*The British Academy*

# E-resources for research in the humanities and social sciences

## A British Academy Policy Review

**April 2005**

*The British Academy*
10 Carlton House Terrace
London SW1Y 5AH

Telephone: 020 7969 5200
Fax: 020 7969 5300
Email: secretary@britac.ac.uk

Web site: www.britac.ac.uk

# Foreword

The maintenance of Britain's high international standing in advanced research in the humanities and social sciences is critically dependent on the provision of ready access to the materials for research, whether archives, texts, images, books, or journals.

These resources are gathered, held, and funded in a variety of different institutions which operate a variety of different mechanisms. But the long-established mechanisms for accessing traditional, non-electronic materials are proving increasingly inadequate in consequence of the radical changes which are being brought about by current information technology. The new technology offers new kinds of research resources, new ways of exploring them, and new methods for identifying what is electronically or non-electronically available. But provision of electronic resources in Britain is at present both patchy and uncoordinated.

The British Academy, as the country's national academy for the humanities and social sciences, is particularly well placed to consider the resource needs of all of the humanities and social sciences and to compare these with the needs of researchers in natural science, technology, and medicine. The Council of the Academy has accordingly decided that a review of the state of research resources provision, and particularly of electronic resources provision, in the humanities and social sciences is urgently needed.

The Academy's review shows that many individual researchers are actively exploiting e-resources and that some valuable initiatives in the provision of such resources are under way in selected institutions. But much more needs to be done. Many complex issues need to be addressed if there are to be in place the policies and mechanisms which will ensure that researchers in the humanities and social sciences have the access to e-resources that they are increasingly going to need.

Lord Runciman, CBE, FBA

*President of British Academy*

# Contents

## Contents cont

# Preface

The Academy's previous policy review, '*That full complement of riches*', laid out the humanities and social sciences' vital contribution to the UK's economic advantage, social development, and cultural enrichment. Research in these areas, just as in science, technology and medicine, depends not only on the researcher's own initiative and insight and on financial support for their work, but on access to the resource materials on which that research is based.

These resources include, notably, written materials, and other representational items, for example images; and they cover not only such primary resources as manuscripts, journal papers, numerical data or image banks, but also secondary resources or discovery tools like catalogues.

Computer and communications technology is radically changing all aspects of information management, from manuscript or map digitisation to archive portals and Web search engines. It has the potential to revolutionise access to research resources as well as the form of resources themselves, and has already begun to change the way researchers work.

At the same time, there are vast existing non-electronic resources which are, and will remain, crucial for research. The balances to be struck between electronic and non-electronic resource provision, and the most effective forms of provision, matter for every researcher. However there are many factors affecting e- and non-e resource provision for UK researchers, from technological developments and opportunities, through intellectual property regimes, to organisational structures and institutional habits.

The information world is changing very rapidly. But ad hoc reaction will not give researchers best value from the resource opportunities that are arriving, or help funders get good outcomes from their support for researchers. Our review, based on data from both individuals and organisations, argues strongly for much more coherent policies and proactive strategies to ensure UK researchers in the humanities and social sciences have access to the resource materials, and especially the new e-resources, that they need and deserve.

Professor Karen Spärck Jones

*Chairman of the Review Working Group*

# Executive Summary

## A. The Review

Information and communications technology (ICT) is having a major impact on the form of, and access to, research resources, as well as on methods and tools. This has been particularly marked in science, technology and medicine. The British Academy wished to discover what impact e-resources and e-access are having on research in the humanities and social sciences (HSS), both in their own right and through their interaction with non-e resources. ICT can make research more efficient and effective in two ways: by making primary resources like text, images and data readily available through digitisation; and by facilitating access to both e- and non-e primary resources through digital secondary resources like electronic catalogues and other discovery tools.

The Academy has consulted relevant national institutions and organisations and individual researchers in all areas of HSS to establish the present state of e-resources and their use, and to identify future needs. This Review reports on what the organisations are doing to support resource provision, and on how researchers are actually working with resources. E-resource provision is a complex area, with many component factors that have far-reaching implications. The Review thus examines the role of technological factors, of resource properties, access means, and holding requirements, and of publication and rights choices. E-resource provision is rapidly advancing, and HSS researchers are already exploiting a wide range of e-materials and tools for reaching and using them.

Thus we have found that HSS researchers are proactive and enterprising in their use of e-resources, even where much of their work is with non-e resources; are conducting their research with as much e-sophistication as workers in science, technology and medicine; are developing HSS research in new ways; and are reaching across subject boundaries within and outside HSS. Much of this gain comes from using generic resources, notably email and Web engines. But much comes from HSS-specific resources, making it clear that e-resources are vital for HSS research health.

However our Review shows that UK resource provision is ad hoc and fragmented, structural options are not being fully analysed, and important policy issues are not being vigorously addressed. More strategic, coordinated and well-targeted action is needed to ensure that the UK's own rich stock of research resources is more accessible than at present, and that UK researchers can deploy the resources they need from both UK and non-UK sources to keep UK research and scholarship in HSS at the international leading edge. This action must, moreover, be grounded in researchers' actual, not deemed, requirements, so e-resources and access to them are designed, from the beginning, for researchers' use.

## B. Recommendations

The Academy's Review has therefore led to both Specific and Overall recommendations addressed to pertinent stakeholders, notably national institutions and bodies, to universities and libraries, and to researchers themselves.

The recommendations, with their motivation, are presented in full in Section 5, along with their target stakeholders.

The **Specific recommendations**, summarised here, fall under headings as follows:

**Resource types** RECOMMENDATIONS 1–8

That provision for non-e resources, especially monographs, be maintained; that coordinated provision for digitising other primary, and secondary, resources, be adopted; and that guidelines for e-resource creation be formulated and promoted.

