

## **Digitised Content in the UK Research Library and Archives Sector**

[This is a Pre-proof version of an article which appeared in the journal,  
Journal of Librarianship and Information Science, 38(2), pp. 105-122]

Barbara Bültmann  
Rachel Hardy  
Adrienne Muir\*  
Clara Wictor

Barbara Bültmann is Librarian at the German Historical Institute, 17  
Bloomsbury Square, London WC1A 2NJ, Tel +44 (0)207 309 2019  
Fax +44-(0)20-7309 2069 / 7309 2072, Email [bbueltmann@ghil.ac.uk](mailto:bbueltmann@ghil.ac.uk)

Adrienne Muir is a lecturer in the Department of Information Science at  
Loughborough University, Loughborough, LE11 3TU, Tel +44 (0)1509  
223064, Fax +44 (0)1509 223053, Email [A.Muir@lboro.ac.uk](mailto:A.Muir@lboro.ac.uk)

Rachel Hardy is a Research Associate in the Research School of Informatics,  
Holywell Park, Loughborough University, Loughborough, LE11 3TU, Tel +44  
(0)1509 635724, Fax (0)1509 223053, Email [r.l.hardy@lboro.ac.uk](mailto:r.l.hardy@lboro.ac.uk)

Clara Wictor works in the Bibliothèque nationale de Luxembourg, 37  
Boulevard F.D. Roosevelt, L-2450 Luxembourg, Tel +352 22 97 55-1, Fax  
+352 47 56 72, Email [bib.nat@bi.etat.lu](mailto:bib.nat@bi.etat.lu)

\*To whom correspondence should be addressed

# Digitised Content in the UK Research Library and Archives Sector

## Abstract

A study of the current provision of digitised 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 digitised 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 digitisation 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 co-ordinated and distributed, rather than centralised, one.

## Keywords

Digitisation, Research Libraries, Funding, National Strategy

## 1 Introduction

In the UK, it became clear in the 1990s that digital information would play a major role in higher education (Whitelaw & 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 digitising their collections. Libraries and archives are digitising 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 & Rieger 2000, p. 1). The commercial

---

<sup>1</sup> UK Data Archive is a centre of expertise in data acquisition, preservation, dissemination and promotion. <<http://www.data-archive.ac.uk/>>.

<sup>2</sup> 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.

<sup>3</sup> Higher Education Digitisation Service (HEDS) <http://heds.herts.ac.uk/> provides advice, consultancy, and a complete production service for digitisation and digital resource development and management to the higher education sector, museums, public and national libraries, archives and other not-for-profit organisations.

<sup>4</sup> The Resource Discovery Network (RDN) <<http://www.rdn.ac.uk/>> is a cooperative network consisting of a central organisation and a number of independent service providers called hubs offering subject portals).

sector is also involved in digitisation with publishers digitising 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 digitising research library collections. Generally digitisation 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) 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 digitisation of the cultural heritage. The European Commission report *Coordinating digitisation in Europe* (European Commission 2002) gives an overview of digitisation funding, collaboration and strategic initiatives in all EU countries. UNESCO is maintaining a register of some significant digitisation 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 digitisation strategy. The digitisation programmes that have been initiated in the UK have been funded by a number of different bodies, such as the JISC and various UK lottery-funding bodies. A recent survey<sup>6</sup> showed that the majority of digitisation 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. Digitisation strategies vary in terms of rationale for digitisation, the aims of projects and the selection criteria used. A recent report has highlighted issues related to digitisation at the national level, including risks of duplication, use of diverse standards and importance and opportunities of collaboration (National Audit Office 2004). However, there are some examples of successful collaborative digitisation 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/>).

---

<sup>5</sup> [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)

<sup>6</sup> Carried out by Bültman.

