JISC

e-Resources

# Digitisation in the UK

## The case for a UK framework

# Digitisation in the UK – the case for a UK Framework

A report based on the Loughborough University study on Digitised Content in the UK Research Libraries and Archives Sector commissioned by JISC and the Consortium of Research Libraries (CURL)

JISC

consortium of curl research libraries in the British Isles

The full Loughborough University study can be found at:
**www.jisc.ac.uk/digitisation**


**Acknowledgements**

| | |
|---|---|
| David Dawson | Head of Digital Futures, Museums, Libraries and Archives Council |
| Stuart Dempster | Programme Manager, JISC |
| Robert Kiley | Head of Systems Strategy, The WellcomeTrust |
| Dicky Maidment-Otlet | Communications Manager, JISC |
| Richard Ovenden | Keeper of Special Collections and Western Manuscripts at the Bodleian Library |
| Neil Parkinson | Author |
| David Pearson | Director of Research Library Services, University of London |
| Jean Sykes | Librarian and Director of IT Services, London School of Economics |
| Simon Tanner | Director, King's Digital Consultancy Services |


This report is available in alternative formats which can be found at: www.jisc.ac.uk/digitisation

# Contents

# 1.0 Executive summary

*A wealth of material in museums, libraries, archives and journals remains undigitised, despite the pressing need to sustain the momentum, to continue to create resources of increasing value and comprehensiveness for the end user.*

**The Loughborough study makes three clear recommendations:**

- Establish a UK framework for digitisation
- Coordinate existing services
- Investigate users' needs

In just a handful of years, an enormous amount of richly detailed and flexible digital material has been amassed in the UK as technology has expanded to make it possible: a conservative estimate suggests £130 million of public money has been spent on the creation of digital content since the mid-1990s. Nevertheless, this growth has been as unstructured as it has been phenomenal, and the material has accumulated in the absence of a UK framework for digitisation to advise on content, standards and sustainability, rather than in response to one. Digital programmes have sprung up in piecemeal fashion, dictated by individual circumstances, and executed locally. And the picture has been confused by the proliferation of standards and formats for digital surrogates, the varying approaches to accessibility, and the considerable number of advisory bodies which encourage take up of one scheme over another, despite minimum standards being outlined in JISC's *Information Environment Architecture Standards*[1] and MINERVA's *Technical Guidelines for Digital Cultural Content Creation Programmes*[2]. Moreover, digital projects have tended to be driven by supply rather than demand, spurred by opportunity instead of actual need.

A wealth of material in museums, libraries, archives and journals remains undigitised, despite the pressing need to sustain the momentum, to continue to create resources of increasing value and comprehensiveness for the end user. The very existence of powerful search facilities is changing users' behaviour and expectations. Future digitisation programmes must respond to this and need to be more clearly informed by researchers' needs, to respond to actual rather than theoretical demand.

1 JISC Information Environment Architecture Standards
**www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards**

2 MINERVA Technical Guidelines for Digital Cultural Content Creation Programmes
**www.minervaeurope.org/publications/technicalguidelines.htm**

There are already plenty of bodies which take a view on digitisation of research library material in the UK, including the Joint Information Systems Committee (JISC), the British Library, The National Archives, the Research Information Network (RIN), the Research Councils UK (RCUK) and the Museums, Libraries and Archives Council (MLA), yet each has different constituents, with none fully able to take a UK-wide overview. Outside the UK there has also been activity in the digitisation field, particularly the EU funded programmes, such as Minerva[3] which, among other aims, is attempting to coordinate digitisation activities, and more recently the European 7th framework[4] and i2010[5] which includes the EU digital library[6] proposal. The involvement of commercial publishers and funding bodies adds to the complexity of the situation. Perhaps most significantly, the announcement of the Google Print Library Project[7] initiative to digitise huge quantities of books from some of the world's leading libraries means the time could not be more opportune for a considered and dynamic public sector response. While only in its earliest stages, Google Print and the Open Content Alliance[8] are changing the world of information provision and portend a revolution in which the sector needs to participate fully.

*The establishment of a UK e-Content Strategy of which a Digitisation Framework would form a part, would contribute to the EU e-Content programmes on Digital Libraries proposal under the i2010 strategy.*

In 2005, JISC and the Consortium of University Research Libraries[9] (CURL) commissioned Loughborough University to undertake an in-depth investigation into the current state of digitisation in the UK, and this document draws on its findings. It charts how far we have come to date and makes the case for strengthened coordination and the establishment of a UK framework to ensure future projects are better executed, more sustainable, and respond directly to the needs of the research community. The results would benefit all participants and stakeholders: the higher the quality and comprehensiveness of digitised resources, the better the value for everyone, both in financial and academic terms.

Loughborough's research uncovered deep fragmentation in all components of the digitisation infrastructure: the records of available material, the provision of e-resources for different disciplines, the metadata and standards used, the advisory and support services, the availability of funding, the differing priorities of funders, and variable hosting, delivery and authentication methods. Yet the very interconnectedness of the elements of the digitisation process, where each impacts on the other, makes it both easier and more essential to place them within a framework which can make formal links that resonate across all operations. All shortcomings identified in Loughborough's study can therefore begin to be addressed, from inadequate metadata to lack of collaboration, by uniting the various sectors through a UK framework for digitisation. A UK-wide strategy would assist in filling gaps in provision, cut across the efforts of individual funders and digitising organisations, reduce overlaps between support services and assist in the provision, take up and use of resources. Fears that any such 'nationalisation' might stifle local innovation can be allayed by emphasising the flexible nature of the framework we envisage; one which would issue clear guidelines rather than prescriptive demands, which would draw

3 MINERVA programme
**www.minervaeurope.org/whatis.htm**

4 EU 7th Framework
**http://europa.eu.int/comm/research/future/documents_en.cfm**

5 EU i2010 programme
**http://europa.eu.int/information_society/eeurope/i2010/index_en.htm**

6 EU digital library
**http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1202&format=HTML&aged=0&language=EN&guiLanguage=en**

7 Google Print Library Project
**http://print.google.com/googleprint/library.html**

8 Open Content Alliance
**www.opencontentalliance.org**

9 Consortium of University Research Libraries
**www.curl.ac.uk**

up 'gold standards' to be regularly reviewed. Such a framework, then, should be coordinated and distributed, rather than centralised, and ensure effective networking of expertise across different sectors.

This report considers in turn the current availability of digitised resources, the support and advisory systems in place, and the operations and priorities of funding bodies. It promotes a vision to safeguard the future of digital resources by placing the process and the results within clear parameters, regardless of the unpredictable metamorphoses of the technology and business models to come.

## 1.1 Loughborough's key findings

Loughborough's key findings can be summarised as follows:

- There is already a wealth of digitised material in the UK and the investment in digitisation projects has amounted to £130 million of public money over 10 years
- Many types of material are now online, from manuscripts to sound and video files, although there has been an emphasis on archives and manuscripts relating to the arts and humanities, and social sciences
- Significant gaps in provision remain in many disciplines, including those seemingly well served
- There is no UK register to map individual digitisation projects and therefore no authoritative resource to aid discovery and prevent duplication
- There are many sources of guidance, some with overlapping remits. Different support services seldom work together and there is scope for collaboration and consolidation here
- The proliferation of standards is beginning to give way to a common consensus regarding metadata schemas and file formats. However, there is no overarching view on standards; guidelines vary according to the stakeholders concerned and the support services consulted
- Metadata creation is becoming a more urgent priority than digitisation itself in some cases. It is a crucial but costly part of the process and its creation must be costed into funding bids for projects, and the implications understood by funding bodies
- Those involved in digitisation projects view the current fragmented funding structure as unsatisfactory, especially the way digitisation has been funded largely on a short-term 'project' basis
- The different funders do not tend to work together to provide joint funding for digitisation projects
- Funding bodies are concerned about issues of long-term preservation and stability of resources; as a result, some are becoming more strategic and making consideration of these areas a condition of funding
- Innovative cross-sectoral collaborations are changing the models of digitisation projects; the Google initiative will hasten this development
- Any future developments involving commercial partners need to avoid the appearance of the library community (and others) being sold its resources at a premium

## 1.2 Recommendations

We are making three clear recommendations on the basis of these findings as follows:

### 1.2.1 Establish a UK framework for digitisation

There is a pressing need for a UK task force which can stand back and see the holes in the patchwork of digital projects, set clear guidelines on standards and take a coherent line on access mechanisms as part of a UK e-Content Strategy. We need a framework not simply to respond to the powerful drivers for digitisation (access, demand and preservation), but because these drivers, and the response to them in the forms of funding, projects and services, are not being coordinated effectively. There are strong currents of goodwill, enthusiasm for digital projects, and a desire to bring increasing volumes of material online, from users and providers alike. These should be tapped by those active in the digitisation or acquisition of e-Content within the UK.

### 1.2.2 Coordinate existing services

Part of the framework's remit would be a much-needed coordination of support and advice on all related issues – from guidance on standards to a comprehensive listing of all projects undertaken – and provision of a portal for information on funding streams. This should map to the proposed EU 7th Framework 'centres of competence'. A single access point to the range of services offering guidance on standards would help foster interoperability and sustainability, and would also benefit funders by enabling them to derive better value from their investment. A comprehensive resource listing all digitised projects, meanwhile, would chart the landscape, promote individual resources and prevent unnecessary duplication. Funding bodies (and the bidders), would benefit from coordination through a single point and this may facilitate better cooperation between funders.