**Access** RECOMMENDATIONS 9–19

That HSS resource provision be maintained against STM competition; that major libraries take the lead in sustaining key non-e resources; and that national mechanisms for access to non-e resources be improved.

That high priority be given to secondary e-resources, e.g. catalogues, to help researcher access to resources; and that this is done in ways that take advantage of general ICT developments, notably the Web, and researchers' familiarity with these.

**Open access** RECOMMENDATIONS 20–21

That HSS pay attention to open access developments.

**Technical support** RECOMMENDATIONS 22–25

That universities and their libraries improve HSS-targeted technical support; and also that HSS researchers develop their own awareness of e-resources.

**Repositories** RECOMMENDATIONS 26–32

That the properties, requirements and implications of different repository models, especially for the very long term, be investigated; and that work on repository models factor in researcher access via general ICT mechanisms and facilities.

**Intellectual property** RECOMMENDATIONS 33–37

That properly equivalent interpretations of intellectual property for non-e and e- resources be developed and applied; that appropriate models for fair dealing for research purposes, especially relating to older and to audio-visual materials and to secondary resources, be formulated; and that HSS researchers consider the importance of e-access to published resources they author.

**Organisation and funding** RECOMMENDATIONS 38–45

That key players collaborate more systematically in e-resource provision, especially for the long term, to improve e-resource provision and support; that in this they address both technical and service aspects; that they pay particular attention to means for resource discovery, and the role of wider ICT mechanisms in enhancing resource access; and that researchers themselves are involved in this cooperation.

The Review's **Overall recommendations**, with the more important starred, are thus as follows:

* **A: that relevant UK institutions and bodies adopt a coordinated and coherent strategic approach to e-resource provision and access, based on research community needs.**

* **B: that e-resource conversion by resource holders pay particular attention to secondary before primary e-provision.**

**C: that all those providing e-resources address means and mechanisms for access from general information discovery systems such as Web engines.**

* **D: that the national institutions, funding bodies and library representatives collectively address the development of licensing and fair use protocols for e-resources that balance the claims of providers and users.**

* **E: that the national institutions and funding bodies conduct an in-depth analysis of the requirements and options for long-term e-resource curation, preservation and use.**

**F: that HE and other research institutions ensure that HSS researchers have sufficient access to appropriately-trained technical support staff.**

**G: that HSS researchers actively seek guidance on access to, and provision of, e-resources.**

**H: that HSS researchers actively promote user community interests to both e-resource funders and providers.**

# Section 1: Introduction

## 1.1 Information technology and research resources

Information and communications technology (ICT) is rapidly changing the nature of, and access to, research resources. This is clearly evident in Science, Technology and Medicine (STM), with indirect consequences for the Humanities and Social Sciences (HSS). But HSS are also directly affected in many different ways.

Electronic (e-) resources are replacing, or enhancing, non-e resources, and are appearing as new types of discovery tool (e.g. Web engines), primary data (e.g. digital maps), or analysis package (e.g. clustering programs). Going electronic may mean no more than converting a library card catalogue to e-form; but this can instantly make the catalogue far more widely available than before. Digitising historic texts or image archives can open up these materials to many more users and hence research purposes than before. Digitised or born-digital documents, or graphical or numerical data, can be analysed on a scale and in ways that could not be attempted before, for example to study social trends or patterns of language use.

E-resources have significant advantages for research. But they may not win on all technical or practical fronts, e.g. the data organisation or presentation format may be unsatisfactory. There are far more non-e than e-resources, and there are many that are vital for HSS research which are unlikely ever to become digital. The switch to e-resources in STM research has been rapid and, because STM research is typically focused on current technical advances and knowledge, is wide-reaching. HSS research is benefiting from ICT advances generally, and from many e-resources that are not field-specific, e.g. major library catalogues or population databases or statistical analysers. HSS may also benefit from the STM example and from institution-wide e-provision.

But competition for resources, most conspicuously for library budgets, is putting a strain on support for non-e resources that are essential for HSS and are unlikely to become digital soon, if ever, e.g. copies of historical language materials, or librarians knowing about these. This is not just a straight clash between funding for access to STM e-resources versus access to HSS non-e resources. Many STM researchers are still insisting on paper journals, whatever the rising subscription costs, and HSS researchers are demanding both e- and non-e versions of major reference works. The situation is complicated by the fact that e- and non-e versions of the same object have different functionality, e.g. for searching or by being bundled with other resources, so it is difficult to do the cost-benefit analyses to make sensible funding choices.

More importantly, funding for research resources is not just funding for access, such as deciding between another paper copy of a reference book or a CD version of it. ICT should, in general, make research resources more accessible, because once a catalogue or a text is digitised it can in principle be reached from anywhere, regardless of where the original is kept. But access to e-resources is thus also a matter of e-resource provision: e-provision enables e-access. Choices about funding can therefore be between money for more access to existing e-resources or for more e-resources, and between competing candidates for digitisation or digital support, as well as between e- and non-e resource access and support.

## 1.2 The British Academy's interest

It is clear that access to and provision of e-resources for HSS research is complex and raises important intellectual, technological, administrative and political issues. These are compounded by the rapid changes in ICT itself: Web search engines are only a decade old, and digital media presented as long-term storage solutions appear and then disappear with unpleasant rapidity.

The Academy examined the role of ICT in the HSS in ITHS93[1], but primarily from the research point of view. Our concerns here, in this, the Academy's third Policy Review, are much wider. The Academy undertakes Policy Reviews on matters which affect HSS as a whole, have a national dimension, are important in both principle and practice, and where the Academy believes that a Review and its recommendations can contribute to formal policies and community activities.

