

---

*Barbara Bültmann is Librarian at the German Historical Institute, London.*

*Rachel Hardy is a Research Associate in the Research School of Informatics, Holywell Park, Loughborough University.*

*Adrienne Muir is a lecturer in the Department of Information Science at Loughborough University.*

*Clara Wictor works in the Bibliothèque Nationale de Luxembourg.*

# Digitized content in the UK research library and archives sector

BARBARA BÜLTMANN, RACHEL HARDY, ADRIENNE MUIR\* and CLARA WICTOR

A study of the current provision of digitized collections for researchers in the UK higher education sector was carried out through desk research, a Web-based questionnaire of research libraries and interviews. The study identified a great deal of digitized material in the sector and there has been considerable expenditure of UK public funds in the creation of digital material in the last ten years. However, funding of digitization has been piecemeal and uncoordinated. It is clear that there is a need for coordination, but no agreement on how it should be implemented. Any future national approach would have to be a coordinated and distributed, rather than centralized, one.

**KEYWORDS:** digitization; funding; National Strategy; research libraries

---

## 1. INTRODUCTION

In the UK, it became clear in the 1990s that digital information would play a major role in higher education (Whitelaw and Joy, 2000). There has been considerable investment in a national digital network for the UK tertiary education and research sector. Since then, the UK library world has made significant advances in the development of digital content services. Support services tackling issues such as access to and the preservation and maintenance of digital resources have been established. Examples include the UK Data Archive,<sup>1</sup> the Arts and Humanities Data Service (AHDS),<sup>2</sup> the Higher Education Digitisation Service (HEDS)<sup>3</sup> and the Resource Discovery Network (RDN).<sup>4</sup>

Many knowledge institutions worldwide are digitizing their collections. Libraries and archives are digitizing books, manuscripts, images and other types of material on the basis that 'one is convinced of the continuing value of such resources for learning, teaching, research, scholarship, documentation, and public accountability' (Kenney and Rieger, 2000: 1). The commercial sector is also involved in digitization with publishers digitizing their own material or material held elsewhere. Examples of initiatives include Eighteenth Century Collections Online and Early English Books Online. Even Google, through its Google Print programme, is involved in digitizing research library collections. Generally digitization serves one (or more) of three purposes: enhanced access to physical information artefacts, preservation of original artefacts through the creation and provision of access to surrogates, and commercial exploitation of information assets.

The UK Government has provided large sums of money to initiate, maintain and support information and communication technology (ICT)

\*To whom correspondence should be addressed.

---

Address: Department of Information Science  
Loughborough University  
Loughborough  
Leicestershire LE11 3TU  
UK  
Tel: +44 (0) 1509 223064  
Fax: +44 (0) 1509 223053  
E-mail: a.muir@lboro.ac.uk

---

Copyright © 2006 B. Bültmann, R. Hardy, A. Muir and C. Wictor  
(London, Thousand Oaks, CA and New Delhi)  
Vol 38 (2): 105–122  
[DOI: 10.1177/0961000606063893]



innovations for the benefit of the research community. Where a national dimension exists, the HE Funding Councils have set up joint subcommittees to deal with particular issues. These include the Joint Information Systems Committee (JISC), which is committed to continuing its central role in providing a world-class infrastructure and promoting innovation through development programmes for the community. Strategic initiatives include the Information Environment, which aims to provide a platform for the provision of access to digital content for learning, teaching and research. Another government-funded activity is the Core e-Science Programme, which aims to enable e-science through ICT solutions and there is also the Arts and Humanities Research Council's ICT Strategy Projects Scheme.

Frameworks are emerging to support collaborative research and development of digital content, including national and multinational strategies for the digitization of the cultural heritage. The European Commission report *Coordinating Digitisation in Europe* (European Commission, 2002) gives an overview of digitization funding, collaboration and strategic initiatives in all EU countries. UNESCO is maintaining a register of some significant digitization efforts worldwide.<sup>5</sup> While there has been coordination in the provision of the infrastructure and born digital content in the UK, there is currently no UK-wide digitization strategy. The digitization programmes that have been initiated in the UK have been funded by a number of different bodies, such as JISC and various UK lottery-funding bodies. A recent survey<sup>6</sup> showed that the majority of digitization projects are small scale and carried out in isolation. This survey showed that projects have used a variety of standards and formats and there has been some duplication in the selection of material to be converted. Digitization strategies vary in terms of rationale for digitization, the aims of projects and the selection criteria used. A recent report has highlighted issues related to digitization at the national level, including risks of duplication, use of diverse standards and importance and opportunities for collaboration (National Audit Office, 2004). However, there are some examples of successful collaborative digitization projects involving UK participants, including the Scottish Cultural Resource Access Network (SCRAN <http://www.scran.ac.uk>), and the International Dunhuang project (<http://idp.bl.uk/>).

Given the fragmented nature of UK digitization efforts until now, this is an opportune moment to stand back and review the situation, with the aim of assessing the needs of researchers, how well they are being met and how provision of digitized collections should be managed in the future. In August 2004, the JISC and CURL Digital Content Creation & Curation Task Force issued an invitation to tender for a study of the current provision of digitized collections for researchers in the UK higher education sector. The study had a number of objectives:

- Produce a high level survey of digitized material, both already available and in the process of being created, held in UK research collections across all disciplines.
- Survey demand for digitized material and identify gaps in existing provision.
- Develop a mechanism for identifying future digitization priorities.
- Review funding structures and opportunities and assess possible ways of funding priority areas.
- Recommend standards and formats for future digitization projects.
- Provide an outline action plan for a national digitization strategy for the UK research community.

The Joint Information Systems Committee and Consortium of Research Libraries (CURL) commissioned a team of researchers from the Department of Information Science at Loughborough University to carry out this study between November 2004 and March 2005.

## 2. STUDY METHODS

The objectives of the study were addressed through desk research, a Web-based questionnaire of research libraries and interviews with key informants. One of the main outcomes of the desk research was a list of digitized resources available to UK-based researchers. The desk research revealed gaps in publicly available information on availability of resources, details of planned projects and wish lists of institutions. Therefore the desk research was supplemented by a questionnaire survey of UK research libraries and archives focusing on past, current and future digitization projects, reasons for and against digitization and experience of in-house and outsourced digitization and collaborative efforts. The questionnaire was hosted on the Loughborough University website and sent to three respondents as a pilot study. A point was raised regarding the inconvenience of providing detailed information and this proved to be a telling one for the identification of material to be digitized in the future. Many survey respondents were not able to provide the level of detail hoped for. A Web survey approach was deemed the most appropriate given time and resource constraints. Ideally, the questionnaire would have been sent directly to individuals that had been identified as having the requisite experience and expertise. The Web-based approach may also have deterred respondents from seeking out detailed information on resources created and funding. However, the project team's previous experience of questionnaire surveys covering similar topics suggested that respondents would find this request time consuming and difficult to collate whatever the questionnaire format (Astle and Muir, 2002; Muir, 2004). Fifty-one replies were received from 47 institutions, which included the major research libraries and archives in the UK. The questionnaire also reached the

Library of the Society of Antiquaries in London and the Archaeological Data Service (ADS).

Thirty-six in-depth interviews were conducted with representatives of different research disciplines and stakeholder groups, including:

- members of JISC and CURL
- representatives of institutions with digitization experience
- the three UK national libraries and The National Archives of the UK and Scotland
- publishers
- support services
- scholarly societies

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. Scholarly societies were targeted in an attempt to gain some insight into the needs of researchers.

This article reports the findings of the primary research carried out for this study. However, where appropriate, findings from the desk research are used to clarify or amplify the primary research results. The article focuses on digitization activities in the UK, including collaborative efforts. Funding sources and future plans in UK research libraries and archives are reported as are the views of digitizers, support services, publishers and funders on the future of digitization in the UK. The possibility of developing a national strategy for digitization of research material in the UK is considered. Finally, conclusions and recommendations for future action toward such a strategy are offered. It is important to note that the creation of digitized content and the management of and provision of access to digital content in general is a fast moving area. New digitization efforts have been announced since the research was carried out and new projects under programmes such as JISC's Digital Repositories programme have begun. The results reported here represent a snapshot of activities at the time the research was carried out.

### 3. DIGITIZATION ACTIVITIES IN UK RESEARCH LIBRARIES AND ARCHIVES

Survey respondents were asked about their current and past digitization activities. The results can be seen in Figure 1.

All 51 respondents answered this question; two-thirds have engaged in digitization activities. Seven respondents are at the planning stage of their first project. Three respondents did not consider the amount of digitization done so far to be large enough to tick 'yes', and the remit of the ADS does not include digitization.