*A single access point to the range of services offering guidance on standards would help foster interoperability and sustainability, and would also benefit funders by enabling them to derive better value from their investment.*

### 1.2.3 Investigate users' needs

Future developments in digitisation need to respond more directly to user demand rather than supply, yet researchers' and other user needs (including searching behaviours) are still not fully understood. Insight is particularly lacking into requirements in the science and social science fields. Various channels, including subject associations, Learned societies and academies, can be used to gauge need and focus the general goodwill and enthusiasm from the research community into a targeted programme that fulfils specific needs and plugs information gaps. The British Academy *E-resources for Research in the Humanities and Social Sciences* report[10] into researchers' needs could be supplemented by further investigation by Research Councils, while the Research Information Network (RIN)[11] has a key role to play in investigating user requirements. Once surveys have been undertaken, the results should be shared with JISC and CURL, and the response coordinated and used to inform policies, procedures and strategies.

10 British Academy '*E-resources for Research in the Humanities and Social Sciences*' report **www.britac.ac.uk/reports/eresources/index.html**

11 Research Information Network **www.rin.ac.uk**

# 2.0 Introduction

The capacity for new methodologies in education and research, and the availability of resources, have been transformed by a wave of digitisation programmes, bringing a wealth of heritage material online in the last decade. Digital resources are now available to enrich educational experiences at all stages of the learning journey, from formalised lessons in the primary school classroom to the lifelong learner's casual browsing at home, from undergraduates embedding film clips in their electronic projects to the purposive search of the professional researcher. An enormous amount of richly detailed and flexible digital material has been amassed in a small number of years as technology has expanded to make it possible: a conservative estimate suggests £130 million of public money has been spent on the creation of digital content since the mid-1990s.

Nevertheless, this growth has been as unstructured as it has been phenomenal, and the material has accumulated in the absence of a UK framework for digitisation to advise on content, standards and sustainability, rather than in response to one. The combination of advances in Information and Communications Technology (ICT) and new funding opportunities has guaranteed the continuous development of digital programmes, but these have sprung up in piecemeal fashion, dictated by individual circumstances, often small in scale and carried out in isolation. The approach to digitisation in the UK seems systematically scattershot: facilitated and executed by various stakeholders with overlapping remits, with opportunistic projects springing up in response to available funding, often managed and undertaken with only one eye on the need to produce a future-proof resource. The proliferation of standards and formats for digital surrogates, and the considerable number of advisory bodies which encourage take up of one scheme over another, has confused matters. Different groups involved in digitisation are often unclear on how best to apply metadata, for example, and there are misunderstandings over who has ultimate responsibility for the sustainability and long-term management of projects and their resulting e-resources.

The UK's impressive accumulation of a body of digital material to date is only a first step, a product born out of the heat of new technology. There remains a wealth of material in museums, libraries, archives and journals currently undigitised, despite the pressing need to sustain the momentum, to continue to create resources of increasing value and comprehensiveness for the end user. In fact, the digital revolution is *accelerating*, as the faltering and sluggish experience of the dial-up connection is replaced with the seamless, quick-fire functionality of broadband, enabling a far richer learning experience, faster searches, and a wealth of media accessible through common interfaces. The physical world of monographs on shelves is now complemented by instantaneous desktop access, and the very existence of powerful search facilities is changing users' behaviour and expectations. Future digitisation programmes must respond to this and need to be more clearly informed by researchers' needs and demands. Librarians have an excellent track record in identifying useful material for digitisation and it is in their nature to take the long-term view. But as the indiscriminate digitisation of *all* material is not a possibility, there is a clear need to gather more researcher and user input, for digitisation programmes to respond to actual rather than theoretical demand and contribute in a targeted way to the development of the UK knowledge economy.

*There is a clear need to gather more researcher and user input, for digitisation programmes to respond to actual rather than theoretical demand and contribute in a targeted way to the development of the UK knowledge economy.*

There are already plenty of bodies that take a view on digitisation of research library material, including the British Library, The National Archives, RIN, JISC and the MLA, yet each has different constituents and agendas, none with the remit to take a UK overview. A recent National Audit Office report[12] confirms there is no UK policy on the selection of material for digitisation or the creation of a UK-wide digital library. The picture is further confused by the involvement of commercial publishers and the continuous evolution of business models for projects. Some digitisation funding is internal, but mass digitisation tends to be externally funded. Moreover, it is axiomatic that digitisation simply does not happen without significant, usually external, funding, so funding bodies have a key role to play in current and future developments, helping define the context in which policymakers operate. The recent announcement of the Google Print Library Project to digitise huge quantities of books from some of the world's leading libraries, including the Bodleian, means the time could not be more opportune for a considered and dynamic public sector response to this radical development. While only in their earliest stages, the Google Print Library Project and the recently announced Open Content Alliance are changing the world of information provision and portend a revolution in which the sector needs to participate fully. Rather than feeling its thunder has been stolen, the library and information community needs to act soon, in a coordinated way of size and significance, to capitalise on the boost provided by the initiatives of Google and Open Content Alliance. Although much digital activity is likely to continue within the private sector, the figure of £130 million spent to date signifies the likelihood that the involvement of the *public* sector is far from over. There has never been a better time to establish the parameters of a UK digitisation framework and guide future developments accordingly.

**12** National Audit Office '*The British Library - Providing services beyond the reading rooms*' report
**www.nao.org.uk/publications/nao_reports/03-04/0304879.pdf**


Despite the sums of money involved only a fraction of analogue collections have been digitised and made available online to date. If the UK is to maintain its position as a key player within the international research community and build upon its commitment to life long learning and the knowledge economy then effective public investment should be continued in order to maximise opportunities and leverage in the creation of viable new e-Content. Many of these analogue collections are not viable for the commercial sector to fund. A mixed economy of private and public funding is therefore called for, especially if the supply of primary and secondary research material is to be maintained.

*The results would benefit all participants and stakeholders: the higher the quality and comprehensiveness of digitised resources, the better the value for everyone, both in financial and academic terms.*

This document draws on the findings of an in-depth investigation into the current state of digitisation in the UK, commissioned by JISC and the Consortium of Research Libraries (CURL) and carried out by Loughborough University. Loughborough surveyed the constituent parts of the digitisation process – libraries, digital service providers, funding bodies and commercial publishers – through questionnaires and interviews, to gauge their priorities and concerns, mapping the current disjoints and the scope for future cooperation. This summary charts how far we have come to date and makes the case for strengthened coordination and the establishment of a UK framework to ensure future projects are better executed, more sustainable and respond directly to the needs of the research community. The results would benefit all

participants and stakeholders: the higher the quality and comprehensiveness of digitised resources, the better the value for everyone, both in financial and academic terms.

The recent creation of strategic initiatives, such as the RIN, the People's Network [13] and the Common Information Environment (CIE)[14], is beginning to address the mismatches among digitisation projects. But their existence makes the opportunity for action even more timely. The RIN has the potential to become a valuable conduit for researchers, to promote itself as a resource for their benefit, and to feed back a clearer picture of user needs, thereby influencing digitisation programmes from the perspective of the individual user. The New Opportunities Fund (NOF)-funded and MLA-managed People's Network, meanwhile, stands to increase access to digital resources by connecting all public libraries to the Internet, thereby increasing the potential for collaborative involvement in digitisation projects. The CIE is creating an open environment for information and resources gleaned from the museum, library, archive, health and education sectors. This is a direct response to the piecemeal creation of digital content, and attempts to ensure the end user can discover information and material which can be used and re-used according to their needs. The CIE is thus directly addressing a perception that digital content in its current infrastructure fails to reach the right people at the right time.

But these consolidations in themselves, while welcome, are not enough. As Loughborough's research shows, there is deep fragmentation in all components of the digitisation infrastructure: the records of available material, the provision of e-resources for different disciplines, the metadata and standards used, the advisory and support services, and the availability of funding and priorities of funders. While the width of these fissures may initially seem unbridgeable, the very interconnectedness of the elements of the digitisation process, where each impacts on the other, makes it both easier and more essential to place them within a framework that can make formal links that resonate across all operations. Funding could be tied to a guarantee of sustainability and maintenance of certain standards, as nof-digitise[15] were for example, or bids prioritised according to their ability to fill gaps in provision.

This document considers in turn the current availability of digitised resources, the support and advisory systems in place, and the priorities of funding bodies. It maps the area of each, charts the shortfalls and mismatches, and makes recommendations for consolidation and practical cooperation across the board. It promotes a vision to safeguard the future of digital resources by placing the process and the results within clear parameters, regardless of the unpredictable metamorphoses of the technology and business models to come.