This Review was prompted by the fact that the nature, provision, and availability of research resources is rapidly changing under the impact of ICT; and that this may have mixed consequences for HSS researchers' access to resources, both e- and non-e. The ICT impact is both direct and indirect. In particular, e-resources may be provided by national institutions or through nationally-motivated policies and initiatives, e.g. the New Opportunities Fund digitisation programme (ENRWWW); and the form of resources may also be affected by national policy responses to views on resource supply in STM, e.g. the open access publication movement.

The Review was therefore designed:

1) to explore the issues raised, for HSS researchers, by the arrival of e-resources, their extent and state, and their likely development;

2) to make recommendations, based on this analysis, about access to and the provision of research resources, especially e-resources, to national bodies, to HE institutions generally, and to researchers themselves.

The Review would thus consider the impact on HSS research of:

a) ICT technical developments affecting resources and access to them;

b) changes in relationships between non-e and e-resources;

c) shifts in STM resource provision that could have consequences for HSS researchers;

d) practices and intentions in national bodies and institutions that host, develop, or promote research resources.

The Review would concentrate on the UK but, since research is international and e-resources cut across national boundaries, would take note of international developments, especially in the US.

The Report, published in 2002 (RSLG02), of the Research Support Libraries Group (RSLG), chaired by Sir Brian Follett, was an important precursor to the Review. It emphasised the need to develop a strategic framework for the future of its research information resources, and called for a proactive collaborative organisation to promote this. Its proposed Research Libraries Network has recently been funded. Other significant inputs include the RSLG User Survey of 2002 (RULI02), showing that HSS researchers depend on a wide range of resources; the work of the SCONUL and JISC committees on scholarly communication (SSCWWW, JSCWWW),

[1] Short citation forms refer to documents or Web sites listed in the *List of background documents and sources of information in* Appendix 5.

indicating a complex and changing electronic publishing situation; and, more recently, the House of Commons Select Committee on Science and Technology's report on scientific publications (HCSC04), encouraging open access models for research outputs, and the Government's response to this (HCGR04).

There have also been recent relevant initiatives, for example the AHRC's new ICT in Arts and Humanities Research Programme (AIAWWW), established to promote effective HSS use of ICT, and the recent establishment of the Digital Curation Centre (DCCWWW) at Edinburgh University, jointly funded by the JISC and the e-Science Core Programme.

The Academy nevertheless believed that an independent, broad-ranging review of e-resources would be timely and useful to the many interested parties.

Other new developments, stimulated by ICT, have occurred during the review period, e.g. Google Inc.'s Scholar system and their striking announcement (GCOWWW) that they are embarking on a gigantic book digitisation project in collaboration with a number of major research libraries. But the pace of ICT change does not make the business of resource review and recommendations just a forlorn chase after moving targets. The resource world for HSS researchers is richer and more varied than the book holdings even of major libraries. More importantly, the best way forward for those responsible for resource provision and support is to develop proactive strategies for this, and not simply to react ad hoc.

## 1.3 Strategy and procedure

The Review was carried out during 2004 by a Working Group chaired by Karen Spärck Jones, with Geoff Smith as secretary. The Working Group's membership is shown in Appendix 1[2], along with its Terms of Reference in Appendix 2.

There are many HSS researchers in the UK, many bodies concerned with resource access and provision, and many activities, initiatives and studies bearing on e-resources. In some cases e-resource operations are natural extensions of non-e ones, e.g. catalogue digitisation. But the effects of ICT are typically wide-ranging and may imply major changes in the form, use and supply of resources, with ramifications from individual user behaviour to national research funding protocols, as the open access publishing movement shows. These changes are not restricted to the resources themselves. They affect major related areas, for example copyright.

The Working Group recognised that there are many parties with an interest in the area of their Review, and a large mass of background literature. The Academy is not able to support major studies, for example ones involving large-scale surveys. The Working Group therefore decided to adopt a focused strategy by:

a) choosing a particular angle of attack on the topic;

b) consulting with key national organisations;

c) gathering further information from other bodies that could be taken, informally, as representative;

d) conducting a user survey designed to cover a wide subject range and a population with research experience and activity;

e) amplifying the information gathered through (a) to (d) by reference to background documents and other ad hoc discussions.

[2] The Review's recommendations are those of the Working Group as a whole, independent of its individual members.

## Angle of attack

We decided to approach our investigations and discussions from the *user* point of view, as represented by five questions, as follows:

Given that HSS researchers traditionally use a wide range of non-e resources,

1) are these non-e resources becoming less accessible, i.e. being squeezed or disappearing (but could they be e-saveable)?
2) are these non-e resources being replaced by e-versions, and if so how accessible are these replacements?
3) are the non-e resources being superseded by better e-forms with enhanced functionality, and if so how accessible are the enhanced resources?
4) are the non-e resources being added to by wholly new e-material and are these novel resources accessible?
5) Then, further, what is needed to ensure accessibility – is it institutional policy or individual behaviour?

## Key national organisations

We met with senior staff from relevant bodies, including the AHRC, ESRC, the British Library, the National Archives, JISC and MLA. Full details are given in Appendix 3. We are very grateful to all of the people who gave us their time.

## Other bodies

We consulted, in particular, a sample of university librarians chosen to cover a range of historical types, physical locations, and institutional sizes. We also consulted some learned societies in different subject areas, and of different sizes, for further perspectives on resources. These bodies are listed in Appendix 3. Again, we are grateful to those who helped us.

## User surveys

There was no existing, widely-based survey data pertinent to our topic on which we could draw. We could also not undertake, for practical reasons, a large-scale survey ourselves. We therefore decided to draw on researchers associated in one way or another with the British Academy itself. We therefore sought to obtain information from three Academy-related classes of user: the Academy's Fellows; the Academy's Postdoctoral Fellows, and the Academy's Small Grant holders.