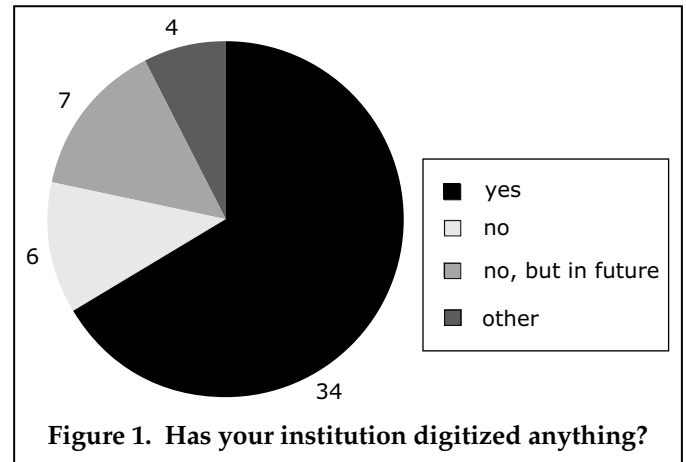


Figure 1. Has your institution digitized anything?

#### 3.1 Reasons for digitizing

All of the 34 institutions with digitization experience gave reasons for doing this. Respondents were able to select multiple responses and were asked to rank their responses. See Figure 2 for results.

Improved access was selected most frequently and ranked most highly, reduced handling comes second whereas building 'virtual' collections was seen as less important. Frequent 'other' responses were: to showcase collections, support (distance) learning, teaching and research.

#### 3.2 Reasons for not digitizing

Eight of the 13 institutions that had not been involved in digitization gave reasons for this (see Figure 3). Three digitizers also answered this question. Respondents were

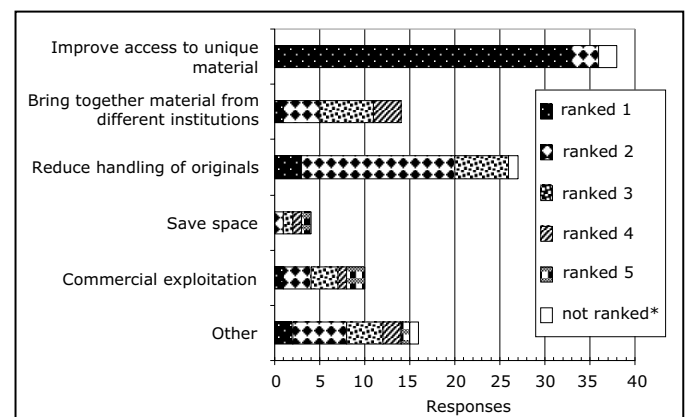
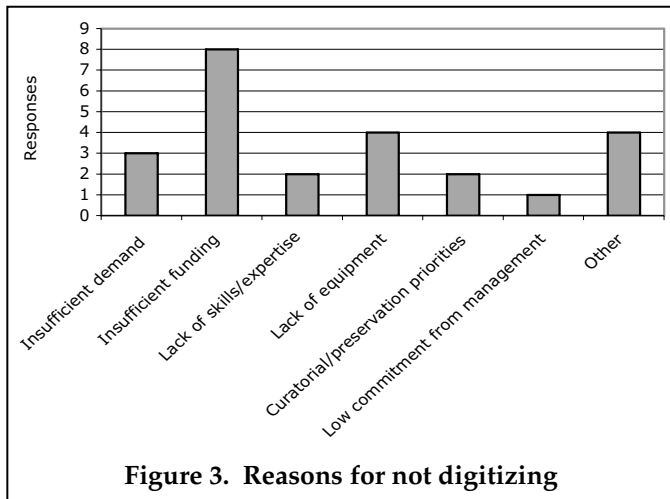


Figure 2. Reasons for digitization (Ranked)

\* There is a count for non-ranked votes because a few submissions selected criteria but did not rank them.



able to give multiple responses. The main reason for not digitizing was a general lack of resources, mainly funding, but also equipment and expertise. Copyright restrictions (2) and low priority (2) were 'other' reasons given.

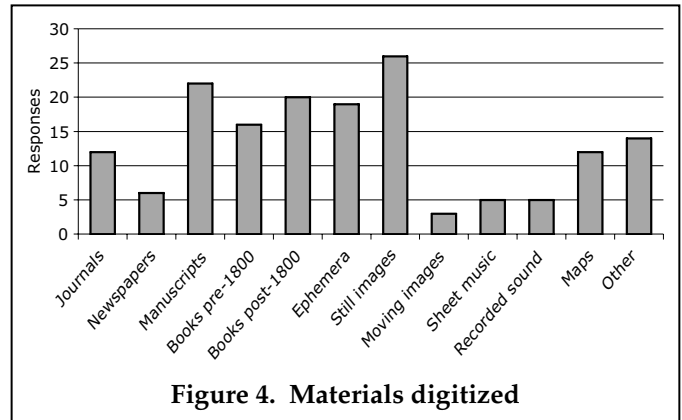
Interviewees suggested similar reasons for not digitizing material. Data ownership issues and prioritizing digitization of finding aids were also mentioned.

### 3.3 Materials digitized

The survey included a question on the types of material digitized by UK institutions. The responses revealed that still images and manuscripts were most frequently digitized. This is possibly because their conversion provides the best return on investment; the capture procedure for both materials is relatively simple but dramatically improves access to the materials. The desk research also produced this finding. Artefacts and artworks were mentioned five and three times, respectively. A few projects covered educational material such as reports, theses and exam papers. The most unusual original materials included shoes, needlework and bindings. Two respondents had digitized the entire range of materials shown in Figure 4.

### 3.4 Subject content of digitized resources

From the desk research it was clear that a large proportion of digitized resources are relevant for the arts and humanities research community. Fewer are relevant for social scientists and there is little in the natural and physical sciences area. The nature of research in the different sectors no doubt contributes to this distribution. It is probably safe to say that while arts and humanities researchers often use older materials, the pure scientists usually require more current information, much of which is born digital. Much of the digitization activity in science involves journal back files.

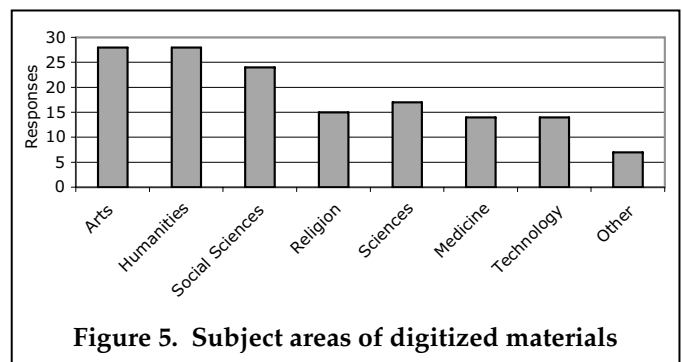


The survey responses confirmed the findings from the desk research. Again the predominant subjects are arts, humanities and social sciences.

Other subject areas include law and jurisprudence (2 respondents), genealogy and leisure pursuits. In one instance, the digitization project was intended as a 'taster', hence covered the whole diversity of special collections. Two respondents indicated that a wide spectrum of subjects is covered by their activities. See Figure 5 for results.

### 3.5 Selection criteria

All the representatives of projects and digitizing institutions interviewed approached digitization in different ways and reported different experiences. Early digitization projects were mainly small scale and involved one specific resource. As the digitization of resources has grown, some organizations have established strategies and criteria for the selection of material to digitize, while others continue to digitize according to market need (particularly publishers) and user feedback. We found little data in the literature on user demand for digitized content. Interviewees gauged demand through surveys and evaluations, while others commented that particular print collections were thought of as being better served by being in an electronic format. One organization looks



first at what can be made available, talks to the user community, holds focus groups and looks at general academic trends.

For some projects/organizations, the selection of material for digitization was linked with funding opportunities, cost and resource requirements. For certain material, for example, fragile, rare or unique, it was easier to apply for and be successful in funding applications. The funding bodies had varying selection criteria for funding digitization projects. Some provided funding in responsive mode and responded to each individual application, some were just beginning to introduce strategies, while others had specific strategies for funding in place:

Must enhance resources to scholarship. Must be led by scholars. Must involve more than one institution and more than one institution's material. Results/end result must be available to scholars that wouldn't be any other way. Results/end result must be of benefit to scholars. Audience must be the scholarly community. (Funding body)

Clear mechanism for making available material to scholars. Clear business model to manage and disseminate resources. Project must be sustainable. Legal arrangements taken care of e.g. rights to disseminate material, with technical issues addressed. (Funding body)

The survey included a question on selection criteria for digitization (see Figure 6 for results). All but one of the 34 digitizers responded to this question. Multiple responses were possible. The most frequent response was relevance to aims and objectives of the institution. Uniqueness or rarity was also a frequent response as were demand and the existence of coherent collections. It is interesting that it does seem, from the responses, that the majority of

respondents were selecting material for digitization according to good collection management principles rather than responding to the aims and objectives of funding sources (three respondents).

Other selection criteria given by survey respondents included:

- represent collection(s) (3)
- test methodologies (1)
- criteria still under discussion (1)

One organization interviewed had established a digitization approval board where each individual project was required to submit certain information regarding who would fund it, the timescale, IPR issues and how it would be delivered. Once submitted, this was then considered by the committee.

Another organization had established a priority list. Interviewees raised a number of issues in relation to the selection of material for digitization. Some organizations had too much material and found it difficult to prioritize, others found that although a list of criteria had been established, there was still too much material that fits the criteria.

