning and funding digitisation projects. Several digital archives already exist in the sector, so libraries do not necessarily have to develop all the systems and infrastructure to store and manage material in the long-term or have to find on-going resources to support these activities.
3.5 Cooperation and Coordination in Digitisation

3.5.1 Funding for Digitisation

There is some, but limited, evidence of cooperation between funding bodies and some are only beginning to think in a strategic way about funding digitised resource creation. It became clear during the course of the study that co-ordination is needed. We tentatively suggest that any “national strategy” has to be formulated at a very high level and centralised implementation may not be feasible. It is probably not realistic to expect the various UK public sector funding bodies, never mind other independent and international funders, to develop a unified strategy for funding digitisation in the UK. However, it may be possible to improve co-ordination. Funders could work co-operatively, through existing fora, such as Research Councils UK, or through a new forum including the main funding bodies, to co-ordinate activities. It should certainly be possible for the UK public sector funding bodies to do this. The JISC and the research and funding councils should be able to work together. The JISC could act as the link to the research libraries, through CURL and/or the Research Libraries Network. The Museums Libraries and Archives Council advises the UK government and could be the link to the lottery funding bodies.

3.5.2 Cooperative Digitisation

There is cooperative activity between research libraries and archives and publishers and other commercial organisations. Some of this activity is subsidised by funding bodies such as JISC, Andrew W. Mellon and the Wellcome Trust. While libraries are interested in cooperative digitisation and working with commercial partners, there is some concern at the price of some cooperative, but commercially produced, digitised resources. There are always costs involved in digitisation and someone has to pay. There are various business models for the provision of digitised content that vary in
terms of access conditions and pricing models, depending on whether the
digitising organisation is for or non-profit, and the type of material digitised.
The thing to avoid is the need for the research community to pay very large
sums of money to access digitised research material, particularly if that
material has originated in the research library and archive sector. Commercial
publishers, understandably, need to make an acceptable return on any
investment they make.

Commercial publishers are cautious about subsidised and open access
models. The Google Print initiative is currently an unknown quantity, but could
have a huge impact on business models and research library interest in
digitisation. Publishers who participated in this study are concerned about the
implications of Google Print for future commercial digitisation activities, while
libraries are cautiously hopeful. The Google initiative has the potential not only
to facilitate the digitisation of library materials, but for the existence of the
digitised material to become easily discoverable through Google services. As
mentioned by interviewees, the Google initiative will only be useful if material
is digitised to an acceptable standard and if appropriate metadata is created
for digitised material. If this is the case, and Google remains committed to the
project, this initiative may well prove to be a significant boost for the
digitisation of content. Whether this will be systematic digitisation of content to
meet needs or cherry picking of significant collections is another matter.
CURL and/or the RLN could explore the possibility to taking a nationally co-
ordinated approach to the inclusion of UK research libraries in this initiative in
future.

While all of the business models explored in this study had disadvantages as
well as advantages, it is clear from the various developments that there is
increasing scope for public-private partnerships in digitisation.
3.5.3 Standards, Support and Guidance

Librarians have looked for and adhered to standards in digitisation and JISC and other UK and international bodies have had an important role to play in this. This study has shown that whilst individual projects do things a little differently and that standards and file formats depend on materials digitised and purposes, there is a core set of standards and formats used by many projects. As far as metadata is concerned, library-based projects are mostly using some form of Dublin Core or MARC and using XML and METS encoding for metadata. Archives use the EAD and ISAD(G) schemas for records and finding tools to meet their own needs. Library of Congress Subject Headings are used for subject access in the library sector. There seems to be less standardisation amongst publishers and digitisation services.