The Academy Fellows are not representative in age, status, and institutional location; some are also no longer research active. But they constitute, by definition, a sample of people with substantial experience of research resources, and as the survey showed, as users of e-resources as well as non-e ones.

The Postdoctoral Fellows are, by definition, currently highly active researchers, and from a wide spread of institutions.

The Small Grant holders include many other researchers, again from a wide range of institutions, both higher education and other, and also those with no institutional base.

Thus though the survey population is not in any sense representative of the whole UK HSS researcher community, and the survey was not a large one, we believed that the responses to our questionnaire could give us a great deal of valuable indicative information. And this has proved to be the case. We are most grateful to all those who did reply and for the very many, and detailed, comments that they made.

### Other sources of information

We took advantage, opportunistically, of any relevant meetings, visitors to the UK, for example Dr Lynch of the US Coalition for Networked Information, and have considered pertinent background documents. For further details of the latter see Appendix 5.

## 1.4 Definitions

### Subject scope

We have taken HSS as covering all of the subject areas that the British Academy itself takes as its scope: these are all areas and periods of language, literature, archaeology, history, philosophy, law, and religious, political, governmental, economic, social, geographical, anthropological and psychological study, as well as art and music other than practice-led research in the creative and performing arts. But note that we sometimes use the conventional academic abbreviation AH for Arts and Humanities in contrast with SS for Social Sciences. The Review scope is the same as that of the AHRC and ESRC combined, with the exception just mentioned. The forms of research methodology, and research practice covered, and hence resource implications, are correspondingly wide.

### Researcher affiliation and standing

We have focused primarily on researchers in higher education, and in our survey, on researchers of postdoctoral standing. However we have also incidentally been able to gain some information about researchers in similar contexts, e.g. museums, or those doing independent research, and also about doctoral students. We believe that our studies and findings are relevant to doctoral student resource needs, setting aside any special requirements for training resources and, as importantly, to other serious HSS researchers regardless of their formal standing and affiliation, though the latter may have particular, noted, access problems.

### Electronic and digital

Following general usage, we interpret electronic, *e-*, extremely broadly. We use it to refer to anything which is within the scope of ICT, and hence to anything which is held on, treated by, or conveyed through, a computer.

We use *digital* to refer both to materials that have been digitised and to those born digital, except where this particular distinction matters.

### Resources

We recognise that there is no tight, agreed definition of a research resource. Resources are in principle anything that can be used to do research, from a potsherd to a poem to a pen. But we have made some broad distinctions that are helpful for our purposes.

Focusing on e-resources restricts resource scope to representational ones, primarily text or image and hence, for the comparison with non-e resources to analogous ones, excluding objects like pots or miners' picks. We also distinguish *primary resources*, e.g. books, from *secondary resources*, e.g. catalogues. This is a rough and ready distinction but one which is useful because the two types may have very different implications in relation to costs of creation or support, sizes of user community, etc. Thus in particular, many important secondary resources may be e-available while primary ones are not.

Both primary and secondary resources cover a wide range of types. Primary resources include both unpublished materials, like manuscripts, and published ones, like journals. Objects of particular sorts, like maps, may be either. Primary resources include both sets, for example data sets, and individual objects, for instance a specific text. In general here the presumption is that

primary resources are collections, perhaps just aggregations of independent objects, perhaps complex assemblies with a rich internal structure. As indicated, primary resources can include images and also recorded music, though we treat museum-quality prints as outside our scope.

Secondary resources include not only catalogues but also bibliographies, abstract journals, citation indexes, i.e. in general, types of resource that lead to, or characterise, primary resource collections. Thus our broad distinction is between resources *on* which research is done, i.e. primary resources, and resources *through* which primary resources are reached, i.e. secondary resources. As this implies, both types of resource are essential for research.

Our working definitions of primary and secondary resources are further illustrated in the table below. Our distribution of particular resource types may be disputed: surely a colleague is a primary, not a secondary type? Individual objects may shift type: thus a historic figure's library catalogue may be a primary resource. Nor is our list of types exhaustive. But we have found the broad categorisation valuable. Equally, while books or papers are the products of research and are thus derivative, we have chosen to treat them as the primary resources they are for other readers. Other derived objects, for example annotated text corpora, are also taken as primary.

| *Primary resource examples:* | *Primary examples continued:* | *Secondary resource examples:* |
|---|---|---|
| Correspondence | Photographs | Library catalogues |
| Institutional records | Maps | Archive catalogues |
| Field notebooks | Music | Museum catalogues |
| Manuscripts | Video recordings | Bibliographies |
| Working drafts | Audio recordings | Abstracts journals |
| Theses | Films | Subject indexes |
| Incunabula | Satellite images | Classification schemes |
| Printed books | Sensor logs | Citation indexes |
| Pamphlets | Questionnaire returns | Web portals |
| Serials | Field surveys | Search engines |
| Conference proceedings | Numerical data | Colleagues |
| Official publications | Statistical analyses | |
| Reference compilations | Formal databases | |
| Reports | Museum object records | |
| Graphics | Text corpora | |

In general, the types of resource mentioned can exist in either non-e or e-form (even colleagues may be remote, never-met, e-colleagues). But ICT has brought some new types, notably through the World Wide Web: these include both primary resources consisting of linked page sets, and secondary resources in the shape of Web portals that are not much like their non-e precursors. Our Review is thus concerned with novel resource types as well as the relation between non-e and e-versions of familiar types.

Some secondary resources, like catalogues, are often referred to as *discovery tools*; and when in electronic form are so intimately tied to the processors, like search engines, that exploit them, that we have included this type of processor as a secondary resource. At the same time, the effective use of primary e-resources may depend heavily on *e-tools* like concordancing or statistical packages. We have taken access to such e-tools as within our remit, though with less

emphasis, since they are not integral to the primary resources, or to the secondary resources without which the primary ones cannot be reached, that are our main concern.