Given the fragmented nature of UK digitisation 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 digitised 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 digitised collections for researchers in the UK higher education sector. The study had a number of objectives:

- Produce a high level survey of digitised material, both already available and in the process of being created, held in UK research collections across all disciplines
- Survey demand for digitised material and identify gaps in existing provision
- Develop a mechanism for identifying future digitisation priorities
- Review funding structures and opportunities and assess possible ways of funding priority areas
- Recommend standards and formats for future digitisation projects
- Provide an outline action plan for a national digitisation 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 digitised 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 focussing on past, current and future digitisation projects, reasons for and against digitisation and experience of in-house and outsourced digitisation 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 digitised 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 & Muir 2002, Ayre & Muir 2004). Fifty-one replies were received from 47 institutions, these 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 digitisation 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 digitisation 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 digitisers, support services, publishers and funders on the future of digitisation in the UK. The possibility of developing a national strategy for digitisation 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 digitised content and the management of and provision of access to digital content in general is a fast moving area. New digitisation efforts have been announced since the research was carried out and new projects under programmes such as the JISC's Digital Repositories programme have begun. The results reported here represent a snapshot of activities at the time to research was carried out.

### **3 Digitisation Activities in UK Research Libraries and Archives**

Survey respondents were asked about their current and past digitisation activities.

Error! Objects cannot be created from editing field codes.

### **Figure 1 Has your institution digitised anything?**

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

### **3.1 Reasons for Digitising**

All of the thirty-four institutions with digitisation experience gave reasons for doing this. Respondents were able to select multiple responses and were asked to rank their responses.

Error! Objects cannot be created from editing field codes.**Figure 2 Reasons for digitisation (Ranked<sup>7</sup>)**

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 Digitising**

Eight of the thirteen institutions that had not been involved in digitisation gave reasons for this. Three digitisers also answered this question. Respondents were able to give multiple responses. The main reason for not digitising is a general lack of resources, mainly funding, but also equipment and expertise. Copyright restrictions (2) and low priority (2) were “other” reasons given.

Error! Objects cannot be created from editing field codes.**Figure 3 Reasons for not digitising**

Interviewees suggested similar reasons for not digitising material. Data ownership issues and prioritising digitisation of finding aids were also mentioned.

### **3.3 Materials Digitised**

---

<sup>7</sup> There is a count for non-ranked votes because a few submissions selected criteria but did not rank them.

The survey included a question on the types of material digitised by UK institutions. The responses revealed that still images and manuscripts were most frequently digitised. 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 digitised the entire range of materials shown in the figure below.

**Error! Objects cannot be created from editing field codes. Figure 4 Materials digitised**

### **3.4 Subject content of digitised resources**

From the desk research it was clear that a large proportion of digitised 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 digitisation 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.

**Error! Objects cannot be created from editing field codes. Figure 5 Subject areas of digitised materials**

Other subject areas include law and jurisprudence (2 respondents), genealogy and leisure pursuits. In one instance, the digitisation 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.

### **3.5 Selection criteria**

All the representatives of projects and digitising institutions interviewed approached digitisation in different ways and reported different experiences. Early digitisation projects were mainly small scale and involved one specific resource. As the digitisation of resources has grown, some organisations have established strategies and criteria for the selection of material to digitise, while others continue to digitise according to market need (particularly publishers) and user feedback. We found little data in the literature on user demand for

digitised 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 organisation 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/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. The funding bodies had varying selection criteria for funding digitisation 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 digitisation. All but one of the thirty-four digitisers 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 digitisation according to good collection management principles rather than responding to the aims and objectives of funding sources (three respondents).

Error! Objects cannot be created from editing field codes.**Figure 6 Selection criteria (Ranked<sup>8</sup>)**

Other selection criteria given by survey respondents included:

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

---

<sup>8</sup> There white sections represent non-ranked responses because a few submissions selected criteria but did not rank them.



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

Another organisation had established a priority list. Interviewees raised a number of issues in relation to the selection of material for digitisation. Some organisations had too much material and found it difficult to prioritise, 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 Digitised Material

Commercial companies in general charge for access to their digitised 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 digitised, 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 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.