Long-term maintenance and preservation are issues of concern to digitisers and funders. Support and guidance on digitisation is sought from a variety of sources. Internal sources of technical, legal and collection management advice are sought, as are external sources. There are many different support services available in the UK for digitising organisations. At their own admission, there is a degree of overlap between these services. Now that digitisation is becoming more established, the time may be ripe to review the services available in order to identify any areas of overlap. It may be useful to have a single point of access to guidance and advice on different aspects of digitisation, including technical, legal and management guidelines and case studies. Again, this should have an international focus to include the expertise or organisations such as Cornell University Library and the Library of Congress. The advice may be provided by different services, but the users would have one access route.
3.5.4 Discovery of Digitised Material

There is an issue surrounding the discovery of the existence of digitised material, so that duplication of effort is avoided and researchers can find material of relevance. The results of this study indicate some issues in the creation of metadata records for digitised material. Only two thirds of survey respondents said there were records for all material digitised and some respondents mentioned that a lack of bibliographic records was an issue. It seems that, in some cases, records do not exist for the originals and metadata creation is a higher priority than digitisation. Metadata creation is an expensive part of the digitisation process and this activity is one that could have a negative effect on existing library operations. It therefore seems sensible that metadata creation is costed into funding bids and that funding bodies be prepared to fund metadata creation. It seems pointless to digitise without providing the means to retrieve digitised resources. It would also be unfortunate if digitisation of useful resources is delayed or does not take place because of a lack of metadata. Automation of metadata creation and re-use of existing metadata records would also ease this situation, this is an area currently under investigation.

The issue of lack of awareness and resistance to use of resources available on the part of UK researchers needs to be addressed, otherwise large investments in digitising material will be wasted. JISC is already working on this with research councils through its ICT awareness programmes. While digitised collections are likely to be included in institutional catalogues and Web sites, information on digitised resources should also be covered in the search tools used by researchers, including the Research Discovery Network. Some resources already are, but coverage needs to become more comprehensive. Our survey found little evidence of OAI-PMH compliance.

A comprehensive listing of existing digitised resources could facilitate the analysis of gaps in the provision of digitised content. The creation of new digitised resources to meet identified needs could also be facilitated by a list,
not only of what has already been digitised, but also of what is in the process of being digitised. One of the deliverables of this study was a list of digitised resources available to the UK research community. The study included a comprehensive search for digitised resources which was complicated and may well have missed important resources. There is a need for a better mechanism for identifying relevant projects and collections. There is a precedent here in preservation microfilming. The Mellon Microfilming involved cataloguing material microfilmed and submitting records to various national and international registers.

Registers and catalogues for digitised material already exist. The systematic submission of information on digitisation projects and material digitised to a national and perhaps international register should be investigated. A UK Register of Digital Surrogates, similar to the National Register of Archives, could facilitate greater collaboration and cooperation. As the register develops, gaps in provision will become increasingly clear. The register could also help in the identification of relevant projects and collections. The appropriateness of existing registers, for example the UK register of preservation surrogates and the OCLC/Digital Library Federation registers should be investigated, as should the nature of the information to be submitted and the best methods for submission. It may be necessary to modify existing registers to allow for information on projects and digitised resources, so the registers in other countries mentioned in this report should be examined as models. Any system would need to be simple and inexpensive to contribute to, in order to maximise participation.

Digitising organisations may well need to be motivated to submit information on projects and digitised material. This may be difficult in the private sector, although publishers may find benefits in a wider awareness of their digitised products and services. There are precedents for the submission of records to registers by commercial publishers (ProQuest). Funding bodies could stipulate that recipients of grants should submit records as a condition of funding. How information could be submitted retrospectively is an issue that needs to be explored.
3.6 Conclusion: A National Framework for Digitisation

It is clear from discussions with various players that there is a need for coordination in digitisation activities within the UK, but no agreement on how this should be implemented. The findings of this study indicate that any future national approach would have to be a co-ordinated and distributed, rather than centralised, one. A UK-wide strategy would assist in filling gaps in provision, cut across the efforts of individual funders and digitising organisations, reduce overlaps between support services and assist in the provision, take up and use of open access resources. A UK-wide approach would assist in overcoming institutional issues, such as successful project management being impeded by costs, varying file and metadata formats and preservation problems. A crucial aspect of any national strategy is that it should reflect researchers’ priorities.

Acknowledgements

The authors wish to thank everyone who participated in the study and supported us in this work. The views expressed here are those of the authors and do not necessarily reflect those of the funders, JISC and CURL.