**13** People's Network **www.peoplesnetwork.gov.uk**

**14** Common Information Network **www.common-info.org.uk**

**15** nof-digitise **www.mla.gov.uk/action/pn/nof-digitise.asp**


# 3.0 Digitised resources in the UK: supply and demand

## 3.1 Success stories

**There is already a wealth of digitised material in the UK and the investment in digitisation projects amounts to £130 million of public money in 10 years**

**There is no UK-wide register to map individual digitisation projects. The introduction of such a system would keep track of resources, aid discovery and prevent duplication**

Loughborough's conservative estimate of an investment of £130 million of public money in digitisation projects in the past decade is a useful figure, which helps gauge the abundance of digital surrogates brought online in the early, experimental days of the technology. The actual expenditure may be even higher as individual projects have not been aggregated in a central directory.

The wealth and eclecticism of material digitised is hinted at by the large sums of money involved. While archives, manuscripts, artworks and photographs are typical materials for digitisation, the variety overall is as great as the holdings of any contributing museum or special library, extending to monographs, artefacts, maps, newspapers, government publications, grey literature, moving images and music (the last named has been digitised in the form of sheet music *and* digital sound files).

Current high-profile projects give a flavour of the state of digitisation in the UK at this time. For example, JISC is supporting a range of comprehensive, innovative resources as part of its Digitisation Programme[16], including Newsfilm Online[17], which brings together 3,000 hours of digitised news footage from the ITN archives and makes the files accessible and editable through the desktop, while British Newspapers 1800–1900[18] is creating searchable surrogates of complete runs of major newspapers published throughout the 19th century.

16 JISC Digitisation Programme
www.jisc.ac.uk/digitisation_home.html

17 Newsfilm Online Project
http://temp5.bufvc.ac.uk/newsfilmonline/public_html/index.php

18 British Library Newspapers 1800-1900 Project
www.bl.uk/collections/britishnewspapers1800to1900.html

Such visible examples only give a taster of the diversity of resources already available or currently in development. Others include cross-sectoral collaborative projects, international initiatives, and ventures with commercial publishers. Among the cross-sectoral collaborations is Moving Here[19], an online resource on the topic of migration to the UK in the last 200 years and involving 50 partners from museums, libraries and archives. The service, spearheaded by the National Archives, provides 200,000 digital objects, drawn from the partners' collections, and includes audio-visual files, maps and images. Similarly wide ranging is the Scottish Cultural Resources Access Network (SCRAN)[20], a charity with finance from the Scottish Executive, which provides educational access for schools and libraries to digital materials representing Scottish culture. SCRAN has cooperated with 450 institutions in the UK, including libraries, museums and archives.

As digital resources can transcend UK boundaries, the list does not stop at UK-based projects, and there are plenty of international resources which are available in the UK, swelling the volumes of material on offer, but curated abroad. JSTOR[21], for example, created by the Andrew W Mellon Foundation[22], is an independent, not-for-profit entity, building a trusted digital archive of important scholarly journals (nearly 450 in total), and involving many international higher education (HE) institutions and a large number of scholarly, society and commercial publishers as participants.

Private sector involvement, often in the form of public–private collaboration, is also part of the landscape, and publishers including Reed Elsevier and Routledge are actively involved in the digitisation of backlists. Recent projects include Early English Books Online (EEBO)[23] and the Text Creation Partnership (TCP)[24], which in the UK is a collaborative partnership between ProQuest, JISC and the British Library.

Nevertheless, it is difficult to map all digitisation projects with authority as so many have been undertaken on an ad hoc basis and have not been recorded in a central register. It should be noted that the Multilingual Inventory of Cultural Heritage in Europe (MICHAEL) project[25] aims to reveal all the digitised resources within the cultural heritage sector; however, this does not map current and future digitisation projects and initiatives. There is a clear need for a better mechanism for identifying relevant projects and collections in order that they can be fully indexed and described and, by extension, discovered and used. Any such system will need to be simple and inexpensive, to facilitate universal contribution, and could also be a requirement of receiving the funding.

**19** Moving Here **www.movinghere.org.uk**

**20** SCRAN **www.scran.ac.uk**

**21** JSTOR **www.jstor.org**

**22** Andrew W. Mellon Foundation **www.mellon.org**

**23** EBBO **http://eebo.chadwyck.com/home**

**24** EBBO-TCP **www.odl.ox.ac.uk/eebo/eebo.html**

**25** MICHAEL Project **www.michael-culture.org/index.html**

## 3.2 Demand for digitised materials

The relationship between availability of digital material and demand is axiomatic. As technology increases ease of access to a given resource, its user base inevitably grows correspondingly, as the British Academy has pointed out in its recent study *E-resources for Research in the Humanities and Social Sciences* (2005).

Indeed, one of the two key drivers for digitisation is the ability to increase access to material via an enhanced and innovative form of resource discovery. As digital resources become more commonplace, they set a precedent which future developments must acknowledge and exceed. Desktop access is now the norm, and the availability of material online is changing the way research is being done. New forms of access, such as keyword searching, the complex indexing and cross-referencing that electronic cataloguing allows, and the facilities of onscreen comparison of physically disparate material, produce an increasing demand not only for the availability of digital material itself but for an instantaneous and satisfactory response to search terms. At its worst, this can foster an attitude that 'if it's not on the Internet, it doesn't exist'; at best it acts as a powerful driver to bring increasing quantities of material online to keep up with demand, breathing new life into heritage material and opening it up to new lines of research.

The second key driver for digitisation is the ease with which fragile and inaccessible materials can be preserved and conserved by transferring the burden of use to a high-specification digital surrogate. Turning the Pages[26], a commercial application devised by Armadillo Systems[27] and promoted by the British Library, for example, has allowed the general public and researchers alike detailed interactive exploration of digitised rare books in quantities which would be unthinkable in the real world.

These drivers are self-evident to anyone involved in research and digitisation, but it is unfortunate that their influence on projects is ad hoc and localised, with no one, to date, able to produce a formalised response to these needs, thereby ensuring that demand is matched to supply.

**26** Turning the Pages Project **www.bl.uk/onlinegallery/ttp/ttpbooks.html**

**27** Armadillo Systems **www.armadillosystems.com/**

## 3.3 Types of material

**Many types of material are now online, from manuscripts to sound and video files, although there has been an emphasis on archives and manuscripts relating to the arts and humanities, and social sciences**

**Significant gaps in provision remain in many disciplines, including those seemingly well served**

According to Loughborough's findings, manuscripts and archives are the most frequently digitised type of material in libraries and archives (the capture procedure is easy and dramatically improves access to materials), although other types, including artefacts, have also been digitised. The bulk of the material available electronically is most relevant to the fields of arts and humanities, and the social sciences, although there are still significant gaps in provision in many disciplines, including those seemingly well served. The gap in the sciences may reflect different research methodologies between hard science and the humanities, with scientists preferring current journals and articles which are often born digital. Nevertheless, the issue of digitisation of back-runs of journals is relevant to researchers in the sciences.

In addition, many UK digitisation projects have focused on specific themes or topics, or provided a taster for special collections overall, with digitisation effectively becoming a marketing and publicity tool.

## 3.4 Reasons for digitising

**It will never be possible to digitise everything, so future developments must respond to researchers' needs, which are currently poorly understood. Further investigation into their specific requirements is urgently required**

Both drivers identified above (adding to the resources and preserving the originals) come into play for individual digitisation projects. There is often a question of whether material should be digitised just because it is rare or valuable or whether there should be demonstrable need. Librarians and archivists have a good track record in making sound judgements as to what content will be useful, and tend to take a long-term view. Nevertheless, digitisation projects are seldom initiated as a direct response to researchers' needs, and this is a matter for concern. Indeed, little has been written on the current and future requirements of researchers for digitised material, although the Research Support Libraries Group's (RSLG) *Researchers' Use of Libraries and other Information Sources* (2001)[28] provides useful insights, as does the aforementioned British Academy report. RSLG found that biological and medical research relies heavily on e-journals and active full-text databases,

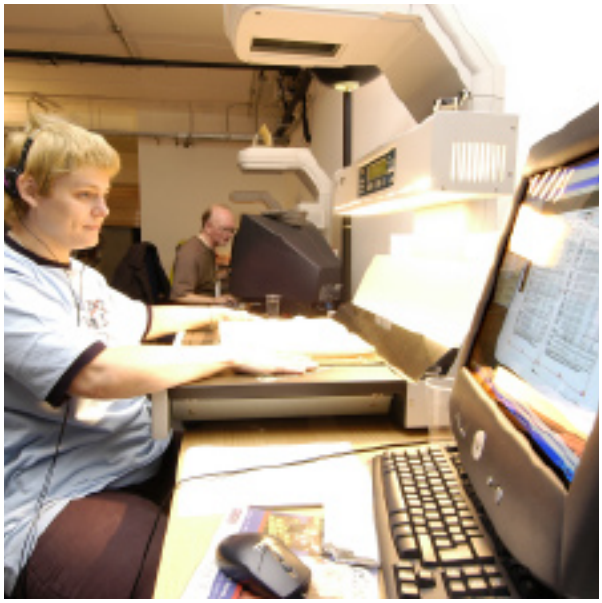
28 RSLG Researchers' Use of Libraries and other Information Sources
www.rslg.ac.uk/research/libuse


while arts and humanities researchers prefer physical access and value serendipitous browsing. Pure scientists prefer focused searches. Another finding was that non-conventional research resources, such as moving images and broadcast materials, are rarely used. This may be because they are not currently well established and it remains to be seen whether the creation of resources such as Newsfilm Online will influence and alter research methodologies by providing new forms of digital source material.