### 3.6 Access to digitized material

Commercial companies in general charge for access to their digitized material. Three main publisher charging models were identified by the study:

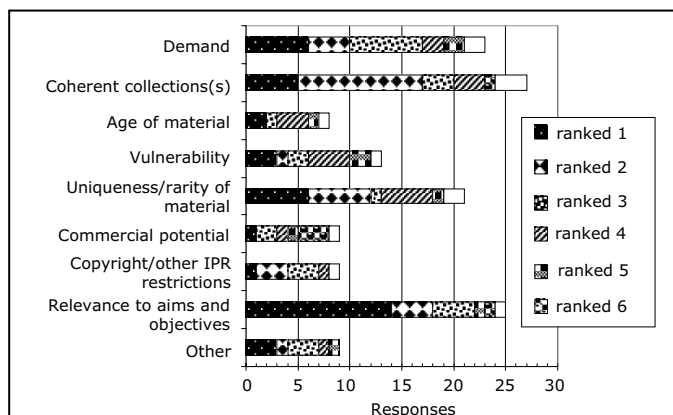
- payment of a lump sum for general access (subscription)
- payment on a usage basis (pay per view, or pay per download)
- paying to own content (outright purchase).

Publishers often offer different models for the same content. They may reserve particular payment modes for particular types of material. For example, journals are in general offered on a subscription basis, whereas large bodies of textual material might be offered for outright purchase. There are exceptions to charging for access to content, for example when public funding is given or when:

Societies themselves also pay for some journals to be digitized, and have asked that we make them freely available alongside any current subscription to the journal. (Publisher)

The publishers interviewed said that they regard it as important to respond to customer feedback, or even involve them in developing access models.

Library interviewees generally felt that their role was to provide free access to their resources. However, they struggle to apply this rule to digital material due to its costs, both in creation and maintenance.



**Figure 6. Selection criteria (Ranked)**

White sections represent non-ranked responses because a few submissions selected criteria but did not rank them.

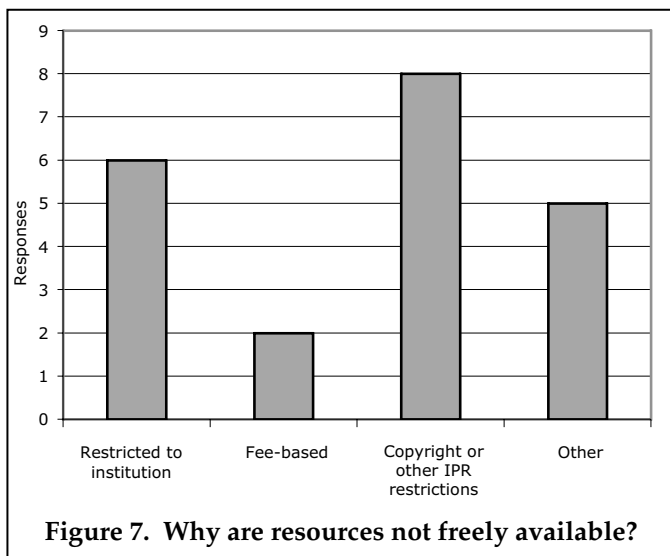
Survey respondents were asked about the accessibility of their digitized material. There were also questions on the existence of descriptive metadata for discovery purposes and access restrictions. Out of 34 digitizers, 32 replied to a question on how freely accessible their resources are. Twenty-one responded that they provide free access and 11 provide fee-based access. Additionally, six respondents selected both methods.

All of these respondents indicated what restrictions they placed on access. Multiple responses were possible. Respondents indicated that copyright is the most frequent hindrance to public access. Other responses included:

- Projects are not completed yet, but will be public in future (2).
- Externally funded resources are made freely available but internally funded resources are considered as 'institutional assets' to which access is restricted or resold (1).
- Not yet decided how to make it available, for some material it may just be available within the University, for other material we have to investigate copyright and determine the method of making it available (1).
- No formal delivery mechanism, all funded projects are freely available (1).

See Figure 7 for results.

The interviews revealed some examples of library and archives charging models for access. For example, the National Archives of Scotland (NAS) allow free viewing of their digitized wills, but charge for downloads. This model is also applied by the National Library of Wales (NLW) and the Wellcome Trust to some of their material. This model is based on the charging model for

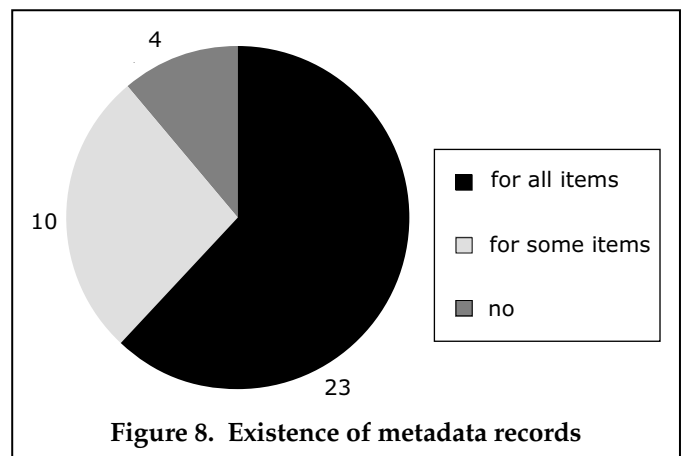


Access method	No. of responses
Website listing	25
Institutional Web catalogue	19
Portal	8
Consortium Web catalogue	6
Search engines index content	4
OAI-PMH compliant	1
Other indexing/listing mechanism	1
Not applicable	1

reproduction services. The benefit of charging is seen as the potential for covering maintenance costs and further development of the project.

There were 35 survey responses to a question on how resources available to external users, free or for a fee, are made accessible (see Table 1). Website listings and institutional catalogues were the predominant finding aids. One respondent said that they make resources available through the project website. Surprisingly, only one has made its resources OAI-PMH compliant.

All 34 digitizers (as well as three non-or not-yet digitizers) answered a question on metadata (see Figure 9). The aim of the question was to find out if metadata records were available for all digitized material. Around two-thirds indicated that metadata was created for all digitized items. Comments here included that metadata creation depended on the project, or was restricted to manuscript material. In one instance, there were backlogs, but eventually all items should have metadata records. One institution that created metadata for all items indicated that metadata creation was 'very time-consuming'. This issue is explored further in Section 3.1.1.1.



### 3.7 Management and preservation of digitized material

Interviewees thought that, in general, collection management was the responsibility of the organization/institution that hosts the digitized material. Some funding agreements specify that long-term management of the digitized collection/resource should be planned for, while others do not require or ask for collection management information. Support services offer assistance with planning collection management, for example some provide case studies of digitization projects to aid others in the planning and management of projects. Interviewees found such information useful.

Interviewees felt that good project management was a vital component of all digitization projects, and planning project management should be incorporated into the initial stages. Many respondents discussed the difficulty of successful project management due to the different roles and factors involved in any digitization project.

The long-term management of digitized collections was raised by those involved in digitization projects, in particular the cost and also the preservation of the collections/resources. Some interviewees felt neither they nor others had addressed these issues and that guidance was required in these areas.

Another issue raised was the need to add value to digitized resources. In the past, many projects only involved digitization. However, many stressed the need for resources to have added functionality appropriate to the user group.

#### 3.7.1 Preservation

The Association of Research Libraries has recently announced its endorsement of the production of digital surrogates as a method of preserving non-digital material. While ARL points out the advantages of digitization over methods of producing surrogates, such as photocopies or microforms, and describes the progress made in digital preservation, it is clear that there is still a lot of work to be done before the preservation of born digital and digitized material will be assured.

Digital preservation requires both technical strategies and supporting infrastructure. Technical strategies include migration and emulation. Simple strategies include 'refreshing' information and media migration to combat deterioration and obsolescence in storage media respectively. Conversion strategies to combat software obsolescence may rely on backward compatibility of new application software or interoperability of different software. Equally, it may involve more complex conversion processes. Digitizers can influence the 'preservability' of the resources they create through the standards they follow. While the use of standard formats may simplify the migration task, migration is still likely to be required because even standard formats change over

time. For more complex digitized resources, emulation may be required. The aim of emulation is to retain the look, feel and functionality of digital information through the use of software that allows new technological platforms to mimic the behaviour of older technology platforms. See Figure 9 for results.

Formats used by digitizers are discussed below. Survey respondents were also asked which technical preservation strategies they planned to use (see Figure 9). Thirty-eight respondents answered this question, including all digitizers. Multiple responses were possible. Most respondents were willing to refresh media, a short-term preservation measure. No respondent chose to emulate obsolete technology. One respondent said that they had a system in place, with a storage area network (SAN), daily back-up procedures, off-line and near-line archiving.

Currently some institutions are planning a change of storage methods:

- from CD to a SAN and local digital repository;
- from bit-stream maintenance by Computer Services to local digital repository;
- from tapes in different locations to new media.

Another institution is investigating the use of LOCKSS<sup>7</sup> software and of storage resource brokering software (SRB) 'for distributed replication'. Four institutions have yet to decide on preservation measures. One institution devolves preservation to the Archaeological Data Service, which has implemented an 'OAIS-based preservation programme'. Finally, three respondents do not envisage long-term preservation for their digital resources, which are 'ephemeral', digitized exam papers 'intended for a cohort of students' or considered as 'access rather than preservation copies'.