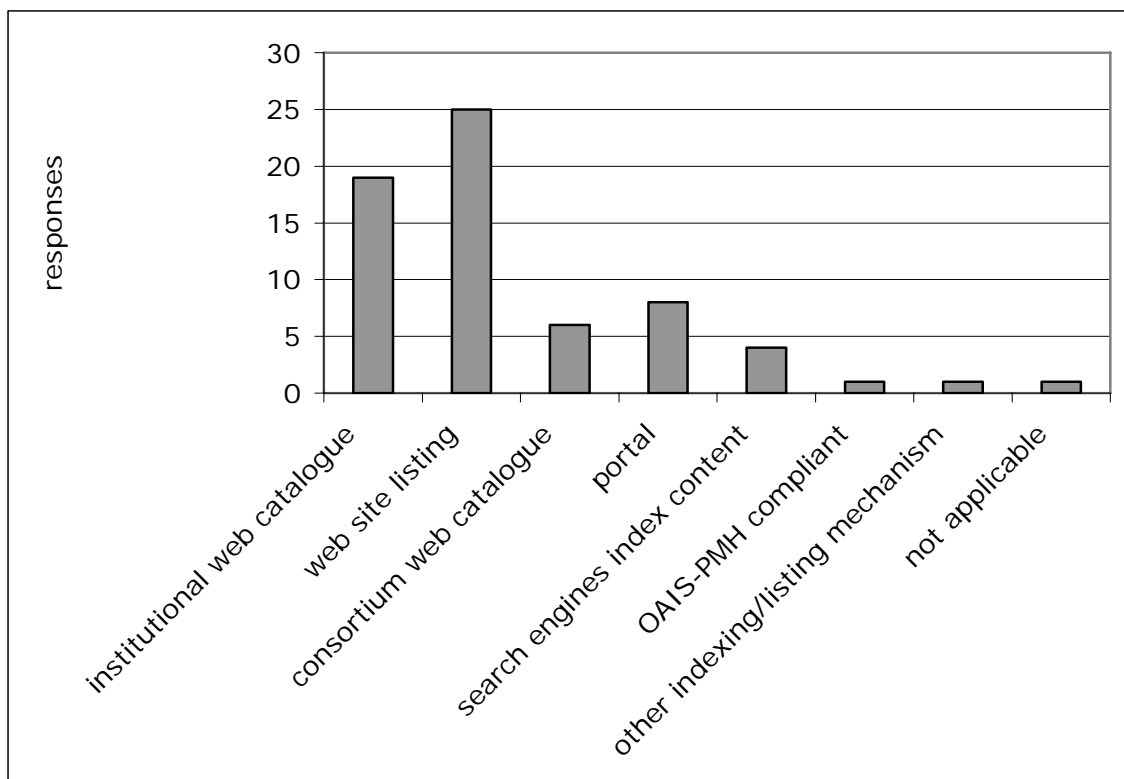
Survey respondents were asked about the accessibility of their digitised material. There were also questions on the existence of descriptive metadata for discovery purposes and access restrictions. Thirty-two out of thirty-four digitisers replied to a question on how freely accessible their resources are. Twenty-one responded that they provide access for free and eleven provide fee-based access. Additionally, six respondents selected both methods.

Error! Objects cannot be created from editing field codes.**Figure 7 Why are resources not freely available?**

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 that aren't 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)

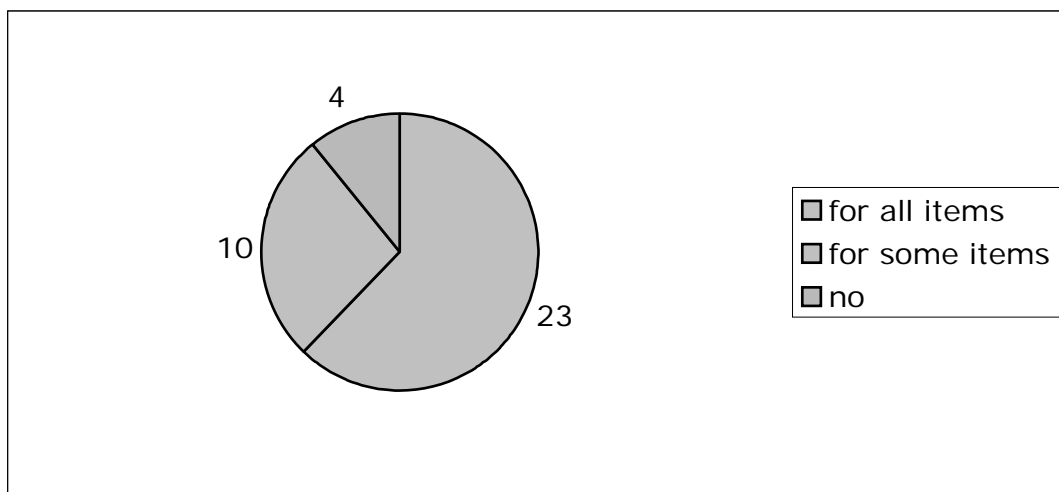
The interviews revealed some examples of library and archives charging models for access. For example, the National Archives of Scotland (NAS) allow free viewing of their digitised wills, but charge for downloads. This model is also applied by the National Library of Wales (NLW) and the Wellcome Trust to some of their material. This model is based on the charging model for reproduction services. The benefit of charging is seen as the potential for covering maintenance cost and further development of the project.