Loughborough's questionnaire survey of 34 institutions with digitisation experience found that 'improved access to unique material' was by far the most commonly cited reason for undertaking such projects. Selection criteria, however, vary among institutions; some organisations have established strategies while others digitise according to market need and user feedback. The most frequent response in considering past projects was 'relevance to aims and objectives of the institution', reinforcing the sense that digitisation continues to take place at an ad hoc, localised level, despite the boundless reach of online resources and the potential to link projects in order to build a picture that transcends individual institutions. Selection criteria are also influenced by a wide range of variables which, again, vary from one institution to the next. These might include user feedback, focus-group opinion, response to market trends, and popularity of courses in a given discipline for which relevant material can be digitised. In some cases, collaboration between publishers, libraries, academics and curators has determined the content to be digitised.

*The most significant barrier to digitisation was, inevitably, lack of funding.*

Respondents cited 'value for research and teaching' as the most prominent reasons for digitising material in the future. The most significant barrier to digitisation was, inevitably, lack of funding. Put simply, digitisation cannot happen without significant financial support, usually from an external body. Many of Loughborough's interviewees pointed out that their institutions held many more resources that should be digitised and made available to the research community, but that this would only happen if further funding could be secured. The survey found no predetermined or generally accepted methodology for developing future digitisation projects, although some institutions have taken practical steps such as establishing appropriate posts or ring-fencing funding for opportunities that might arise.



# 4.0 Standards, formats and guidelines

**Too many standards?**

**All materials should be digitised to the highest specifications for flexibility and sustainability in the future, and interoperability and resource discovery in the present**

For a technology which is both in its infancy and rapidly developing, the proliferation of standards for digital surrogates, and the wide variety of file formats in which these surrogates can be preserved, is inevitable. This is partly to be welcomed: no single body would wish to impose standards which might quickly be outstripped as technology accelerates and matures. The widespread adoption of high-speed Internet access, for example, is revolutionising the amount of data that can be transferred, allowing access to quantities and formats of material unimaginable in the now-fading dial-up era. Moreover, with so many digitisation projects occurring at a local level and in relative isolation, there must be flexibility in adoption of recommendations in line with individual circumstances. The proliferation of standards is not only inevitable but in some ways even desirable, allowing the most robust, flexible and future-proof formats to rise to the top over time.

*The proliferation of standards is not only inevitable but in some ways even desirable, allowing the most robust, flexible and future-proof formats to rise to the top over time.*

Nevertheless, the now familiar joke that 'the good thing about standards is that there are so many to choose from' masks a genuine risk that the adoption of too many standards is the same as having no standards at all. Moreover, as the process of digitisation and the creation of surrounding metadata is a costly business, it undermines the long-term sustainability and interoperability of digital material if the approach to standards and formats is cavalier. In the light of the time and effort expended on any project, it is crucial that the process is done properly, to the highest specification, to ensure continued access as technology metamorphoses around the raw material and master files. JISC and others have taken an advisory line on standards, strongly recommending, without mandating, the adoption of certain formats over others, but the situation would be markedly improved by the establishment of clear guidelines on file formats and metadata. It is noteworthy that JISC has recently

commissioned UKOLN[29] to review and update the Information Environment (IE) architecture standards while MINERVA Technical Standards have been adopted in circa ten EU countries.

By securing the sustainability and interoperability of e-resources, their potential to provide a rewarding and high-calibre learning experience is maximised. However, the wealth of standards in play is mirrored in the large number of advisory and support bodies which are available to counsel institutions and make recommendations on the adoption of one format or schema over another. While it is laudable that there are so many sources of guidance, their proliferation further fragments the UK infrastructure for digitisation, and some collaboration and consolidation is needed in order that the guidance on standards can itself be standardised.

## 4.1 Currently adopted formats and standards

**The proliferation of standards is beginning to give way to a common consensus regarding metadata schemas and file formats. However, there is no overarching view on standards; guidelines vary according to the stakeholders concerned and the support services consulted**

Despite the bewildering array of potential standards, which differ both according to the nature of the governing institution (library, museum, archive) and the type of material being digitised, Loughborough's study uncovered some broad trends. It seems that a consensus is already emerging about metadata schemas and file formats, albeit in an unstructured and uncoordinated way. Library-based projects, for example, tend to use a variant of Dublin Core[30] or MARC[31] (Machine-Readable Cataloguing Record), encoding their metadata in XML[32] (Extensible Mark-up Language) and METS[33] (Metadata Encoding & Transmission Standard). Subject access is generally provided through the detailed Library of Congress subject headings, although the second most popular response from survey respondents was the use of a unique system of the library's own devising, suggesting that the ad hoc approach, with no concern for interoperability, is still thriving in some quarters. Archives use EAD[34] (Encoded Archival Description) and ISAD(G)[35] (General International Standard Archival Description) for records and search tools to meet their own needs, while publishers and providers of digitisation services are more pragmatic and flexible, using whichever standards seem appropriate for a given project at a given time. Many of those surveyed expressed a view that their involvement in the field of standards and metadata had placed them on a steep learning curve, and a number of interviewees stressed a need for further guidelines, specifically on how metadata should be applied. The possibility of automating metadata production was raised by one interviewee, and this issue needs further exploration.

29 UKOLN www.ukoln.ac.uk

30 Dublin Core http://dublincore.org

31 MARC www.loc.gov/marc

32 XML www.w3.org/XML

33 METS www.loc.gov/standards/mets

34 EAD www.loc.gov/ead

35 ISAD (G) www.icacds.org.uk/eng/standards.htm

Variance in file formats is stabilising, with a broad consensus on the use of TIFF[36] for master files and a combination of the JPEG[37] family and PDF[38] for delivery. XML has been used for preservation and delivery, as befits its flexible, bespoke nature. Nevertheless, Loughborough observed that file formats still vary according to material digitised and advice received by the project managers, funding body, or support services. This again reflects the broad and eclectic taxonomy of material being digitised, which might range from papyri to newsreels. In the current landscape, the result is a strange mixture of proliferation and a kind of tacit standardisation, each the result of the different contexts in which individual digitisation projects occur.

## 4.2 Cataloguing of digitised material

**Clear guidelines on standards, formats and metadata should now be established. The community itself would like more detailed advice on issues such as the application of metadata**

**Metadata creation is becoming a more urgent priority than digitisation itself in some cases. It is a crucial but costly part of the process and its creation should be costed into funding bids for projects**

There is a clear 'chicken and egg' issue regarding the relationship between cataloguing and digitisation, and it is not always properly understood that the two processes function in an interdependent way. Loughborough's study found that digitisation can sometimes be held up by cataloguing issues, and that finding aids to collections may need to be digitised in advance of the raw material itself. The British Academy has welcomed such a sequence, however, and noted that, while not the most glamorous candidate for digitisation, catalogues are more widely used than any individual resource named within them. Whether digitisation should always wait until cataloguing has been completed is a moot point.

*Poorly described digital resources fail the user, who cannot locate them...*

While individual resources and projects may be well catalogued, there are sometimes issues concerning inadequate metadata. Poorly described digital resources fail the user, who cannot locate them, and the information community, by increasing the likelihood of duplication of effort, and the digitisation of non-unique material more than once because there are insufficient records to disclose its prior existence online. Two-thirds of survey respondents indicated that metadata was created for all digitised items, but some flagged the lack of bibliographic records as an issue and noted that in some cases metadata creation has become a more urgent priority than digitisation itself. This is important, as digitisation cannot be satisfactorily undertaken if the materials are not underpinned by a sturdy information infrastructure, but metadata creation is expensive and needs to be costed into funding bids and increasingly accepted as a vital part of the process. The New Opportunities Fund (NOF) has funded digitisation to the tune of £50 million in recent years, for example, but its project summary concluded: 'Many projects

36 TIFF www.awaresystems.be/imaging/tiff/faq.html

37 JPEG www.jpeg.org

38 PDF www.adobe.co.uk/products/acrobat/adobepdf.html

underestimated the time, effort and expertise required to create the metadata needed to adequately describe digitised material'. There are some stark shortcomings exposed by Loughborough's study, too. Interviewees were asked if their resources were 'accessible to users with special needs', and only a third confirmed that they were. In part this was explained, once again, by the relative novelty of the technology, and the argument that accessibility guidelines are in themselves a developing area.