Note that in referring to resources, we refer to things that have content, not to the content itself. In general we are not concerned directly with resource content independent of its embodiment: so, for example, two copies of a book with the same content but in different libraries are treated as two resources. Similarly, for our purposes, we treat the *e-forms* that *e-content* takes as resources, regardless of their underlying relationships. As this implies, the relation between resources and intellectual property is complex.

## Provision

We interpret resource *provision* to mean, broadly, making resources available to users, in this case researchers. Thus libraries may provide non-e resources and digital libraries e-resources. Provision is independent of original resource *creation*. The author of a paper is its creator. The author may also provide the paper text by posting on a Web site, a publisher may provide it in some form, and a library may provide it. However as secondary resources like catalogues are vital research resources, we include catalogue *production* and other such list or description compilation within provision rather than creation; but we treat database authors as creators. These distinctions are clearly not watertight and are problematic for intellectual property, but are suited to our main purpose. We do not assume that resource providers are either content creators or intellectual property owners. We are concerned with how resources, once created or produced, are provided (supplied or mediated) for use.

Thus approaching our theme, access to research resources, from the resource point of view, our concern is with the form, and adequacy of provision and, more specifically with e-resource provision as adding value to traditional non-e provision.

## Access

We interpret access broadly. At bottom, a resource is accessible if you can look at it (or hear it). But reaching this point may be a matter of *physical access* e.g. can I get to the library*? logical access*: can I find what library it is in? or *administrative access*: am I allowed in the library? These distinctions of access mode cover a wide range of specific possibilities, e.g. for physical access ranging from going to my department's library for a copy of a journal to crossing the world to reach a special archive. The modes also apply to both non-e and e-resources, though they may take different forms. Thus for e-resources physical access may be to a display on a computer screen, i.e. strictly to a virtual object, not to a physical original. Physical access from the user point of view thus applies to whatever form the resource takes to be usable, regardless of whether it is a unique non-e or e-object or a copy.

Non-e and e-resources obviously present different challenges for physical access, like a long journey versus mastery of a laptop. The same applies to logical access. Thus manipulating an online search engine, for example, requires an ability to drive a computer as well as the ability to read, though the ICT know-how required may not be very taxing. But it also, and more importantly, requires an understanding, e.g., that the way a search engine works when string matching in searching for a title word does not necessarily simulate what humans do when scanning a catalogue. ICT may not merely replicate traditional non-e forms of logical access even when it is assumed to do so. Thus while the discovery tools needed for logical access can in principle be the same whether they are non-e or e-, in practice ICT can offer new ways of manipulating traditional tools as well as some distinctively new tools, like URLs.

Administrative access is taken here to include formal access, as governed by, for example copyright, as well other access controls not necessarily justified on the basis of intellectual property, for example status. Administrative access can in principle take the same form for both

non-e and e resources, and whether this is about permission to access or charges for access. However when, for example, what constitutes a copy may not be quite the same operationally for non-e and e-resources, this can lead to rather different treatment for non-e and e-versions of the same thing. Thus non-e versions may be accessible to all users of a library, including visiting researchers, but e-versions open only to institution members.

The administrative controls on access can differ for a secondary resource and the primary resource to which it refers, for example there are no controls on using a catalogue but restrictions on who can read a book for which it contains an entry. The same applies to e-resources, where anyone may be able to check the table of contents of a new journal issue, but not to access, for free, an article that is only a mouse-click away.

Access to resources for research is the precursor to research and research is about using resources, not just reaching them. The presumption that getting to a position where it is possible to read a text, inspect a set of numbers, or look at an image constitutes using these resources is manifestly inadequate even though, at the same time, reading, inspecting or looking at are themselves resource uses. We have nevertheless taken the rather informal notion of being able to read, inspect, look at, etc., a resource as constituting access to it.

In considering resources for HSS research from the access point of view, therefore, we have focused on the form and severity of barriers to access. We have been especially concerned with barriers to e-access where e-resources are of value to researchers, but also with any untoward effects that a shift to e-access may have for non-e access.

## 1.5 UK and non-UK resources

We have not so far distinguished resources created in the UK from ones created elsewhere. Some national bodies, like the National Archives, may be wholly or primarily concerned with resources originating in the UK, and the UK Data Archive is grounded in UK-developed resources. Moreover though in the non-e case libraries have long been mediators for resources created elsewhere, for example through purchasing foreign books and subscribing to foreign journals, from the user's point of view the operational effect is much the same as if these resources are UK ones.

But ICT is changing the UK/non-UK division in radical ways that have already altered researcher access to resources and have wider implications for the funding and organisation of resource provision and access.

In general, with non-e primary resources, whether they are formally published or not has been most significant for ready access. Individual researchers, or institutions, could obtain copies of published resources and so bring these nearer to their users. With non-published resources, on the other hand, the user had to go to the resource. With secondary resources the situation was similar. Libraries might hold printed catalogues for non-UK resources, or abstract journals, making the information in them somewhat more accessible, but card catalogues had to be visited at their location.

The initial impact of ICT was on secondary resources, so these became more readily available to UK researchers, though not necessarily for free or outside institutions. UK institutions like national or university libraries thus continued to be mediators of non-UK resources, with the addition of JISC as a licence negotiator. Since then there has been a very rapid increase in the supply of both secondary and primary resources, some free but some not free, and this has been accompanied by the spread of network connections so they can be reached from the researcher's desktop.

It is therefore necessary to consider how differences between UK and non-UK origins for e-resources affect institutional roles not only in supporting UK resources but in mediating non-

UK ones. In particular if, as appears to be the case, many more such resources, as e-resources, are becoming available for free and can be reached immediately through Web engines, what form should UK institutional support for research resources take (beyond sustaining internet connectivity through JANET)?