**Figure 8 Access to "published" collections**

There were thirty-five survey responses to a question on how resources available to external users, free or for a fee, are made accessible. Web site 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 thirty-four digitisers (as well as three non-or not-yet digitisers) answered a question on metadata. The aim of the question was to find out if metadata records were available for all digitised material. Around two thirds indicated that metadata was created for all digitised 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 3.10



**Figure 9 Existence of metadata records**

### **3.7 Management and Preservation of Digitised Material**

Interviewees thought that, in general, collection management is the responsibility of the organisation/institution that hosts the digitised material. Some funding agreements specify that long-term management of the digitised 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 digitisation projects to aid others in the planning of and management of projects. Interviewees found such information useful.

Interviewees felt that good project management is a vital component of all digitisation 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 digitisation project.

The long-term management of digitised collections was raised by those involved in digitisation 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 digitised resources. In the past, many projects only involved digitisation. 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 digitisation 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 digitised 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. Digitisers 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 digitised 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.

Error! Objects cannot be created from editing field codes.**Figure 10 Preservation measures**

Formats used by digitisers are discussed below. Survey respondents were also asked which technical preservation strategies they planned to use. Thirty-eight respondents answered this question, including all digitisers. Multiple responses were possible. Most respondents are willing to refresh media, a short-term preservation measure. No respondent chose to emulate obsolete technology. One respondent said that they have 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>9</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”, digitised exam papers “intended for a cohort of students” or considered as “access rather than preservation copies”.

### 3.8 Future Digitisation Plans

Survey respondents were asked about their future digitisation plans. One respondent did not answer this question, but a second response from the same institution did.

Error! Objects cannot be created from editing field codes.**Figure 11 More digitisable holdings?**

A large majority (41) of institutions have holdings that ought to be digitised in their opinion.

Respondents were asked whether they planned to digitise the material they had listed. Twenty-six (63%) stated that there were plans to digitise while fifteen respondents (37%) did not know.

Error! Objects cannot be created from editing field codes.**Figure 12 Reasons for digitising mentioned collections**

Three of the six “non-digitisers” that do not plan to digitise in future also hold collections that could be converted.

When asked why they thought this material should be digitised, forty-three respondents replied. These replies included respondents who did not plan to digitise the material themselves. Multiple responses were possible. Value for teaching and research were predominant reasons for digitising remaining collections, followed by uniqueness/rarity, cultural heritage and access considerations. This is different from the responses given to the earlier

---

<sup>9</sup> “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.” <http://lockss.stanford.edu/>, [accessed 14.02.2005].

question on reasons for digitising material. There, increasing access and reducing handling were the main objectives of digitisation activities.

Both “other” comments pointed to digitising 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]p/larger public libraries/archives and others.”

### **3.9 Priorities for Digitisation**

Interviewees commented that some institutions created digitisation strategies/programmes or lists of collections/items that could be digitised. 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 digitisation.

### **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 processing 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, p. 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 digitisation 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, e.g. 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**

Thirty-three of the thirty-four digitising respondents indicated which standards they used for digital files. 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.