## 4.3 Range of support services

**There are many sources of guidance, some with overlapping remits. Different support services seldom work together and there is scope for collaboration and consolidation here**

There is a formidable range of support services that help shape the adoption and development of standards and advise those managing digitisation projects, and the UK is recognised for its training infrastructure capacity. Support services range from the Library of Congress Digital Formats and Preferences[39] site, which tracks the evolution of formats, to the Arts and Humanities Data

39 Library of Congress Sustainability of Digital Formats **www.digitalpreservation.gov/formats/index.shtml**

Service[40] (AHDS) guides to good practice. The Technical Advisory Service for Images[41] (TASI), meanwhile, covers emerging formats and standards and advises the higher (HE) and further (FE) education communities on the digitisation of images. Other sources of advice include the Digital Curation Centre[42] (DCC), which is primarily concerned with born-digital material, and the Higher Education Data Service[43] (HEDS), which offers consultancy and production services to not-for-profit organisations from any country. UKOLN, meanwhile, is jointly funded by JISC and the Museums Libraries and Archives Council (MLA), and offers advice on standards and digital preservation to the cultural heritage and education communities, while the AHDS is one example of the UK data centres that provide storage and long-term preservation measures for digitised collections. The JISC-funded British Universities Film and Video Council[44] (BUFVC) promotes the production, study and use of film and related media in FE, HE and research through courses and consultancy, and provides guidance on digitisation of sound and film. Meanwhile, the TechDis[45] service, also funded by JISC, aims to enhance provision for disabled students and staff in education and adult learning through the use of technology, and can offer advice on accessibility accordingly. The EU-funded MINERVA initiative collates a range of guidelines from around the world, but the absence of a standard guide to the range of standards is itself telling.

Loughborough interviewed representatives from a range of support services to determine their priorities and gauge their use by, and standing in, the community. There is considerable overlap between the types of support offered by different bodies (many of which are JISC funded) and these include mailing lists, workshops, publications, and guidelines on standards and preservation. Some interviewees stated that while they make strong recommendations regarding metadata, they can only encourage and not mandate use (this point was also echoed by representatives from funding bodies, described in part 5). The remit of support services tends to be defined either by discipline (eg the AHDS), or the type of material digitised (eg TASI). Some services are limited to the FE and HE sectors; others serve the commercial and public sectors, including museums, libraries and publishers. This specialisation is advantageous in certain ways, helping establish consistent standards in a given area, matching up discrete yet related projects, but it also adds to the confusion of the overall picture, increasing the number of organisations that seek to influence any one digitisation project, which might itself have a foot in several camps.

All support services interviewed felt their advisory role was becoming increasingly valued, and they see themselves as having an important role to play in the area of metadata and standards in the future. While there is considerable overlap in the remits of each body, there is no commensurate level of cooperation. The absence of collaboration between like-minded services only fragments the sector, isolating digital programmes which could otherwise be easily linked. The future role for these services, meanwhile, will differ according to resources available and communities served; their key strategic aim is often to secure further funding in order to continue their work.

40 AHDS http://ahds.ac.uk

41 TASI www.tasi.ac.uk

42 DCC www.dcc.ac.uk

43 HEDS http://heds.herts.ac.uk

44 BUFVC www.bufvc.ac.uk

45 TechDis www.techdis.ac.uk

*However enthusiastic the support services may be in providing advice and guidance, they can only be influential and take a proper overview if those managing digitisation projects seek to consult them in the first place.*

However enthusiastic the support services may be in providing advice and guidance, they can only be influential and take a proper overview if those managing digitisation projects seek to consult them in the first place. Loughborough's study revealed a poor take-up of many services offered and found that the most popular source of advice solicited by institutions was that of 'internal technical experts', thus perpetuating the localisation of projects. Respondents expressed concern at the lack of practical courses on offer and noted that simple logistics are powerful barriers to attending some training events: if the cost is high, or the location inconvenient, they will not attend. Nevertheless, TASI was revealed as a popular service, closely followed by AHDS and HEDS. The DCC may become more widely used as it becomes more established. The BUFVC was not well used, but this may reflect the relative paucity of film and video digitised to date. TechDis is also seldom used: a matter for concern when only a third of respondents were able to confirm that their projects were fully accessible.

More generally, Loughborough found a desire among interviewees simply to know what others in similar circumstances were doing. Networking is the key to managing transitions in models and standards; simply turning up at the appropriate conferences helps make the state of current thinking clear, and places local initiatives in a wider context.

The issue of standards should not just be viewed from the point of view of the professional; it impacts on the experience of the end user, too. In the absence of assured and consistent standards for material digitised, it becomes difficult to set universal standards for eventual resources created. The quality and benchmarking of project outcomes is as variable as the different methods used to implement them. In theory, the ability to measure quality – both in terms of materials digitised and the design of the e-resource which gives access to them – would be invaluable, and would help to tie disparate projects into a common framework. Although the link between high-quality metadata/formats and a high-quality experience for the researcher is perhaps indirect, it is nevertheless significant. The more standards and schemas become streamlined and entrenched, which could be facilitated through the UK framework we are recommending, the easier resources will be to use, and the more valuable the learning experience for the end user.


# 5.0 Funding structures and opportunities for digitisation projects

## 5.1 Funding structures

**Those involved in digitisation projects view the current fragmented funding structure as unsatisfactory, especially the way digitisation has been funded largely on a short-term 'project' basis**

**Guidance on the wealth of funders should be streamlined into a 'one-stop shop' for information. This could result in the creation of a portal useful to libraries, archives and others**

Digitisation projects are expensive and lack of funds is the most basic, yet most significant, impediment to getting projects off the ground. Any programme incurs a wide range of costs, including documentation and preparation, conversion, rights clearance, equipment, human resources and maintenance. Projects are often funded through a variety of avenues, including institutional budgets, public grants, corporate sponsorship and private donations. Even among external funding bodies, there is deep fragmentation. Just as emerging standards, types of material digitised and support given have been piecemeal and uncoordinated, so too have the funding structures which have made projects possible in the first place. This multiplicity of funding streams and agendas is not in itself unwelcome as it maximises the avenues for potential digitisers to explore. The resources now at the desktops of individuals and institutions in the UK have accumulated without the support and motivation of centralised funding and a framework for digitisation as part of a UK e-Content Strategy. Work to date has been largely funded on a short-term project basis, with major funders including the Andrew W Mellon Foundation in the USA, the various lottery bodies and JISC, among others. Institutions tend to decide on which funding bodies to approach depending on the nature of the material to be digitised, the target user community, and the volume of funds required. Some funders cover only the HE and FE sectors; others cover museums, libraries and archives. While JISC has tried to fund comprehensive projects likely to reach a large audience, and to work collaboratively with organisations to build a collection of digital surrogates

which all meet CIE standards, it has done so in the absence of an overarching framework to coordinate and link different activities and guide separately funded projects towards common outcomes.

Figure 1 illustrates the complexity and number of both advisory and funding bodies which influence the development of digitisation projects. The diagram as it stands only represents those based in the UK and does not include the HE library and archive sector, or reflect international money flowing to UK projects.

*No satisfactory single and authoritative source of advice on potential funders exists.*

No satisfactory single and authoritative source of advice on potential funders exists. The Technical Advisory Service for Images (TASI) website, however, provides some information on funding bodies which potentially fund digitisation, including the Higher Education Funding Councils, the strategic group Research Councils UK, JISC and the Arts and Humanities Research Council[46] (AHRC). In theory, the AHRC has no specific funding strategy for digitisation and allocates awards solely on the grounds of the quality of research proposals received. Nevertheless, Loughborough found it to be moving away from responsive funding towards a more strategic method. Other sources include National Lottery-related bodies such as the Big Lottery Fund[47], the Heritage Lottery Fund[48] and (in theory) the National Endowment for Science, Technology and the Arts[49] (NESTA). The MLA works on a wide range of initiatives and provides advice and information regarding the digitisation of content. It has had involvement in projects such as Curriculum Online[50], acted as expert adviser for NOF's EnrichUK programme[51], and has funded digitisation projects in its Designated Museums Challenge Fund[52] and through the Renaissance in the Regions programme[53]. Two other notable sources of funding which have made considerable contributions to UK digitisation projects are the US-based not-for-profit Andrew W Mellon Foundation, and the Wellcome Trust, an independent, privately owned charity, dedicated to exploring biomedical issues and the history of medicine. Wellcome has been involved with digitisation projects in these subject disciplines, including the current Medical Journals Backfiles Project[54] (an innovative collaboration between Wellcome, JISC and the National Library of Medicine (USA) which is digitising complete runs of medical journals and placing them on open access through the PubMed Central interface). Given the confusing and large number of potential sources of funding, and the amount of time project managers invest in locating and approaching these, the creation of a 'one-stop shop' of advice and information on funders would expedite the process for everyone.