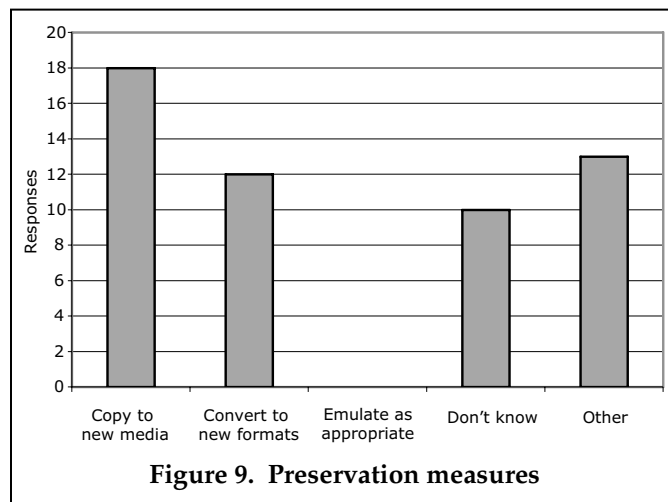


Figure 9. Preservation measures

### 3.8 Future digitization plans

Survey respondents were asked about their future digitization plans. One respondent did not answer this question, but a second respondent from the same institution did. A large majority (41) of institutions have holdings that ought to be digitized in their opinion.

Respondents were asked whether they planned to digitize the material they had listed. Twenty-six (63 percent) stated that there were plans to digitize while 15 respondents (37 percent) did not know. Results are shown in Figure 10.

Three of the six 'non-digitizers' that do not plan to digitize in future also hold collections that could be converted.

When asked why they thought this material should be digitized, 43 respondents replied (see Figure 11). These replies included respondents who did not plan to digitize the material themselves. Multiple responses were possible. Value for teaching and research were predominant reasons for digitizing remaining collections, followed by uniqueness/rarity, cultural heritage and access considerations. This is different from the responses given to the earlier question on reasons for digitizing material. There, increasing access and reducing handling were the main objectives of digitization activities.

Both 'other' comments pointed to digitizing to create 'preservation surrogates' that support 'preservation of heavily used and delicate material'. On commercial value, one respondent commented that 'Some have commercial value but I think this is overplayed – especially b[y] research universities/Russell G[r]ou[p]/larger public libraries/archives and others'.

### 3.9 Priorities for digitization

Interviewees commented that some institutions created digitization strategies/programmes or lists of collections/items that could be digitized. Other big institutions know that they have many collections of interest and hence define priority areas, but have no detailed lists. This point is reflected in the findings of the questionnaire survey, in that most respondents were not able to provide detailed lists. We were not able to get a meaningful idea from interviewees of obvious gaps or priorities areas for digitization.

### 3.10 Standards, formats, guidelines and existing policies

From the literature, good practice appears to be that 'Your design goal should be to hold master versions of all your data in forms that can be converted to meet varying purposes' (Arts and Humanities Data Service, 2003). The master file created from the original item should capture as much of the information content of the item as possible. This approach is likely to result in large file sizes, with implications for the amount of storage space required. The master files will also require more process-

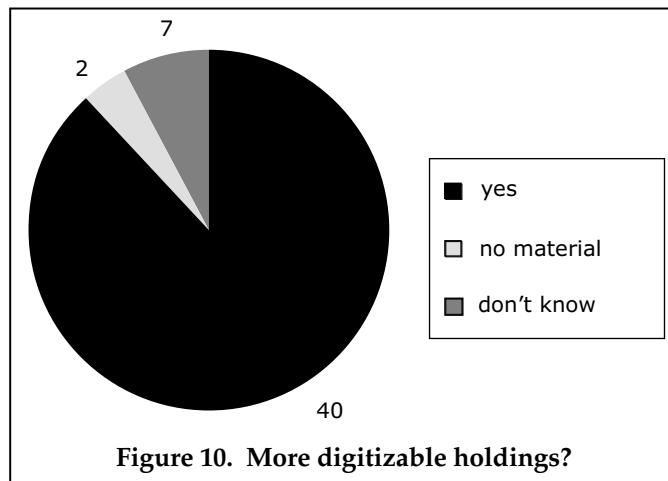


Figure 10. More digitizable holdings?

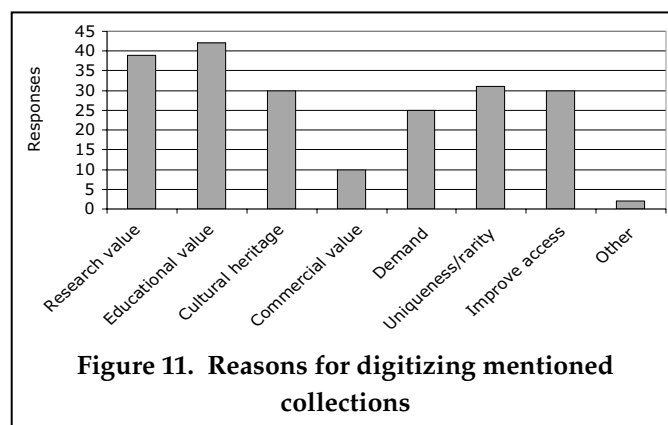


Figure 11. Reasons for digitizing mentioned collections

ing power for online viewing. To save storage space and accelerate downloads, files can be compressed. Lossless compression is recommended for the storage of master files (Hughes, 2004: 188).

'Proprietary' software is typically subject to (often costly) use licences. These invariably prohibit modification and redistribution of the software without express permission. The source code is not disclosed to users, and so the application cannot be adapted to individual systems. 'Non-proprietary' software can be copied, edited and distributed more freely. 'Closed' software gives the user no control over the application. 'Open' software is open to modification. However, while all 'non-proprietary' software is 'open', not all 'proprietary' software is 'closed', as the copyright owner can publish the source code and issue licences that allow copying, tweaking and redistribution. For instance, the PDF specification is freely available on condition that the new application includes specific access control mechanisms (Berglund et al., 2004).

For digitization projects, these notions are important



because closed proprietary systems create a dependency on the system provider, be it for increasing functionality or fixing bugs. Moreover, if the provider goes out of business or ceases to support the system, the user is left with a legacy system that cannot be adapted, since the code is unknown. However, some closed proprietary systems are de facto standards for certain applications, for example, PDF allows online viewing, downloading and printing of text documents while controlling/prohibiting modification. The use of open systems avoids any dependency but requires programming skills. BL's Turning the Pages software from Armadillo Systems is an example of a custom-developed proprietary application that probably will not become a standard but rather serves the niche market for delivery of high-resolution images of manuscripts together with written and spoken commentaries.

### 3.11 File formats

Of the 34 digitizing respondents, 33 indicated which standards they used for digital files (see Figure 12). Multiple answers were possible. There appears to be broad consensus about the use of TIFF (Tagged Image File Format) for master files and the JPG (Joint Photographic Experts Group) format family and PDF (Portable Document Format) for delivery. XML was used for both preservation and delivery more often than XHTML.

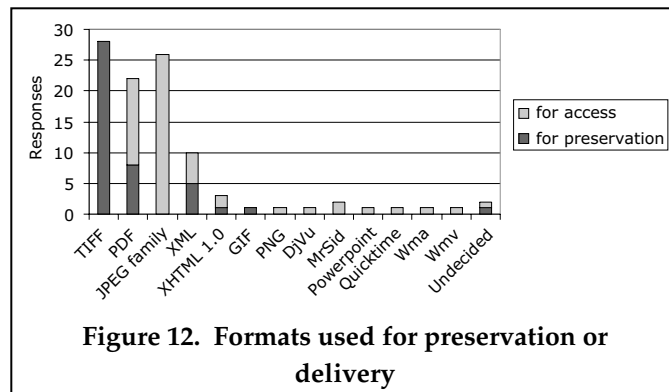
More recent and rarely mentioned formats included the digital negative format DNG, investigated for use as a master file format, and DjVu, MrSid and Luna Insight for delivery. It may be noteworthy that the Portable Network Graphics (PNG) and the more established GIF (Graphic Interchange Format) were mentioned by only one respondent. This may illustrate the slow take-up or ignorance of the promising PNG format or just be because no other respondent had any use for either GIF or PNG.

Not all of the digitizing organizations interviewed provided details of file formats used. Interviewees stated that the requirements for formats were stability, formats that have been used in the past, formats required by users/funding bodies, and similar formats used in the discipline. A small number of projects used formats that were dictated by partners they worked with (e.g. JSTOR). The interviewees used mainly TIFF, PDF, HTML and XML; they also mentioned the use of SGML.

While a number of interviewees discussed the need for standards in relation to file formats, preservation and interoperability, the majority did not.