**Error! Objects cannot be created from editing field codes.**

#### **Figure 13 Formats used for preservation or delivery**

More recent and rarely mentioned formats include 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) are 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 digitising organisations interviewed provided details of file formats used. Interviewees stated that the requirements for formats are 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**

Twenty-eight out of the thirty-four survey respondents that had digitised indicated the metadata schemes they use. 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 standardisation amongst publishers and digitisation services. It does seem clear that the choice of metadata format depends on what is being digitised and for what purpose. While the majority of organisations/projects interviewed manage and produce their own metadata, one used an outside organisation.

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

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 organisations 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 digitisation projects and resources. The possibility of automating metadata production was raised by one interviewee. The organisation 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, Spurgin & Crystal 2005).

### ***3.11.2 Classification Schemes Used***

Twenty-six out of the thirty-four survey respondents who had digitised material indicated the classification schemes used. Multiple answers were possible. Library of Congress Sub-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 specialised thesauri were mentioned. A



number of replies were not detailed enough (“thesaurus”, “subject headings”) to be meaningful.

Less than two thirds (22) of the thirty-four digitisers answered a question on unique identifiers. 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.

<b>Identifier Type</b>	<b>Number of Responses</b>
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

**Table 1 Object identifiers used by survey respondents**

### **3.12 Support Services for Digitisation**

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.

Digitisation 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 organisations. These include:

- Mailing lists
- Advice and expertise in all aspects of digitisation 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, digitisation, sustainability of

collections, including economic sustainability, digital preservation management, assisting with funding applications and conferences. While some organisations offer specific services, others offer assistance covering the entire process of digitisation. Some support services focus on specific disciplines or areas of digitisation, for example 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 are used and requested more often. These include 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 digitisation 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 includes 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), 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 digitisation. TECHDIS, CETIS, The Digital Curation Centre (DCC) is mainly concerned with born digital material, but the expertise gained by this new service could also benefit digitised 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.

Error! Objects cannot be created from editing field codes. **Figure 14 Sources of advice consulted**

Internal source 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 digitised, since most respondents have digitised 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 organisations such as OCLC, the Council on Library and Information Resources, the Digital Library Federation. Conferences, work shops and mailing lists were also mentioned as source of advice.

## 4 Funding for Digitisation Projects

There have been several studies on the costing<sup>10</sup> of digitisation projects. Costs for digitisation are significant and include: documentation and preparation, conversion costs, ensuring copyright status and rights clearance of material, equipment costs, human resources, and ongoing maintenance costs. From the research, it looks like 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 digitisation, while others do not specifically fund digitisation (these bodies are aware that digitisation is included in some of the project funding allocated, but don't fund pure digitisation projects). When projects/groups apply for funds to digitise resources/collections, certain bodies are generally the first port of call. The funding body selected usually depends on what is to be digitised, the subject area, the amount of funds required, and the target user community. Some interviewees named certain funding bodies as regular funders for their digitisation 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 funding bodies 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.

---

<sup>10</sup> For example see Lee (Lee 2001, Chapter 4) and Simon Tanner and Joanne Lomaz Smith (Tanner & 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).

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 digitisation 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 organisation requiring the funding, the material/resources being digitised and the target audience. A number of funding bodies reported collaboration with support services, either through funding a support service or through specific collaboration e.g., 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.

Error! Objects cannot be created from editing field codes.**Figure 15 Main funding source(s)**

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.

Error! Objects cannot be created from editing field codes.**Figure 16 External Funding Bodies**

Twenty-eight respondents, including all twenty-seven that selected external or combined funds from the previous question answered this question. Multiple responses were possible. Public sector funding is most frequent, particularly National Lottery funding. Respondents specified which trusts and other bodies had funded their digitisation efforts.

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)

**Table 2 Other funding sources**

Twenty-six of the twenty-seven respondents that indicated that external funding, alone or combined with internal funds, was the main source of funding (along with two others), responded to a question on the proportion of the digitisation 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, for most, external money makes up the bulk of the digitisation budget.