**46** AHRC **www.ahrc.ac.uk**

**47** Big Lottery Fund **www.biglotteryfund.org.uk**

**48** Heritage Lottery Fund **www.hlf.org.uk/English**

**49** NESTA **www.nesta.org.uk**

**50** Curriculum Online **www.curriculumonline.gov.uk/default.htm?cookie%5Ftest=1**

**51** EnrichUK **www.enrichuk.net**

**52** Designated Museums Challenge Fund **www.mla.gov.uk/action/designation/desig_dcf.asp**

**53** Renaissance in the Regions Programme **www.mla.gov.uk/action/regional/ren_report.asp**

**54** Medical Journal Backfiles **http://library.wellcome.ac.uk/node280.html**

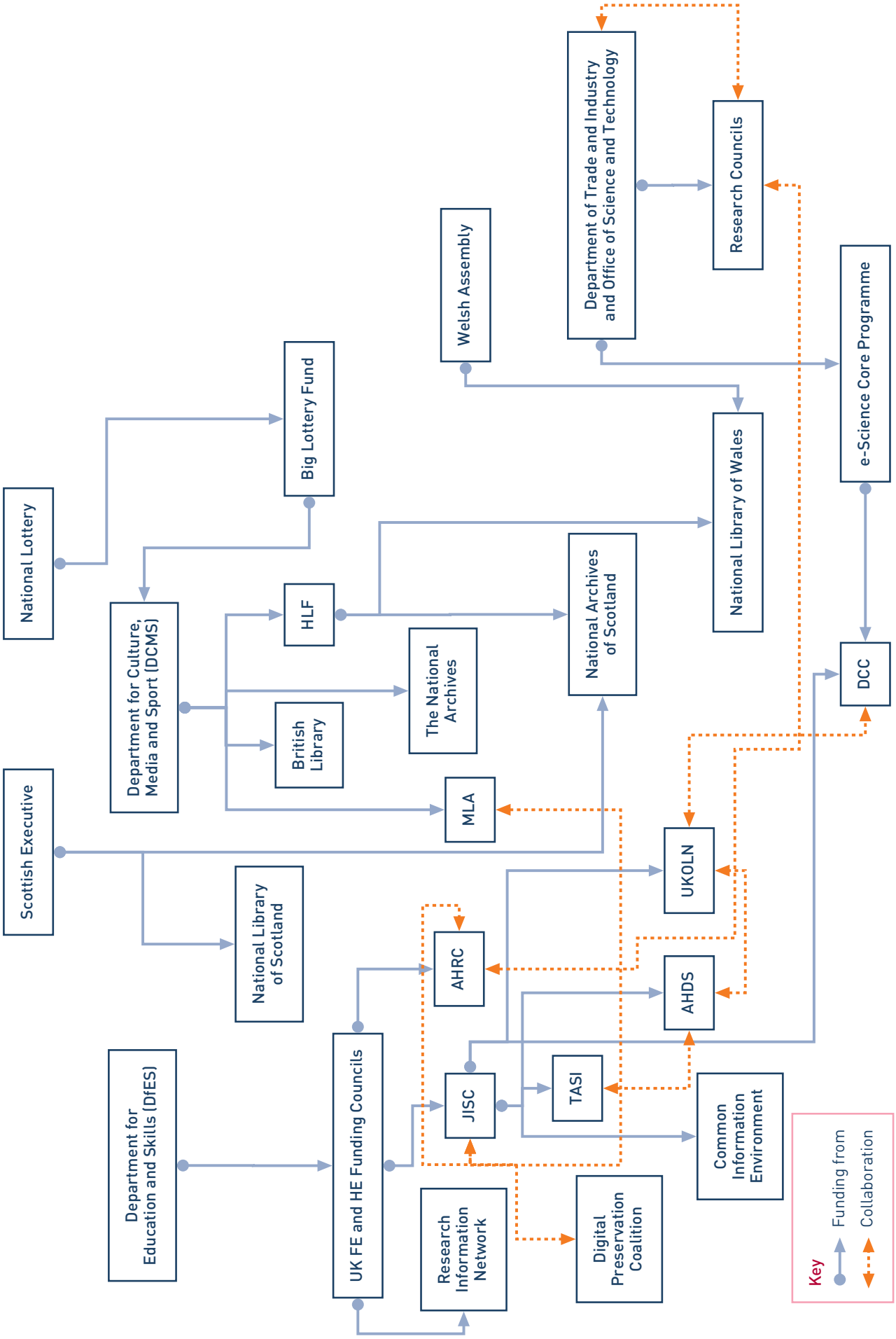



**Figure 1**
*Complexity of advisory and funding bodies influencing digitisation projects*

Loughborough discovered that around half of its survey respondents derive funding for projects from a mixture of internal and external sources but that external sources usually provide the bulk of the funds. Public-sector funding, especially lottery-related, is common. Popular combinations of funding sources include: (i) the research boards/councils and the National Lottery; (ii) JISC and the National Lottery; and (iii) a fusion of all three. Government money for such projects is distributed to the research community through the funds to institutions and to research councils. In addition, the Higher Education Funding Council for England (HEFCE) and the Department for Education and Skills (DfES) pay a block grant to UK universities, from which most of the HE library budgets are derived. In general, external money makes up the bulk of the budget, and funders inevitably play a crucial role in determining the future of electronic resources.

## 5.2 Concerns for the future

**The different funders do not tend to work together to provide joint funding for digitisation projects; funders should be able to work more cooperatively, coordinating bids and grants, and working together to fund major projects**

**Funding bodies are concerned about issues of long-term preservation and stability of resources; as a result, some are becoming more strategic and making consideration of these areas a condition of funding**

When Loughborough's interviewees were asked about their concerns for the future of UK digitisation, there was widespread agreement that the current funding structure is unsatisfactory, particularly the 'project' nature of grants which does not necessarily guarantee long-term maintenance of resources created. The study noted an obvious yet crucial relationship between the future of digitisation and the future of funding opportunities: just as funding facilitates endeavours, its absence stifles their development. The funding bodies themselves have their own priorities and agendas and have generally not worked together to provide joint funds for major projects. Currently, many funders and digitisers alike are concerned that the future of digitisation is precarious because it is entirely contingent on whatever (limited) funds are available at a given time. Yet there is no apparent reason why funders, especially those in the public sector, could not work more cooperatively, coordinating bids and grants through existing or new forums.

Studies conducted for JISC and the National Preservation Office concluded that 'a great deal of money can be wasted if digitisation projects are undertaken without due regard to the long-term preservation of the digital files'. Many funding bodies interviewed by Loughborough expressed concern that the projects themselves did not take into account long-term issues of preservation, sustainability and access, and that when funding runs out, or the project is ostensibly 'completed', the resources created are not properly managed. Those

receiving funds, conversely, felt there was a need on the part of funding bodies to review their strategy accordingly, and to include funding for the longer term. Some acknowledge this and are becoming more strategic, even limiting the size of grants unless projects can be shown to be factoring in sustainability of the resources.

Loughborough found occasional examples of collaboration between funding bodies, but this is usually limited to joint funding of a particular initiative. More common is collaboration between a funding body and a support service, such as the relatively symbiotic relationship (in matters of funding) between the Arts and Humanities Research Council (AHRC) and the Arts and Humanities Data Service (AHDS), the latter providing technical input into funding applications received by the former. Such collaboration is vital in integrating the different responsibilities of the bodies which orbit digitisation projects. Funders, for example, recognise the importance of standards but feel it is not their place to impose them (with only occasional exceptions such as the nof-digitise programme, which did insist on standards as a condition of funding). Direct cooperation between funding body and advisory service rectifies this, seemingly devolving responsibility to the 'experts', while ensuring their advice is contained within the overall channels which facilitate the project in the first place.

## 5.3 Cooperative and commercial digitisation activities

**Innovative cross-sectoral collaborations are changing the models of digitisation projects; the Google initiative will hasten this development**

**Any future developments involving commercial partners need to avoid the appearance of the library community being sold its resources at a premium**

The wealth of digital resources has been amplified and strengthened through a range of innovative collaborations between different parts of the library, education and cultural sectors, and between the public and private spheres. While again many initiatives have been uncoordinated and not formalised, such opportunistic and innovative projects are maximising the amount of material brought online and making ingenious connections of a kind only possible in the virtual world. Collaboration has been particularly strong in Scotland and successful projects like the Glasgow Digital Library, in which Scottish libraries work together to create a corpus of digital content, are evidence of what can be achieved. Participants in Loughborough's study were also asked about their reasons (theoretical and actual) for taking part in cooperative digitisation projects and the most frequent response concerned the potential to build collections from dispersed materials. Clearly, one of the most obvious benefits of digitisation, the virtual collocation of physically disparate material to provide vast increases in the amount of source material on offer to the researcher, remains as potent as ever and must be continuously developed.

*Loughborough found that outsourcing was often done because the expertise, equipment and staff were lacking on site.*

Commercial collaboration is also evident in the digital landscape and most frequently takes the form of outsourced digitisation. Loughborough found that outsourcing was often done because the expertise, equipment and staff were lacking on site. However, those who undertook digitisation in house noted strong reasons for doing so, including better quality control and the opportunity precisely to develop those staff skills that had been identified by others as lacking.

Publishers, including Thomson Gale and ProQuest, are also involved in the digitisation of backlists and have produced popular resources like Eighteenth Century Collections Online[55] (ECCO) and Early English Books Online (EEBO), respectively. ProQuest is also behind the Text Creation Partnership (TCP), a collaborative venture in support of ECCO and EEBO, based at the University of Michigan. It unites a range of scholarly publishers and libraries and is creating manipulable and searchable electronic editions of titles from the Short Title Catalogue of Early English Books. Libraries may join if they purchase EEBO, and participants in TCP are helping to shape the archive. The venture particularly values library input, in fact, and believes it 'provides a model for partnerships between publishers and libraries to serve a common goal: meeting the research needs of end users'. The TCP is seen as a particularly successful new model involving all stakeholders: publisher, library and researcher, created and maintained by the community, for the community. All these projects provide great benefit to libraries by allowing access to high quality and in-demand digitised material. The downside is the price: some resources are hugely expensive, bordering on the cost-prohibitive for some institutions where a JISC UK Licence has not been agreed. Any future developments need to avoid the appearance that the library community is being re-sold its own resources at a premium.