The researcher survey indicates that researchers, with Google's help, are roaming the Internet and finding and using resources regardless of national origin; and an analysis of the British Academy's research PORTAL site (BAPWWW) suggests the same. PORTAL, which covers only free sites, refers to some 600 major free resource sites across all HSS subject areas – 80% of these are outside the UK, with about 51% in the US.

We cannot attempt to review these non-UK resources systematically, though they are increasingly important for UK researchers. But we comment on developments in the US and Europe that are particularly pertinent to our concerns in later sections.

## 1.6 Conclusion

Overall, our review of researchers in Section 3 and of institutions in Section 2 shows that researchers are making a continually increasing use of e-resources and that UK institutions and organisations are working to increase the supply of e-resources and improve access to them. Non-e resources, especially books, remain important for HSS researchers generally and vital for some. But it is becoming easier to discover these with the secondary e-resources that are appearing and are themselves relatively accessible. E-forms of primary resources that are published, especially journal articles and papers, are spreading, though access to these may be more problematic. The great bulk of unpublished primary resources, text and image, remain in non-e form, but most subject areas are benefiting from the arrival of digitised versions and also from new, born-digital resources, and processing tools. The spread and gain is very variable, and there are access problems here too. Since e-resources can originate anywhere, variety in what appears is a fact of life. But this does not preclude more coherence in the creation of UK e-resources and organisation of access to them. The formal and informal connections between the various bodies constitute a somewhat Topsy-like jungle with mixed implications for readership and coherence on, or priorities and coverage of, resource access and provision.

In conclusion, we can illustrate the opportunities that e-resources offer the researcher using two very different subject areas, namely linguistics and language study, and geography. We can also see, from another point of view, how two types of resource, when available in e-form, namely news material and art images, can enhance the research platform across a wide range of subjects.

### E-resource value illustrations

In linguistics and language studies, digital versions of text material of linguistic as well as literary interest were early entries in the e-field, and the supply of material with specific properties, like the *British National Corpus*, which was designed to have genre coverage and balance, is continually growing. Corpora with research-relevant annotations, for example parsing data, are also becoming available, along with electronic dictionaries. Major resources that have been specially designed and created may not come free, but may not be expensive. At the same time, the Web itself constitutes a vast linguistic resource, and is now routinely trawled for examples of language usage. Much material is free, absolutely or on a fair use presumption. These text resources are increasingly available for many languages, and ICT developments generally have made it possible to handle a wide range of scripts in e-form. As well as these primary resources, there are important secondary ones in Web sites and also publicly available tools like concordancing programs.

In geographical studies, digitised maps, satellite photographs, and geographical information systems have revolutionised the field. Conventional resources are more readily available and in a much more flexible form that supports image manipulation and analysis; wholly new types of resource have appeared, like heat-recording images; and there are many programs for data condensation and interpretation, including sophisticated visualisation. Much of this material has the advantage that it naturally covers wide areas, indeed the whole globe, in the same style, so it is much easier to study geographical information without the artificial limits imposed by national boundaries that are often reflected in the style and content of non-e resources. Much of this material originates in the US, but access conditions are not generally perceived as a barrier to research.

If we take the opposite point of view and consider a particular type of e-resource, e-news material, the fact that it is readily available, at least visually if not always with free, downloadable and manipulable source, make it a far more accessible research resource across a very wide range of subject areas, in fact all. Thus quite apart from 'modern' study areas including contemporary history, politics, economics and law, along with literary and language research, it also offers at least some items of interest to, for example archaeologists and anthropologists.

In the same way, an e-image, say of a statue in a public square, can constitute a research item for art and architecture study, and historical, political, social or cultural analysis, not only in terms of the character of the statue itself, but of its setting and use or misuse. Image data, especially when owned by museums and galleries, may be expensive, but not all is.

## Structure of the report

In the next Section, 2, we review the activities of national organisations and other bodies, focusing on their e-resource operations and policies.

In Section 3 we present the results of our researcher survey.

In Section 4 we examine the themes and factors that underlie, or emerge from, the information and analyses of sections 2 and 3.

Finally, in Section 5 we consider the issues that our Review has raised and offer our recommendations for action or further investigation.

# Section 2: Institutions and Organisations

## 2.1 Introduction

There are a number of national bodies that are important for researchers because of the resources they hold; their role in national coordination and collaboration activities; their actual or potential leadership in introducing new methods and technologies; and their status as a factor in making international connections that can enhance resources, improve access to resources, and facilitate technology transfer.

Some of these, like the British Library and the National Archives, are major resource holders with a far wider scope than HSS alone, but they are important, through their historic holdings and continuing intake, for HSS. Other national resource holders, like the UK Data Archive and related Economic and Social Data Service and the Arts and Humanities Data Service, are important specifically for HSS.

Other bodies are important primarily as funders in varied and sometimes complex ways. Thus the Arts and Humanities Research Council and the Economic and Social Research Council are important because of their support for individual researchers, with their resource consumption and production; because they mandate or encourage resource deposit in archives; and because they support resource facilities in various ways. They are also concerned with generic issues and infrastructure.

In all this, the Joint Information Systems Committee, as a technical agency with a substantial budget, plays a central role through its responsibilities for the national computer network and for data archives support, and through its role in a range of resource-related initiatives and projects.

There are other national bodies, like the Museums, Libraries and Archives Council, which have a strategic advice role that bears on research resources.

The formal and informal connections between the various bodies are complicated and seem to have developed, ad hoc, over time. We note the substantive formal connections where appropriate, but comment on the implications of this complex situation for resource access and provision in Section 5.