Error! Objects cannot be created from editing field codes. **Table 3 Percentage of external funding in "digitisation" budget**

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 are covered by the government, while project partners (and anonymous donations to these) cover 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 digitisation is the significant cost associated with digitisation. Another cost is the need for added functionality of resources to meet the increasing expectations of users. Many stressed that without significant external funding, digitisation and the management of those digitised 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 digitisation projects factor in sustainability of the resources. Some interviewees involved in digitisation projects felt that funding bodies needed to look at their strategies and provide funding for the preservation and maintenance of digitised material.

As the survey respondents, interviewees commented that their institutions held many more resources that should be digitised, 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 digitisation of material. Though viewed as important, many feel the future of digitisation and digitised 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 digitisation of resources and struggle with establishing strategies to manage priorities.

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

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 co-operative endeavour of a number of Scottish libraries<sup>11</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; digitisation is done by the participating institution. Access to material is chargeable. The (UK) National Archives is involved in a number of co-operative projects including *Moving Here*, an online service on migration to the UK over the last 200 years.

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

### 5.1 Partners in Cooperation

Thirty-seven respondents answered a question on cooperation partners. Multiple answers were possible. Other UK and overseas libraries, museums

---

<sup>11</sup> Glasgow Caledonian University, Glasgow City Libraries and Archives, Glasgow Colleges Group, University of Glasgow, University of Strathclyde.

and archives were the preferred partners. Some respondents had used digitisation bureaux and cooperated with commercial publishers.

Error! Objects cannot be created from editing field codes. **Figure 17 What partners would you work or have you worked with?**

(A/L/M = archive/library/museum)

- 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 digitisation service.

Error! Objects cannot be created from editing field codes.

**Figure 18 Nature of collaboration with commercial partners**

Nineteen respondents provided information on their commercial partners. Multiple responses were possible. The most frequent commercial collaboration is outsourced digitisation. One respondent commented that the nature of collaboration varied from project to project.

Error! Objects cannot be created from editing field codes. **Figure 19 In-house versus outsourced digitisation**

Respondents were asked an explicit question on whether they carried out digitisation in-house or whether they outsourced this activity. Thirty-seven respondents answered this question. When asked for reasons for these decisions, the following were given.

Reasons given in favour of outsourcing	Reasons given in favour of in-house digitisation
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 digitisation” (1)</li> </ul>

**Table 4 Reasons outsourcing or digitising in-house**

Cost is a factor for in these decisions; this may be depend on the materials to be digitised.

Error! Objects cannot be created from editing field codes.**Figure 20 Reasons for cooperation**

This question was aimed at digitisers that had cooperated and received thirty-three responses. 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” digitisation experiences.

Error! Objects cannot be created from editing field codes.

**Figure 21 Quality control procedure in collaborative projects**

A question on quality control procedures received thirty-four responses. Nineteen came from respondents who had already cooperated and fourteen from other respondents. Multiple responses were possible. Most of the respondents relied on in-house checking, although some did use external companies and automated checking. “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 applicable” responses came from institutions that only had “solitary” digitisation experience.

## **6 The Future of Digitisation**

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

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 digitisation activities. They see the digital realm as the future but recent developments in the sector, for example, the Open Archives Initiative and Google Print, infringe on the publishers' traditional roles. A number of libraries have successfully run digitisation programmes which to some extent supersede publisher's digitisation 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 organising 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 recognise 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 digitisation in the UK was a cause for concern. One interviewee summed up the present approach to digitisation:

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 a collections survey and then think about what it is that from that to, to digitise based upon an analysis of their user needs. Who are their user communities?  
(Digitisation service)

Some digitisers 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 digitisation 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 co-ordination, including standards, selection criteria, funding allocations, and co-operation. 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 digitisation 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. They would be achieved by having clear guidelines.

Another interviewee said:

I think you've gone a step on whereas 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, so how on earth you think you can shape what people are going to be doing and you don't 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 digitisation. One digital library expert thought that while it should be co-ordinated, it should be distributed. It is clear from discussion with various players that there is a need for co-ordination in digitisation activities, even though they do not all agree on how this should be done.