## 5.4 Responses to Google Print

Unsurprisingly, publishers surveyed by Loughborough viewed the Google Print Library Project initiative with concern and felt it undermined their business model of charging for the supply of digitised content. Libraries received the news far more positively, with some expressing a desire to become directly involved. One interviewee reported that a funding body had expressed interest in funding projects that contribute directly to the Google initiative, which, it seems, is already impacting on business models and research library interest in digitisation. Google itself claims: 'What we are doing is not intended to replace or discourage funding for the efforts of others working to digitise library collections. We hope that our entry into this arena will attract needed attention to digital library initiatives worldwide.'

55 ECCO www.gale.com/EighteenthCentury/about.htm

## 5.5 Roles and responsibilities of stakeholders

As the number and diversity of stakeholders involved in different projects, from the British Library to commercial publishers, charitable trusts to research councils, has increased, the landscape has become entangled to the point that no one clearly perceives their wider roles and responsibilities towards individual projects and the future of digitisation overall. Loughborough uncovered a wide variety of arrangements for the organisation and implementation of digital projects. For example, the National Library of Wales (NLW) funds, delivers and maintains digitised material, and is thus a largely self-sufficient content and service provider. The British Library (BL), by contrast, has undertaken a large number and wide range of digitisation projects, increasingly on a very large scale. These have been mainly funded from external sources – public, private or charitable – and, for the biggest projects, digitisation and hosting arrangements have tended to be outsourced. Hosting will increasingly be taken in-house following completion of the BL's national digital library development. Since the Loughborough study was carried out, the BL has updated its digitisation strategy and also announced its partnership with Microsoft, who will be funding, for delivery by both Microsoft and the BL, the digitisation of out-of-copyright BL content.

A similar, equally fundamental, issue to resolve is where the responsibility for sustainability of resources lies: whether funding bodies should provide for ongoing maintenance or whether it is the responsibility of digitisers. Loughborough have made several recommendations here. Digitisers need guidance on long-term management and pointers to the support services that can assist. The Digital Preservation Coalition[56] and the newly established Digital Curation Centre can help. Funding bodies and recipients of grants should consider the use of existing data archives to facilitate safe storage and preservation of digitised resources. Several digitised archives already exist in the sector and could be added to with data from ongoing and future projects with relative ease.

## 5.6 New models for collaboration

Companies like ProQuest, in contrast to most libraries, have the means and the expertise to digitise but lack the content and, of course, can only operate satisfactorily on the basis of profitability. Publishers face the challenge of emerging new models in which libraries will work increasingly with companies like Google to digitise and provide content. As one put it: 'We need to work more collaboratively with the research community and with JISC in order to identify what the projects are, what the content is, that the research community needs.'

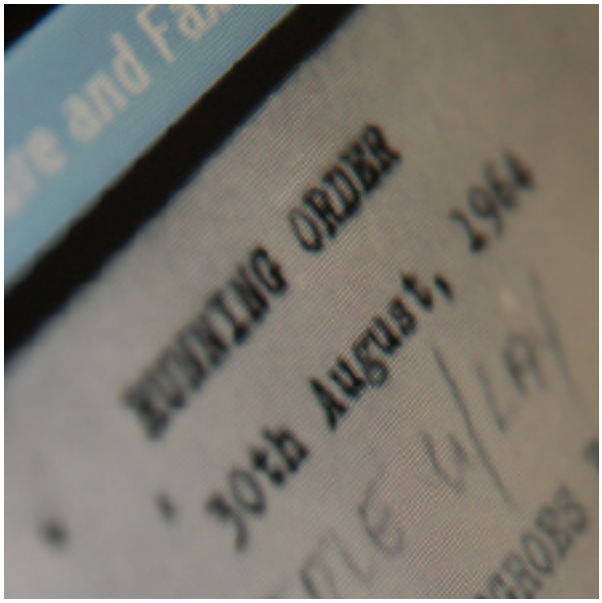
A radical and productive partnership can be found in the JISC/Wellcome Trust/ National Library of Medicine collaboration on the Medical Journals Backfiles digitisation project. The combination of the three bodies is a potent cocktail of initiative, expertise and funds, with the content itself drawn from commercial

56 Digital Preservation Coalition www.dpconline.org/graphics/index.html

publishers and made available through open access. The project is a powerful example of a new model of supporting research provision by strategic funding of digitisation.

As business models change, and publishers consider their response to the Google initiative, one constant is that the research sector considers the most acceptable models to be those which secure free access at point of use, or, at worst, for a modest subscription. With funding often drawn from public money and the content from research libraries, this seems only fair. Nevertheless, the private sector is sometimes able to digitise resources for which funds cannot be found elsewhere, and these commercial publishers need to make an acceptable profit. What should be avoided, however, is the requirement for research libraries to pay substantial subscriptions for access to privately digitised material drawn originally from their own collections.

Loughborough's study, and this resulting document, is not calling for a unified strategy for funding UK digitisation. It would be unrealistic to expect the various UK public sector funding bodies, let alone independent and international sources of funding, to coordinate efforts and subsume all their individual agendas and needs to the 'greater good' of a UK digital picture. Nevertheless, it is possible to make recommendations in the light of best practice successful models of collaborative funding, such as the public–private partnerships described, and to streamline sources of advice. Moreover, funders should be encouraged to work more cooperatively and strategically, especially on large projects. This could be done through existing forums, such as Research Councils UK, or a new forum which includes the main funding bodies.




# 6.0 Conclusions and recommendations

In just a handful of years, and against the challenging backdrop of rapidly evolving technology, the UK has created some remarkable e-resources, from the Old Bailey Session Papers to Turning the Pages, with more to come in the forms of Newsfilm Online and the Medical Journals Backfiles digitisation project, among others. Internationally, highly praised initiatives, such as Early English Books Online and the Text Creation Partnership, have won widespread support from the research community. Nonetheless, it is an inescapable fact that most UK-based projects have been funded and created in a vacuum, executed locally and within a fragmented infrastructure. Much more could be achieved with a framework in place to coordinate all constituents, weaving together the number of bodies and agendas which surround a project, placing each digitisation programme in a UK context, and responding to challenges such as Google Print by capitalising on the opportunities they throw up. It is time to turn the rough-hewn and loose-linked 'crazy paving' of individual digitised resources in the UK into a seamlessly integrated mosaic, where each project informs the others, and all contribute to a bigger picture.

*The key theme running through Loughborough's study is the lack of coordination among all constituent parts of the digitisation process, and it seems remarkable that the great successes of recent years have been achieved despite the schisms in evidence.*

The key theme running through Loughborough's study is the lack of coordination among all constituent parts of the digitisation process, and it seems remarkable that the great successes of recent years have been achieved despite the schisms in evidence. There is still much more material currently undigitised, and a great demand that this be addressed. Moreover, the more content that can be digitised, the greater the opportunity to streamline standards, which will in turn lead to better value (for users, funders and providers), enhanced research, and richer learning and teaching experiences. As one funding body told Loughborough: 'What material exists that wouldn't benefit the discipline? I can't think of any that would not be useful if available in digital form.'

There is now an opportunity to address all shortcomings identified in Loughborough's study at once, from inadequate metadata to lack of collaboration, by uniting the sector through a framework for digitisation as part of UK e-Content Strategy. A Digitisation Framework would assist in filling gaps in provision, cut across the efforts of individual funders and digitising

organisations, reduce overlaps between support services and assist in the provision, take up and use of resources. Fears that any such 'nationalisation' might stifle local innovation can be allayed by placing all digitisation activities within the flexible framework envisaged; one which issues clear guidelines rather than prescriptive demands, which draws up a series of 'gold standards' kept under regular review. Such a framework, then, should be coordinated and distributed, rather than centralised.

While librarians and archivists have sought to find and adhere to standards, and JISC has supported this, a UK approach would also assist in overcoming institutional issues, such as successful project management being impeded by costs, confusion over file and metadata formats, and preservation problems. Any formalised response to all these issues must still reflect researchers' priorities; as the British Academy has observed in their report *E-resources for Research in the Humanities and Social Sciences* (2005), too much of the recent provision of digitised materials has been led by supply, not demand.

Successful models from which the UK can learn are readily found abroad. In the United States, the US Registry of Digital Masters[57] has been established and allows digitisation project managers to see if a specific item has already been digitised, and to check whether the standards and formats are acceptable. The Registry produces interoperable catalogue records which can be extracted by libraries at the local level. Australia's national library offers a fee-based 'digitisation on demand' for materials not yet digitised and has produced a sophisticated resource, with an overall framework and systems architecture listing 82 projects. And in New Zealand, the National Digital Forum[58] composed of cross-sectoral organisations, identifies opportunities for collaboration, cooperation and information; it also negotiates funding and grant applications at the national level.