#### 3.11.1 Metadata

Out of the 34 survey respondents that had digitized, 28 indicated the metadata schemes they used. Multiple answers were possible. Dublin Core was the most frequent response, followed by MARC. However, there are many other schemes in use, which is likely to impact



on interoperability. One reply explained that metadata used was 'project specific'. Interviewees also commented on the metadata schemas used in their projects. While there is no one scheme or template that is adopted as standard, from our findings we can tentatively suggest some trends. The library-based projects are mostly using some form of Dublin Core or MARC and using XML and METS encoding. Archives use the EAD and ISAD(G) schemas for records and finding tools to meet their own needs. There seems to be less standardization amongst publishers and digitization services. It does seem clear that the choice of metadata format depends on what is being digitized and for what purpose. While the majority of organizations/projects interviewed managed and produced their own metadata, one used an outside organization.

Some organizations interviewed adopt metadata that is either recommended by the funding body, or required by the user community, for example, libraries. Many stated that in the area of metadata and standards they were on a steep learning curve and one organization in particular had implemented a metadata working party in order to facilitate work in this area.

Dynamism in the field of metadata was also mentioned, making it difficult to decide on one particular standard or scheme. While a small number of individual organizations and projects were aware of, and to some extent, involved in work on interoperability, or adopted standards with interoperability in mind, this was very limited, and many projects stressed they did not have the time or funds to become more involved. A number of interviewees stressed the need for further guidelines, specifically a set of guidelines about how to apply metadata. Some seemed unsure about what metadata is required and what different schemas and sets exist. Others suggested the creation of a list of metadata used by projects to determine whether or not a metadata consensus is building.

Cost was a recurring issue in the interviews and some interviewees argued that the cost of producing the metadata is becoming an obstacle to completion of

digitization projects and resources. The possibility of automating metadata production was raised by one interviewee. The organization was exploring automatic extraction of metadata from files, but this would then have to be entered manually. The AMeGA (Automatic Metadata Generation Applications) project has focused on overcoming this problem (Greenberg et al., 2005).

### 3.11.2 Subject schemes used

Out of the 34 survey respondents who had digitized material, 26 indicated the classification schemes used. Multiple answers were possible. Library of Congress Subject Headings was the most frequent response, 'own' subject access systems was the second most frequent response. Respondents did not say why they were using their own systems or how they were developed. However, several specialized thesauri were mentioned. A number of replies were not detailed enough ('thesaurus', 'subject headings') to be meaningful.

Fewer than two-thirds (22) of the 34 digitizers answered a question on unique identifiers (see Table 2). There was no obvious trend in responses here. Respondents were either following an in-house protocol or using identifiers issued by the library management or content management system.

## 3.12 Support services for digitization

The desk research identified a number of support services in the UK and representatives of a number of services were interviewed. The library and archives survey included questions on use of support services.

Digitization support services are generally funded in one of three ways: self-funded, funding received from JISC and/or other funding bodies, for example AHRC, or a combination of both. Some services initially received funding but are now self-sustaining. There is much similarity between the services offered by the different organizations. These include:

- mailing lists
- advice and expertise in all aspects of digitization projects
- training and workshops
- print and Web documents
- guides and standards information
- preservation information.

Other services offered by particular services include consultancy, project management, project management training, digitization, sustainability of collections, including economic sustainability, digital preservation management, assisting with funding applications and conferences. While some organizations offer specific services, others offer assistance covering the entire process of digitization. Some support services focus on specific disciplines or areas of digitization, for example

**Table 2. Object identifiers used by survey respondents**

Identifier type	No. of responses
Unique number/ID/filename	4
Database identifiers	3
In-house protocol	3
ISBN	3
Not applicable	3
ISSN	2
Library reference codes	2
Control numbers and technical metadata	1
Institutional reference number	1
Item records/barcodes	1
POI	1
Undecided	1
URL	1
None	1

images or manuscripts, arts and humanities or sciences. Some services acquire and curate digital collections while others limit services to advice and training.

While interviewees stated that all the services offered were used frequently, some services were used and requested more often. These included hands-on workshops, Web documents and other advice on all aspects of project management, help desk services and ongoing support for projects following training.

The majority of support services promote and offer guidance and advice on metadata standards and file formats. While some make strong recommendations for metadata and standards, the majority only make projects/clients aware of standards and formats and cannot do more than encourage use. Each digitization project and discipline has different needs and therefore one set of standards cannot be recommended. However, all of the support services interviewed stated that they felt their role in providing guidance and advice in relation to metadata and standards had become more important and valued. Some felt they provided an increasing amount of detailed information on metadata and standards. Because of this, many of the services consult with other bodies in the creation/setting of metadata standards.

Some examples of services include the Arts and Humanities Data Service (AHDS), British Universities Film and Video Council (BUFVC), Higher Education Data Service (HEDS), Technical Advisory Service for Images (TASI) and UK Office for Library and Information Networking (UKOLN). HERON is a copyright clearance and document delivery service. The Centre for Data Digitisation and Analysis (CDDA) undertakes digitization.

The Digital Curation Centre (DCC) is mainly concerned with born digital material, but the expertise gained by this new service could also benefit digitized collections.

### 3.12.1 Use of support services by survey respondents

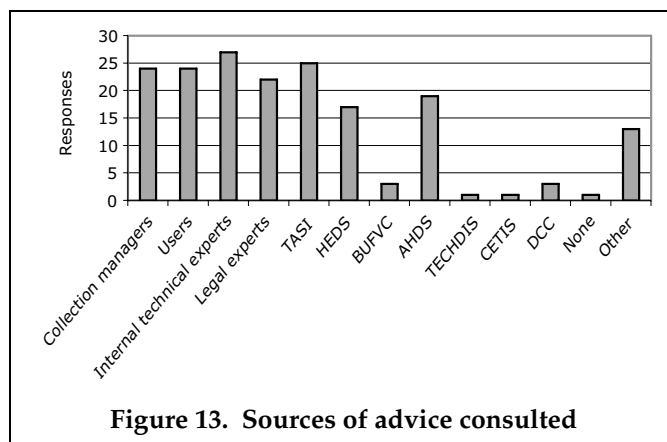
Survey respondents were asked which sources of advice they had used. Thirty-seven responses were received. Multiple responses were possible. Results can be seen in Figure 13.

Internal sources of advice, including collection managers and technical experts were frequent responses. Users were also used as a source of advice, presumably on selection. JISC-funded services, such as TASI and AHDS were also used. The Digital Curation Centre may become more widely used as it becomes more established and institutions have more need of support in preserving resources. The lack of use of the BUFVC may reflect the nature of the source materials digitized, since most respondents have digitized still rather than moving images. The accessibility consultancy service TECHDIS is also little used. The single answer for 'none' is contradicted by a second submission from the same institution, which indicates that several sources of advice were indeed consulted, most of them internal. Some respondents used overseas sources of advice including other research libraries, or organizations such as OCLC, the Council on Library and Information Resources, and the Digital Library Federation. Conferences, work shops and mailing lists were also mentioned as sources of advice.

## 4. FUNDING FOR DIGITIZATION PROJECTS

There have been several studies on the costing<sup>8</sup> of digitization projects. Costs for digitization are significant and include: documentation and preparation, conversion, ensuring copyright status and rights clearance of material, equipment, human resources and ongoing maintenance. From the research, it looks as though UK projects obtain funds from a range and combination of sources, including donations and sponsorship as well as institutional budgets and public grants.

Some funding bodies have strategies for funding digitization, while others do not specifically fund digitization (these bodies are aware that digitization is included in some of the project funding allocated, but do not fund pure digitization projects). When projects/groups apply for funds to digitize resources/collections, certain bodies are generally the first port of call. The funding body selected usually depends on what is to be digitized, the subject area, the amount of funds required and the target user community. Some interviewees named certain funding bodies as regular funders for their digitization activities (such as the Andrew W. Mellon Foundation). Some mentioned that applications to certain funding



bodies had been unsuccessful, and therefore in most cases these were not contacted again. Other funding bodies were avoided due to lengthy processes of application and difficulty in receiving funding (for example the European Commission).

Funding bodies have two general approaches to allocating funding. A number of them have strategic priorities, while others operate in responsive mode, that is, on an ad hoc basis, depending on applications received. A few funding bodies stated that funding was driven by the research interests of the community. A number commented that because of this, the allocation of funding is uneven and some bodies are therefore considering determining some strategic priorities.

Representatives from some funding bodies reported collaborative activity. This was usually, however, in the joint funding of a particular project or initiative and did not involve further collaboration in relation to funding strategies or input into standards or formats. Millions of pounds have been spent on digitization projects in the UK, and a number of project representatives reported receiving funding from a number of sources. Again, who was approached for funding varied depending on the organization requiring the funding, the material/resources being digitized and the target audience. A number of funding bodies reported collaboration with support services, either through funding a support service or through specific collaboration, for example, AHRB and JISC fund AHDS, and AHDS provides technical input to the selection of applications for AHRB funding. Individual projects collaborate with their funding bodies as specified by the individual funding body.

### 4.1 Funding sources used by UK libraries and archives

Thirty-eight respondents to the survey indicated their funding sources. These respondents included institutions in the planning stages of digital projects.

Just under half of the respondents had combined external and internal funding, but this was not the only funding model. Some institutions relied solely on internal funding and others only used external funding (see Figure 14).

Twenty-eight respondents, including all 27 that selected external or combined funds from the previous question, answered this question (see Figure 15). Multiple responses were possible. Public sector funding was most frequent, particularly National Lottery funding. Respondents specified which trusts and other bodies had funded their digitization efforts (see Table 3).