We consulted the bodies just mentioned as clearly important for our Review. There are many other bodies that are significant resource holders, including the National Libraries of Scotland and Wales; the other copyright deposit libraries; other major research libraries; specialist bodies like the Ordnance Survey. Our concern with research resources in higher education means that university libraries in general matter, whether as resource holders or channels of access. Museums and archives across the country are valuable contributors to the resource pool, along with some learned societies.

We could not attempt to cover representative samples of all these bodies. But we have consulted an informal sample of libraries of various types, and of learned societies, through which we believe we have obtained a fair picture of the resource situation and one which covers a wide range of user access needs and resource types. Thus in particular, while we have not consulted museums to more than a very limited extent, nor local archives, we have obtained information bearing on non-text and archive resources. The complete list of bodies consulted is given in Appendix 3.

In this section we review:

1) the major national resource holders and supporters;
2) the situation in university and other research libraries, as seen from exemplar libraries;
3) the position of learned societies, as seen from an illustrative sample;
4) other data on resources, from various sources.

It is evident from the survey reported in Section 3 below that UK researchers make heavy use of resources outside the UK. Some UK resource centres, e.g. ESDS International, provide pointers to non-UK resources, and a recent review of the British Academy's PORTAL site shows that more than three-quarters of the sites included are outside the UK, and about half from the US. We could not attempt to explore external resources, but emphasise their importance to HSS researchers and the consequent need to ensure that UK researchers can access these resources and can, reciprocally, make the resources they create available to others in their wider research communities.

In the sub-sections which follow we consider the different institutions and organisations only from the research resource point of view and so ignore, for example, the British Library's role as a cultural flagship, JISC's functions in relation to further education, and university libraries' role in teaching support.

We much appreciate the time and informative detail that those we consulted gave us, and are grateful for their help.

## 2.2 Major national resource holders and supporters

### a) The British Library (BL)

The British Library is the national library of the UK and its leading research library. Research support, primarily for the academic and business sectors, is taken as one of its primary functions. Its roles are both direct, through its own holdings and services, and indirect through leadership in example and collaboration. It is currently engaged in major strategy reviews, for STM, for SS, and for AH. The impact of e-materials and e-operation on the Library's basic functions, collecting, preserving, cataloguing and making available materials, both individually and collectively, is an important element in this. The aim is to bring these strands together to produce a single corporate strategy in 2005.

The BL has already enhanced its operations, using ICT, in ways that are recognisably of value to the whole research community, notably in introducing its new online integrated catalogue, which for the first time provides cross-searching of its collections at both St Pancras and Boston Spa. It is also operating scan-on-demand digitisation across its collections for its document supply services, with electronic delivery of items on a pay per view basis, and with online ordering. The relative distribution of the Library's e-activities between primary and secondary resources is an important question.

HSS materials, including journals, monographs and report literature, but also sound, music and cartographic materials, consume some 45% of its current acquisitions budget, compared to some 55% for STM. The volume of material received through legal deposit has continued to grow. After cutbacks in the 1990s the BL has managed to maintain its purchased acquisitions levels though as the number of publications produced worldwide is increasing, the proportion of international output the Library takes is declining. At the same time, the Library recognises that the character of its historic holdings means that HSS users will continue to be a major customer constituency.

Most of the BL's materials are, and are likely to remain, in non-e form. The Library is not yet prepared to transfer resources from existing activity to fund large-scale digitisation of its collections. While significant digitisation is taking place, it has been as an additional activity, mainly funded externally and often in partnership, e.g. a current newspaper digitisation project in collaboration with JISC, or digitisation of its Shakespeare Quartos with a commercial partner. The Library also receives digital material through licensing, purchase, and voluntary deposit.

The BL's new catalogue management and access system has much greater potential for interoperability and linking than its predecessor, and the Library continues to develop its article-level resource discovery tools based on table of content information. It is expected to continue to improve its electronic catalogues and to support national union catalogue efforts.

The BL is currently engaged in major activities exploring appropriate ways of managing its digital materials over the long term, given the constraints imposed by publishers' control over sources for which they own intellectual property rights (IPR). Similarly, it is exploring ways of acquiring digital material automatically from the Web through selective harvesting, and also the automatic acquisition of metadata from digital materials. The distinction between image and character files is practically important for content search and has significant implications for metadata quality and the BL's long-term strategies for content access.

The BL aims to ensure long-term access to digital material which it has selected for retention by purchase, deposit or internal creation. It recognises that digital material preservation differs from non-e preservation not only because it requires far more active and continued intervention to ensure survival, but also because it requires preservation decisions when beginning the life cycle management of the digital object. The BL aims to avoid loss of access to digital material, either through media deterioration or through technological obsolescence in hardware or software. This means that both the physical and the information aspects of the material have to be taken into account. The Library was a founder member of the UK Digital Preservation Coalition and is actively collaborating with UK and international organisations that are leading digital preservation research and development.

Digital rights management is already an issue for the BL, and can be expected to become a more major one, given both the Library's scale and its status as a copyright repository. The current legal situation, and developments which are increasing source owners' rights, have ridiculous consequences for users through the arbitrarily different constraints and charges they apply to reading an item under paper and digital regimes. The legal position also has a deleterious effect on efficient resource management. Thus the digital image that is made for the scan-on-demand service cannot be preserved even as an internal master to generate new on-demand copies and to reduce wear and tear on the non-e originals. Similarly, while electronic publications will be received under voluntary deposit and subsequently under legal deposit arrangements, access to this material will be limited to users in the Library's reading rooms.

The BL is clearly seeking to combine its traditional role as a holder of non-e materials with a new role as a holder and provider of e-materials, and seeking to enhance both through improved cataloguing and catalogues. It aims to support the HSS community and this community can benefit from what it is doing. The Library is naturally cautious about simply becoming a part of the seamless Web world, rather than a resource that can be reached via the Web. How innovative its future e-strategies should, can, or will be are important questions.