## **7 Conclusions and Recommendations**

It is clear from discussions with the various players in digitisation that there is a need for co-ordination in digitisation activities, even though they do not all agree on how this should be done. This would assist in filling gaps in provision. A UK-wide strategy could 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. 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 organisations that could lead this strategy are the Research Libraries Network and the JISC. However, other organisations, 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 coordination of digitisation 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 digitised material, particularly in arts and humanities. These results highlight the need for a co-ordinated 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 co-ordinated.

## **7.2 Priorities for Future Digitisation**

The research libraries surveyed for this study provided some information on collections that remain to be digitised. 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 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.

## **7.3 Identification of Existing Digital Collections**

The survey indicated some issues in the creation of metadata records for digitised material. It seems that in some cases, records do not exist for the originals and metadata creation for these is a higher priority than digitisation. Metadata creation is an expensive part of the digitisation 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 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.

Information on digitised 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 digitised resources could facilitate the analysis of gaps in provision. 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. 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 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/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 digitised 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 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.

#### **7.4 Standards and Formats and Collection Management Issues**

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

The survey carried out for this study showed that digitisers 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 digitisation, 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 digitised resources, both in terms of 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 digitisers. Digitisers 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 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.

## **7.5 Funding Opportunities**

The study found that lack of funding was a major deterrent to digitisation. At the moment there are a plethora of funding bodies and opportunities and there is a hint from the study that organisations planning to digitise have to spend time identifying and exploring funding opportunities. It would seem sensible to have a more co-ordinated 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 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, but it should be possible to improve co-ordination.

The Google initiative is currently an unknown quantity, but could have a major impact on business models and research library interest in digitisation. Publishers who participated in this study are clearly concerned about the implications of Google 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 for libraries, 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 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.

## 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.

## References

- Arts and Humanities Data Service (2003) *Introduction to creating digital resources*. <http://www.ahds.ac.uk/creating/information-papers/creating-introduction/index.htm>, [accessed 21.12. 04].
- Berglund, Ylva *et al* (2004) *An investigation into free ebooks: final report*. <http://ahds.ac.uk/litlangling/ebooks/report/FreeEbooks.html#ebooks-div-id2669731>, [accessed 21.12.2004].
- European Commission (2003) *Coordinating digitisation in Europe: progress report of the National Representatives Group: coordination mechanisms for digitisation policies and programmes*. <http://www.minervaeurope.org/publications/globalreport/globalrep2002.htm>, [accessed 03.09.2004].
- Greenberg, J., K. Spurgin & A. Crystal (2005). *Final report for the AMeGA (Automatic Metadata Generation Applications) project*. [http://www.loc.gov/catdir/bibcontrol/lc\\_amega\\_final\\_report.pdf](http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf), [accessed 29.4.05].
- Hughes, Lorna M. (2004) *Digitizing collections: strategic issues for the information manager*. London: Facet.
- Kenney, A.R. & O.Y. Rieger, eds (2004) *Moving theory into practice: digital imaging for libraries and archives*, 2000, <http://www.minervaeurope.org/publications/globalreport/globalrep2002.htm>, [accessed 03.09.2004].
- Lee, S.D. (2001) *Scoping the future of the University of Oxford's digital library collections: final report*. <http://www.bodley.ox.ac.uk/scoping>, [accessed 05.09. 04].

National Audit Office (2004) *The British Library - providing services beyond the Reading Rooms*. [http://www.nao.org.uk/publications/nao\\_reports/03-04/0304879.pdf](http://www.nao.org.uk/publications/nao_reports/03-04/0304879.pdf), [accessed 03.09.2004].

Tanner, Simon & Joanne Lomax Smith (1999) *Digitisation: How Much Does It Really Cost? Paper for the Digital Resources for the Humanities 1999 Conference, Sept. 12-15*, <http://heds.herts.ac.uk/resources/papersl.html>, [accessed 6.9.04]

Whitelaw, A. & G. Joy (2000) *Summative evaluation of Phase 1 and 2 of the eLib initiative: final report*. [www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/elib-fr-vl-2.pdf](http://www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/elib-fr-vl-2.pdf), [accessed 03.09.2004].