There are three clear recommendations to make on the basis of Loughborough's findings as follows:

## 6.1 Establish a UK framework for digitisation

There is a pressing need for a UK body which can stand back and see the holes in the patchwork of digital projects, to recommend standards and to take a coherent line on access mechanisms. We need a framework as part of a UK e-Content Strategy, not simply to respond to the powerful drivers for digitisation (access, demand and preservation), but because these drivers, and the response to them in the forms of funding, projects and services, are not being coordinated effectively. There are strong currents of goodwill, enthusiasm for digital projects, and a desire to bring increasing volumes of material online, in the community, from users and providers alike. These must be tapped.

**57** US Registry of Digital Masters
**www.oclc.org/news/announcements/announcement149.htm**

**58** New Zealand National Digital Forum
**http://ndf.natlib.govt.nz**

## 6.2 Coordinate existing services

Part of the framework's remit would be to review the current provision and possible consolidation of support and advice on all related issues (while recognising the diverse remits of support services and the needs of their end users), from guidance on standards to a comprehensive listing of all projects undertaken, and the provision of a portal for information on funding streams. The framework should also consider whether a single access point to services offering guidance on standards would help foster interoperability and sustainability, and whether or not this would benefit funders by enabling them to derive better value from their investment. A comprehensive resource listing all digitised projects, meanwhile, would not only authoritatively chart the landscape, enabling better resource discovery and increased use of materials, it would also prevent unnecessary duplication, as project managers could check to make sure their materials had not already been digitised. Funding bodies (and the bidders), meanwhile, would benefit from coordination through a single point and this may facilitate better cooperation between funders. It is important to note that the EU 7th Framework proposes to set up seven centres of competence which will complement existing services and promote digital preservation.

## 6.3 Investigate researchers' needs

Future developments in digitisation need to respond more directly to user demand rather than library supply, yet researchers' needs (and searching behaviours) are still not fully understood. Insight is particularly lacking into requirements in the science and social science fields. Various channels, including subject associations, royal societies and academies, can be used to gauge need and focus the general goodwill and enthusiasm from the research community into a targeted programme which fulfils specific needs and plugs information gaps. Additionally, the Research Councils UK should find out more about researcher needs, executing their surveys through the individual councils. The Research Information Network (RIN) also has a key role to play in investigating user needs. While researchers may be more likely to respond to the research bodies than the RIN, there is an opportunity here for the RIN to make an impact. The library community, in turn, can do much to encourage RIN awareness. Once surveys have been undertaken and findings analysed, the results should be shared with JISC and the Consortium of University Research Libraries, the response coordinated, and used to inform policies, procedures and strategies. Knowledge of needs and demands should be periodically updated to coincide with, for example, strategic reviews or content creation programmes. Funders should seek real evidence of user need as a part of application for future digitisation projects.

*While researchers may be more likely to respond to the research bodies than the RIN, there is an opportunity here for the RIN to make an impact.*

# Appendix:

## Loughborough's recommended points for action

1. Study the results of the British Academy survey and identify the implications for resource provision in the arts and humanities research community (CURL, RIN).
2. Continue to focus on raising awareness of, and training and tools for using, digitised content (JISC, AHRB, e-Science programme, etc).
3. Establish ongoing and systematic research into researcher needs, particularly in the sciences and social sciences where this is a high priority (JISC to discuss with Research Councils UK, RIN to discuss with associations and societies).
4. Coordinate findings of investigations into researcher needs at a UK level (Research Councils UK, JISC, CURL).
5. Establish ongoing and comprehensive gap analysis to identify priorities for the digitisation of material (CURL, JISC, RIN).
6. Examine alternative approaches to speed up and reduce the cost of metadata creation, including: funding body support for this activity, automation and possibly outsourcing (CURL, RIN, JISC).
7. Establish a UK Register of Digital Surrogates, similar to the National Register of Archives, to facilitate greater collaboration and cooperation (CURL, JISC, RIN).
8. Create a single point of information on current and previous digitisation projects (CURL, JISC, RIN).
9. Funding bodies should include provision of information to digitisation registers as a condition of funding (Research Councils, JISC to discuss with other major funding bodies).
10. Investigate how to improve discovery of digitised materials (JISC and the MLA through the Common Information Environment).

11. Examine potential for consolidation of existing JISC support and advice services where appropriate (JISC).

12. Encourage the use of current standards as far as possible (Common Information Environment).

13. Create a single point of information on funding opportunities (CURL, JISC, RIN, MLA).

14. Create 'best practice' exemplars for public–private partnership collaboration (JISC).

15. Continue to focus on raising awareness of, and training and tools for, digital preservation through the Digital Preservation Coalition and Digital Curation Centre (JISC).

16. Encourage the recipients of funding to use existing data archives to facilitate safe storage and preservation of digitised resources when planning and funding digitisation projects (JISC and CURL to discuss with funding bodies and research libraries, respectively).

17. Hold a symposium with key UK and international representation on how a UK-wide digitisation strategy could be coordinated, including the creation of a forum for the ongoing sharing knowledge, developing policy and implementation plans.

# Glossary

## Bit rate

In digital telecommunication, the bit rate is the number of bits that pass a given point in a telecommunication network in a given amount of time, usually a second.
www.whatis.com

## Digitisation

The process of converting information into a digital format. In this format, information is organised into discrete units of data (called bits) that can be separately addressed (usually in multiple-bit groups called bytes). This is the binary data that computers and many devices with computing capacity (such as digital cameras and digital hearing aids) can process.
www.whatis.com

## Document Type Definition

A Document Type Definition (DTD) is a specific document-defining and -constraining definition or set of statements that follow the rules of the Standard Generalized Markup Language (SGML) or of the Extensible Markup Language (XML).
www.whatis.com

## Download

Downloading is the transmission of a file from one computer system to another, usually smaller, computer system. From the Internet user's point of view, to download a file is to request it from another computer (or from a Web page on another computer) and to receive it.
www.whatis.com

## Dublin Core

Dublin Core is an initiative to create a digital 'library card catalogue' for the Web. Dublin Core is made up of 15 metadata (data that describes data) elements that offer expanded cataloguing information and improved document indexing for search engine programs.
www.dublincore.org

## EAD

Encoded Archival Description: the EAD Document Type Definition (DTD) is a standard for encoding archival finding aids using Extensible Markup Language (XML).
www.loc.gov/ead

## Encoding

In digital technology, an encoding is a highly compressed video or sound file that preserves the quality of a video or audio recording.
www.whatis.com

## ISAD(G)

General International Standard Archival Description: this standard provides general guidance for the preparation of archival descriptions. It is used in conjunction with existing national standards or as the basis for the development of national standards.
www.icacds.org.uk/eng/ISAD(G).pdf

## International Press Telecommunications Council

The IPTC, based in Windsor, UK, is a consortium of the world's major news agencies and news industry vendors. It develops and maintains technical standards for improved news exchange which are used by virtually every major news organisation in the world.
www.iptc.org

## JISC Information Environment

The JISC IE Architecture specifies a set of standards and protocols designed to realise the vision of delivering digital resources and services to users in an integrated way.
www.jisc.ac.uk/ie_home.html

## JPEG

Joint Photographic Experts Group: a term for any graphic image file produced by using a JPEG standard. A JPEG file is created by choosing from a range of compression qualities (actually, from one of a suite of compression algorithms).
www.whatis.com

## MARC

Machine-Readable Cataloguing Record: the MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form.
www.dublincore.org

## Mbps

Mbps stands for *millions of bits per second* or *megabits per second* and is a measure of bandwidth (the total information flow over a given time) on a telecommunications medium.
www.whatis.com

## Metadata

Meta- is a prefix that in most information technology usages means 'an underlying definition or description'. Thus, metadata is a definition or description of data.
www.whatis.com

## METS

Metadata Encoding & Transmission Standard: the METS schema is a standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium.
www.loc.gov/standards/mets

## OCR

OCR (Optical Character Recognition) is the recognition of printed or written text characters by a computer.
www.whatis.com

## Open Access

The Open Access research literature is composed of free, online copies of peer-reviewed journal articles and conference papers as well as technical reports, theses and working papers. In most cases there are no licensing restrictions on their use by readers. They can therefore be used freely for research, teaching and other purposes.
www.jisc.ac.uk/index.cfm?name=pub_openaccess

## PDF

PDF (Portable Document Format) is a file format that has captured all the elements of a printed document as an electronic image that you can view, navigate, print, or forward to someone else.
www.whatis.com

## TIFF

TIFF (Tag Image File Format) is a common format for exchanging bitmap images between application programs. It is one of the most common graphic image formats. TIFF files are commonly used in desktop publishing, faxing, 3-D applications and medical imaging applications.
www.whatis.com

## XML

XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets and elsewhere.
www.whatis.com

# Digitisation in the UK – the case for a UK Framework

**This document is available in alternative formats**
*For more information: www.jisc.ac.uk/digitisation*

JISC

Further information about JISC:
Web: www.jisc.ac.uk
Email: info@jisc.ac.uk
Tel: 0117 954 5083

consortium of curl research libraries in the British Isles

Further information about CURL:
Web: www.curl.ac.uk
Email: robin.green@curl.ac.uk
Tel: 0121 415 8106

Version 1.1, November 2005