Of the 27 respondents that indicated that external funding, alone or combined with internal funds, was the main source of funding 26 (along with two others) responded to a question on the proportion of the digitization budget made up by external funding. Multiple answers were possible, because proportions would depend on the project and some institutions had undertaken several projects. While many used internal money,

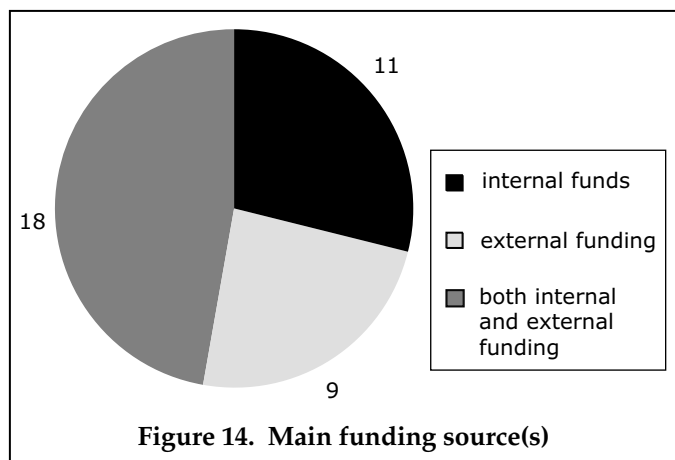


Figure 14. Main funding source(s)

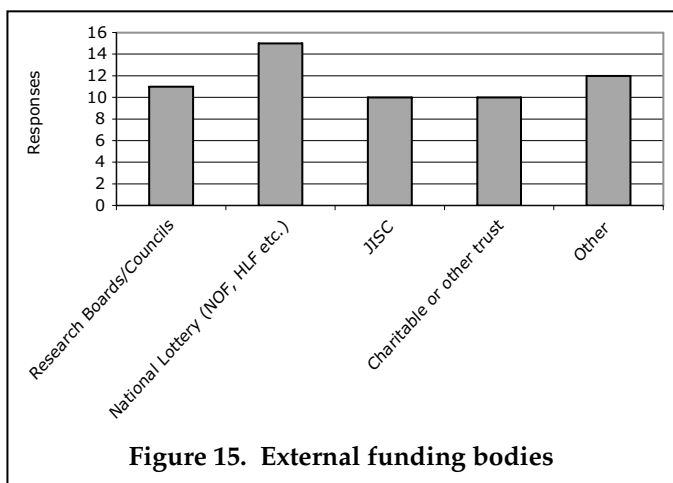


Figure 15. External funding bodies

Other trusts	Other funders
Andrew W. Mellon Foundation	(Anonymous) donors/sponsors
Catherine Cookson Trust	'external industrial contract'
Corson bequest	British Library
Getty Foundation	Commercial partners (3)
Hansard Trust	Department of Trade and Industry
Leverhulme	East Midlands Museums Libraries and Archives Council
Pilgrim Trust	Genealogical Society of Utah
SCRAN	Irish Government
Wellcome Trust	Readers
	Research Support Libraries Programme (2)

for most, external money made up the bulk of the digitization budget (see Figure 16).

Comments under 'other' included the fact that the percentage depended either on the project or on the collection. In one institution, most of the costs were covered by the Government, while project partners (and anonymous donations to these) covered some costs. At another, 'We obtain contract- and project-based work from clients who have received funding from a variety of courses. We do a lot of sub-contracted work for HEDS'.

### 4.2 Funding issues

A key concern of those that had received funding for digitization was the significant cost associated with it. Another cost was the need for added functionality of resources to meet the increasing expectations of users. Many stressed that without significant external funding, digitization and the management of those digitized

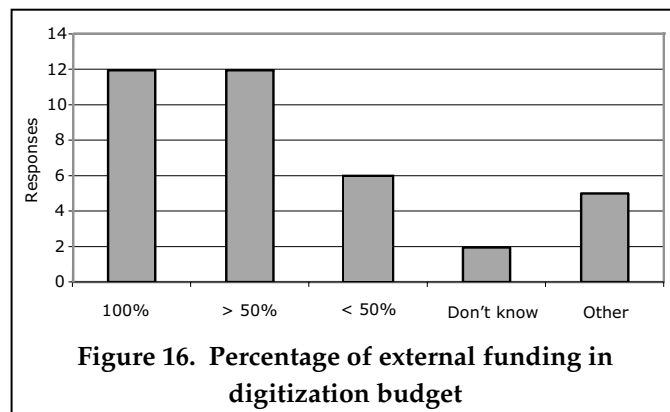


Figure 16. Percentage of external funding in digitization budget

resources would not be possible. Many funding bodies interviewed felt that projects they funded did not take into account the long-term issues such as preservation and sustainability as well as access to the resources. Because of this, a number of funding bodies are limiting funding unless digitization projects factor in sustainability of the resources. Some interviewees involved in digitization projects felt that funding bodies needed to look at their strategies and provide funding for the preservation and maintenance of digitized material.

Like the survey respondents, interviewees commented that their institutions held many more resources that should be digitized, but pointed out this could only be if further funding could be secured. However, another main concern of the funding bodies is decreasing or limited budgets available for the digitization of material. Though viewed as important, many feel the future of digitization and digitized resources is precarious because it depends on the limited funding available. Others stated that it was increasingly difficult to decide where the priorities lie for the digitization of resources and struggle with establishing strategies to manage priorities.

Future plans of funding bodies do include ensuring digitization projects take into account the preservation and sustainability of resources, ensuring digitized resources are accessible and add value to the appropriate user community, considering how best to serve user needs, and working on joint funding for larger important projects they cannot fund alone. One funding body interviewed stressed the necessity for a needs assessment of digitized material in the UK and felt future funding coordination should be strategic. Other funding bodies were concerned about the Google digitization plans and felt that their future digitization funding would depend on the amount of digitization activity assigned to Google.

A number of interviewees commented that their funding did not come from UK funding bodies, but international foundations. Others had begun fundraising activities to secure funds.

## 5. COOPERATIVE ACTIVITIES IN DIGITIZATION

There is some evidence of collaboration between higher education institution libraries, learned societies, museums, archives and trusts. In particular, there is a great deal of cooperative activity in Scotland. Prominent examples include the Glasgow Digital Library, a cooperative endeavour of a number of Scottish libraries.<sup>9</sup> The Scottish Cultural Resources Access Network (SCRAN) is a charity financed primarily by the Scottish Executive. It is a service for libraries and schools in Scotland and provides educational access to digital materials representing Scottish material culture and history. SCRAN acts principally as a standards centre, a funder, a project manager and a host for material. SCRAN acts as a

network; digitization is done by the participating institution. Access to material is chargeable. The (UK) National Archives is involved in a number of cooperative projects including Moving Here, an online service on migration to the UK over the last 200 years.

Thirty-nine survey respondents answered a question on cooperative digitization activities. Four of these had already indicated they had not, up until now, actually carried out any digitization. Twenty respondents (56 percent) had been involved in cooperative activities, 17 respondents had not. Twenty-one institutions were interested in future cooperation.

### 5.1 Partners in cooperation

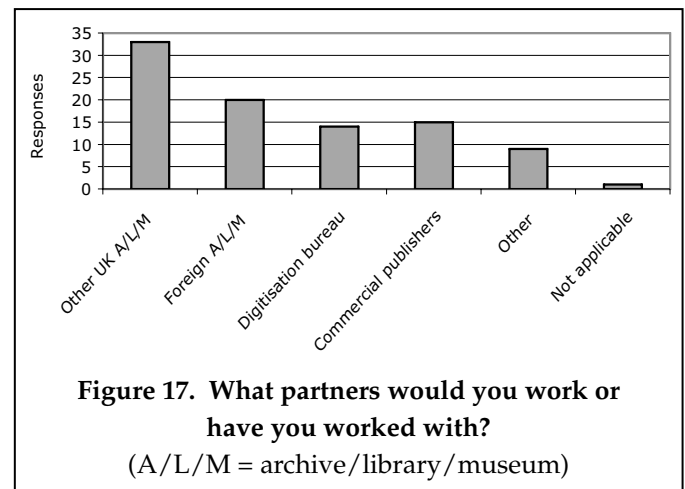
Thirty-seven respondents answered a question on cooperation partners (see Figure 17). Multiple answers were possible. Other UK and overseas libraries, museums and archives were the preferred partners. Some respondents had used digitization bureaux and cooperated with commercial publishers.

Few respondents provided additional details about partners. Those that did mentioned other libraries and archives, specialist technical experts and commercial publishers.

One respondent stated that they were interested in potentially any collaboration. Another respondent was not sure whether to reply to this question as the institution itself offered a comprehensive digitization service.

Nineteen respondents provided information on their commercial partners (see Figure 18). Multiple responses were possible. The most frequent commercial collaboration was outsourced digitization. One respondent commented that the nature of collaboration varied from project to project.

Respondents were asked an explicit question on whether they carried out digitization in-house or whether they outsourced this activity. Thirty-seven



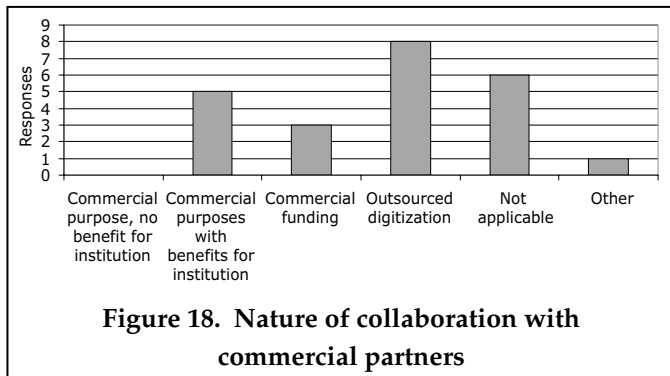


Figure 18. Nature of collaboration with commercial partners

respondents answered this question (see Figure 19). For reasons for these decisions see Table 4. Cost was a factor in these decisions; this may depend on the materials to be digitized.

A question aimed at digitizers that had cooperated, asking about their reasons, received 33 responses (see Figure 20). Multiple answers were possible here. (The structure of the questionnaire did not prohibit answers from respondents who had not actually been involved in cooperative activities.) The most frequent reason given for cooperating was building virtual collections from dispersed materials, followed by sharing expertise and infrastructure.

Some funders (two respondents) required that the project be collaborative in order to be eligible for support. In one case, the cooperation was seen as a means to achieve wider dissemination of the project results. The seven 'not applicable' responses came from institutions that only had 'solitary' digitization experiences.

A question on quality control procedures received 34 responses, 19 from respondents who had already cooperated and 14 from other respondents (see Figure 21). Multiple responses were possible. Most of the respondents relied on in-house checking, although some did use external companies and automated checking.

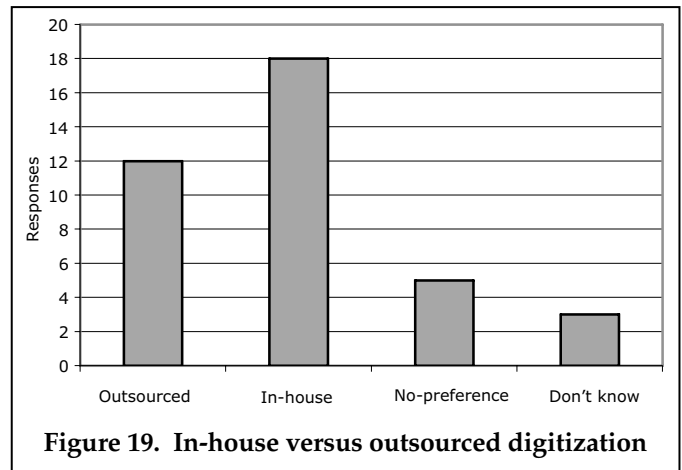


Figure 19. In-house versus outsourced digitization

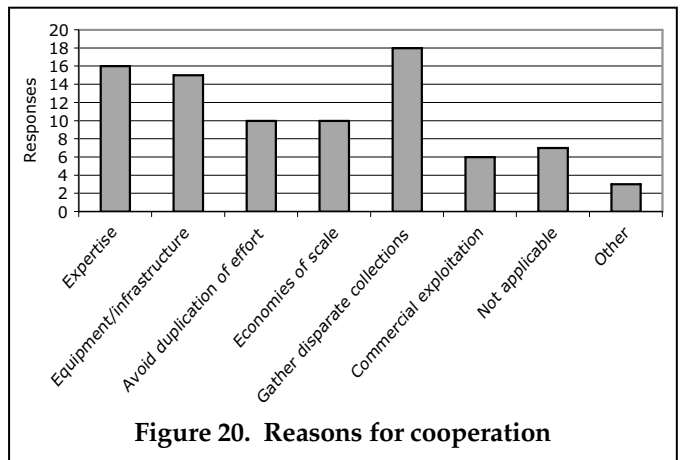
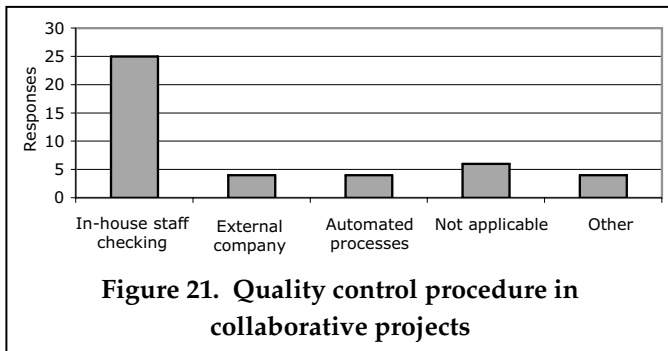


Figure 20. Reasons for cooperation

'Other' replies were varied. In one instance, quality was controlled by project partners, as the respondent's institution acted as content provider only. In addition to in-house checking and automated control, the ADS relies on 'user reportage' to detect remaining flaws. The six 'not

Table 4. Reasons outsourcing or digitizing in-house

Reasons given in favour of outsourcing	Reasons given in favour of in-house digitization
Lack of <ul style="list-style-type: none"> <li>• equipment (4)</li> <li>• staff/time (3)</li> <li>• expertise (3)</li> <li>• space (1)</li> <li>• resources for copyright clearance (1)</li> <li>• money ('If we have more than 100 items, it is more cost effective and efficient to use external agency')</li> </ul>	<ul style="list-style-type: none"> <li>• better control of procedures, handling of the originals or quality (5)</li> <li>• they preferred or were required to keep the originals on-site (4)</li> <li>• cost (4)</li> <li>• develop staff skills (3)</li> <li>• small-scale project (2)</li> <li>• presence of internal expertise (2) and facilities (2)</li> <li>• 'weed out duplicate material prior to digitization' (1)</li> </ul>



applicable' responses came from institutions that only had 'solitary' digitization experience.

## 6. THE FUTURE OF DIGITIZATION

There was a consensus among interviewees that there will be more digitization in the future. Some libraries and archives have started to make provisions for ongoing digitization activities. For example, they have established appropriate posts, have ring-fenced funding and have policies and strategies to deal with future digitization.

Publicly funded service providers are aware of their temporary status and are to some extent unsure about their future, whereas self-sufficient service providers are more confident.

Publishers interviewed were cautious on future involvement in digitization activities. They see the digital realm as the future but recent developments in the sector, for example the Open Archives Initiative and Google Print, impinge on the publishers' traditional roles. A number of libraries have successfully run digitization programmes which to some extent supersede publishers' digitization activities. Publishers are aware that this might have repercussions for some of their business and think tentatively ahead:

I do think that ... as more and more content becomes available, freely available, through publicly funded projects and through initiatives like Google, the importance of organizing information, giving access to information, having good metadata, having good indexing tools, having good finding tools is increased and that's something that we can contribute. (Publisher)

Funding bodies recognize the increasing importance of digital resource provision and are responding to it. Some are doing so in a proactive mode, through incorporating sustainability, open access and preservation in their programmes and developing more strategic initiatives. Others are acting more reactively.

Interviewees thought that the lack of an overall strategy for digitization in the UK was a cause for

concern. One interviewee summed up the present approach to digitization:

It's haphazard, it's ill-focused ... it's cherry picking collections as opposed to strategic, well-planned. ... It's not based on analysis. Institutions don't do an analysis of their holdings which I think they really ought to do ... to digitize based upon an analysis of their user needs. Who are their user communities? (Digitization service)

Some digitizers may see this as contentious and some of our respondents did say they looked into user needs.

The majority of interviewees agreed that having a national strategy on digitization would be desirable. Representatives from libraries, archives and subject representatives were particularly supportive, whereas there was hesitation among some of the funding bodies. The opinions of service providers was divided.

Positive reasons given for the creation of a national strategy involved coordination, including standards, selection criteria, funding allocations and cooperation. There was no clear idea among interviewees of the content of a national strategy or who should develop it.

One interviewee spoke vehemently against a national digitization strategy, saying that it would stifle innovation and would be nearly impossible to achieve. While this interviewee agreed with what a national strategy would aim to achieve, they argued that this would be done better in a non-regimented environment. It would be achieved by having clear guidelines.

Another interviewee said:

... you're actually saying that we want to shape what people are doing whereas I'm saying that at the moment we don't even know what they're doing ... (Library)

The aim of this study was to help provide this overview of activities and existing content, although more needs to be done.

Finally, some interviewees also commented on a possible national infrastructure for digitization. One digital library expert thought that while it should be coordinated, it should be distributed.

## 7. CONCLUSIONS AND RECOMMENDATIONS

It is clear from discussions with the various players in digitization that there is a need for coordination in digitization activities, even though they do not all agree on how this should be done. A UK-wide strategy could assist in filling gaps in provision, cut across the efforts of individual funders and digitizing organizations, reduce overlaps between support services and assist in the provision, take up and use of open-access resources. While librarians and archivists have sought to find and adhere to standards and JISC has supported this, 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. The main organizations that could lead this strategy are the Research Information Network and JISC. However, other organizations, including the research funding councils and CURL also have an interest. The remit of the Museums, Libraries and Archives Council covers all library and archives sectors as well as museums and the MLA already has a role in the co-ordination of digitization in the European Union.

### 7.1 Researcher needs

One of the questions considered in this study was subject areas where there is significant demand from researchers. Although it was not possible to carry out a systematic survey of user needs, the study team approached a small selection of research bodies and societies. While interviewees provided some suggestions of gaps, the study team were not able to get a strong feel for the nature and levels of demand. One point that was raised was the *lack* of demand for digitized material, particularly in arts and humanities. These results highlight the need for a coordinated and systematic survey of user needs, particularly in the sciences and social sciences. This could be carried out through research funding bodies, a more comprehensive survey of the views of subject associations, academies and royal societies, or through the newly established Research Libraries Network.

The findings of these studies could inform policies and strategies of the research councils and be shared with other interested bodies, including JISC and CURL and/or the Research Libraries Network so that the response to the findings can be coordinated.

### 7.2 Priorities for future digitization

The research libraries surveyed for this study provided some information on collections that remain to be digitized. These 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 Research Information Network. However, there is a question of whether material should be digitized 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 digitization 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.

### 7.3 Identification of existing digital collections

The survey indicated some issues in the creation of metadata records for digitized material. It seems that in some cases, records do not exist for the originals and metadata creation for these is a higher priority than digitization. Metadata creation is an expensive part of the digitization process. It therefore seems sensible that metadata creation is costed into funding bids and that funding bodies be prepared to fund it. It seems pointless to digitize without providing the means to retrieve digitized resources. It would also be unfortunate if digitization 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.

Information on digitized resources should also be covered in the search tools used by researchers. Some resources already are, but coverage needs to become more comprehensive. Our survey found little evidence of OAI-PMH compliance; harvesting of metadata records and the provision of search services would be worth exploring.

A comprehensive listing of existing digitized resources could facilitate the analysis of gaps in provision. The creation of new digitized resources to meet identified needs could also be facilitated by a list, not only of what has already been digitized, but also of what is in the process of being digitized. There is a need for a better mechanism for identifying relevant projects and collections. There is a precedent here in the Mellon Microfilming Programme, which involved filming material to preservation standards and creating and submitting bibliographic records to various registers, both in the UK and overseas. Project workers could avoid duplication of effort by identifying material that had already been microfilmed.

Registers and catalogues for digitized material already exist. The systematic submission of information on digitization projects and material digitized 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/DLF 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 digitized resources, so the registers in other countries should be examined as models. Any system would need to be simple and inexpensive to contribute to, in order to maximize participation.



Digitizing organizations may well need to be motivated to submit information on projects and digitized material. This may be difficult in the private sector, although publishers may find benefits in a wider awareness of their digitized 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.

#### 7.4 Standards and formats and collection management issues

There are several sources of guidance on standards and formats relevant to digitization. This study has shown that whilst individual projects do things a little differently and that standards and formats depend on materials digitized and purposes, there is a core set of standards and formats used by many projects. There seems to be less standardization amongst publishers and digitization services. It does seem clear that the choice of metadata format depends on what is being digitized and for what purpose.

The survey carried out for this study showed that digitizers were using a number of different services and sources of advice. It may be useful to have a single point of access to guidance and advice on different aspects of digitization, including technical, legal and management guidelines and case studies. The advice may be provided by different services, but the users would have one access route.

Respondents to the questionnaire survey also seemed concerned about the long-term management of digitized resources, in terms of both funding and expertise. 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 digitizers. Digitizers need guidance on long-term management and preservation. They need to be aware of what 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 Digital Curation Centre should also be able to help here. 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 digitized resources when planning and funding digitization 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.

#### 7.5 Funding opportunities

The study found that lack of funding was a major deterrent to digitization. At the moment there is a plethora of funding bodies and opportunities and there is a hint from the study that organizations planning to digitize have to spend time identifying and exploring funding opportunities. It would seem sensible to have a more coordinated approach to the identification of funding opportunities. Support bodies 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.

It has become clear during the course of the study that coordination is needed. We tentatively suggest that any national strategy has to be formulated at a very high level and centralized 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 digitization in the UK, but it should be possible to improve coordination.

The Google initiative is currently an unknown quantity, but could have a major impact on business models and research library interest in digitization. Publishers who participated in this study are clearly concerned about the implications of Google for future commercial digitization activities, while libraries are cautiously hopeful. The Google initiative has the potential not only to facilitate the digitization of library materials for libraries, but for the existence of the digitized material to become easily discoverable through Google services. As mentioned by interviewees, the Google initiative will only be useful if material is digitized to an acceptable standard and if appropriate metadata is created for digitized material. If this is the case this initiative may well prove to be a significant boost for the digitization of content. Whether this will be systematic digitization of content to meet needs or cherry picking of significant collections is another matter.

#### ACKNOWLEDGEMENTS

The Joint Information Systems Committee and the Consortium of Research Libraries funded this work. The authors would also like to thank all participants in the research and JISC, CURL and Loughborough University colleagues who provided support.

#### NOTES

1. *UK Data Archive is a centre of expertise in data acquisition, preservation, dissemination and promotion.* <<http://www.data-archive.ac.uk/>>
2. The Arts and Humanities Data Service (AHDS) <<http://ahds.ac.uk/>> is a UK national service aiding

the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities.

3. Higher Education Digitisation Service (HEDS) <http://heds.herts.ac.uk/> provides advice, consultancy and a complete production service for digitization and digital resource development and management to the higher education sector, museums, public and national libraries, archives and other not-for-profit organizations.
4. The Resource Discovery Network (RDN) <<http://www.rdn.ac.uk/>> is a cooperative network consisting of a central organization and a number of independent service providers called hubs offering subject portals.
5. [http://portal.unesco.org/ci/en/ev.php-URL\\_ID=1538&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=1538&URL_DO=DO_TOPIC&URL_SECTION=201.html)
6. Carried out by Bültmann.
7. 'LOCKSS is open source, peer-to-peer software that functions as a persistent access preservation system. Information is delivered via the web, and stored using a sophisticated but easy to use caching system', URL (consulted February 2005) <http://lockss.stanford.edu/>
8. For example see Lee (2001, Chapter 4) and Tanner and Lomax Smith (1999). Tanner and Smith from the HEDS mention costs per unit item of between £0.10 and £1.50 depending on quality (conversion cost only).
9. Glasgow Caledonian University, Glasgow City Libraries and Archives, Glasgow Colleges Group, University of Glasgow, University of Strathclyde.

## REFERENCES

- Arts and Humanities Data Service (2003) *Introduction to Creating Digital Resources*, URL (consulted December 2004): <http://www.ahds.ac.uk/creating/information-papers/creating-introduction/index.htm>
- Astle, P.J. and Muir, A. (2002) 'Digitization and Preservation in Public Libraries and Archives', *Journal of Librarianship and Information Science* 34 (2): 67–9.
- Berglund, Ylva et al. (2004) *An Investigation into Free Ebooks: Final Report*, URL (consulted December 2004): <http://ahds.ac.uk/litlangling/ebooks/report/FreeEbooks.html#ebooks-div-id2669731>
- European Commission (2002) *Coordinating Digitisation in Europe: Progress Report of the National Representatives Group: Coordination Mechanisms for Digitisation Policies and Programmes*, URL (consulted September 2004): <http://www.minervaeurope.org/publications/globalreport/globalrep2002.htm>
- Greenberg, J., Spurgin, K. and Crystal, A. (2005) *Final Report for the AMeGA (Automatic Metadata Generation Applications) Project*, URL (consulted April 2005): [http://www.loc.gov/catdir/bibcontrol/lc\\_amega\\_final\\_report.pdf](http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf)
- Hughes, Lorna M. (2004) *Digitizing Collections: Strategic Issues for the Information Manager*. London: Facet.
- Kenney, A.R. and Rieger, O.Y. (eds) (2000) *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Mountain View, CA: Research Libraries Group.
- Lee, S.D. (2001) *Scoping the Future of the University of Oxford's Digital Library Collections: Final Report*, URL (consulted September 2004): <http://www.bodley.ox.ac.uk/scoping>
- Muir, A. (2004) 'Digital Preservation: Awareness, Responsibility and Rights Issues', *Journal of Information Science* 30 (1): 73–92.
- National Audit Office (2004) *The British Library – Providing Services beyond the Reading Rooms*, URL (consulted September 2004): [http://www.nao.org.uk/publications/nao\\_reports/03-04/0304879.pdf](http://www.nao.org.uk/publications/nao_reports/03-04/0304879.pdf)
- Tanner, Simon and Lomax Smith, Joanne (1999) *Digitisation: How Much Does It Really Cost? Paper for the Digital Resources for the Humanities 1999 Conference, Sept. 12–15*, URL (consulted September 2004): <http://heds.herts.ac.uk/resources/papersI.html>
- Whitelaw, A. and Joy, G. (2000) *Summative Evaluation of Phase 1 and 2 of the eLib Initiative: Final Report*, URL (consulted September 2004): [www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/elib-fr-v1-2.pdf](http://www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/elib-fr-v1-2.pdf)