In reviewing the BL's subject-based responsibilities, it appears possible that, in relation to its constituent subject areas, it may emerge with rather different modes of operation, e.g. DIY in one, partnership with, say, large commercial partners in another. It is clearly essential, however, not to cast any such differences which appear currently reasonable in such operational 'iron' form that it hampers research, thus reinventing in ICT form divisions that formerly existed in physical space.

The BL's leadership may be evinced by specific exemplar projects such as Web archiving, but it sees its leadership as much more importantly manifested by taking the initiative in strategic collaborative activities that benefit the provision of research resource for the UK as a whole, or in acting as a representative on behalf of the whole community.

The Library is well placed to do this, given the clout of its importance as a major library, its formal position as national library and a copyright repository, and its human and financial resources. It can also seek to ensure that national strategies take account of the importance of public libraries and specialist subject libraries such as those of learned societies, as collectively significant holders of research resources, to complement the higher education library community.

Thus it is important that the BL uses its muscle and experience: (a) to tackle the increasingly bizarre intellectual property (IP) situation and (b) to inject longer term thinking (i.e. realism) about sustainability into the many UK resource and service activities that are proliferating under every kind of hat. Preservation is one element in ensuring continuing value for e-resources and tools. But far more than this is required for proper sustainability. Thus while, for example, JISC is seeking to innovate, primarily from the technology point of view, in the area of resources and tools, the Library could supply the crucial long-term perspective.

### b) The National Archives (former Public Record Office and Historic Manuscripts Commission) (TNA)

The National Archives are very positive about the importance and benefits of e-resources, and are committed to them. Most current national government records are now produced in e-form ('born-digital'), and those which are scheduled for preservation are transferred to TNA in that form. Some (e.g. documents for public inquiries or required under Freedom of Information provisions) have to be able to be made available immediately and not simply preserved for future public access. Older records are being digitised where there is a clear user need and funding or business model to support the necessary investment, e.g. the 1901 Census. Such funding is often by specific partnerships. Much of the access, and related willingness to pay, is from family historians, and although the material they use is also important for other researchers, recommendations and requirements of the HE sector were not implemented in the access service for the 1901 census.

TNA staff emphasise the crucial role that being digital plays in making their material, defined as (actually or logically) unique objects, more accessible to users. They are able to keep charges down, at a level analogous with paper copier ones, and to provide a high level of support service.

Resource discovery is through TNA's catalogues. Government records are received with cataloguing and other information reflecting the business processes of the originating body. This is reused by TNA, with the addition of necessary housekeeping information. Some entries may reproduce significant quantities of source text which can be searched by the catalogue tools. The catalogue, PROCAT, is electronic. TNA leads and hosts the nation-wide Access to Archives (A2A) programme. This is extremely valuable, providing information on and allowing searching across the catalogues of over 340 archival repositories in England (though these local archives may not be comprehensively catalogued).

TNA is addressing the issues of preservation for digital material, envisaged as needing a mix of migration and emulation, though it is not clear that the requirements for enabling software support have been fully recognised. They are already preserving Web site snapshots in collaboration with the Internet Archive, and are considering the implications of long-term Web site preservation.

TNA is able to get leverage in acquiring resources and making them available from its national status and legislatively endorsed functions. It has been proactive in going digital, seen as a clear win for users. In TNA's view, the need now is not for the TNA to supply but for the users to demand: users need to wake up and grasp the rich resources that TNA is already making readily available.

### c) The Museums, Libraries and Archives Council (MLA)

MLA has general oversight, responsibilities and core funding from the Department of Culture, Media and Sport, for a large number and wide range of institutions and organisations in England. It has a particularly significant role, for the many local MLA bodies, in encouraging and connecting this collectively important set of resource providers and in facilitating access to their holdings. Thus while MLA is not the prime funder for these bodies, it has a notable impact on e-resource provision and access through specific digitisation programmes such as Enrich UK, through resource discovery efforts such as Cornucopia, and through standards and guidelines development. It works closely in e-matters with JISC.

MLA's particular focus is on the public sector environment, and in supporting access to and the use of knowledge (as embodied in MLA materials) as a public service. It has a mission to encourage learning for the 'ordinary user' and a vision of the future for the many who want to know, where users will be able not merely to reach publicly-available information or knowledge resources but get feedback that is automatically customised to the individual, for example for their time and place. The goal is to bring resources out of the backroom so citizens can relate them to their own needs and thus benefit from them. This resource user community may include higher education researchers, but the way in which resources are organised, presented and reached is not designed primarily for them.

To support its existing activities and enable its future vision of the Knowledge Web, MLA has been working together with other organisations like JISC, TNA, BL and BECTA on the Common Information Environment, aimed at improving interoperability and access to e-resources across sectoral boundaries. The Knowledge Web is a planned MLA community-wide secondary resource for which pilots are currently under way. MLA is also active in work on standards, for example exploiting Dublin Core for metadata. MLA takes it as given that the resource provision it seeks will only be effective if it is based on the assumption that access will be via the kinds of search mechanism, like that embodied in Web engines, with which people are already familiar, even if it is not directly through actual engines, because these will not necessarily relate resources to one another in the way envisaged as helpful to users. MLA's strategy also implies a relatively lightweight view of metadata as ordinary-user-friendly metadata.

MLA is working with many parties, in Europe as well as the UK, placing particular emphasis on coordination, for example in digitising primary resources or in adopting common protocols for e-metadata. Thus its influence on resource provision for HSS researchers, though indirect, is pervasive.

### d) The Arts and Humanities Research Council

The AHRC is one of the two main funding bodies for HSS research in the UK, with an annual budget of ca £85 million in 2005. The AHRC funds research and postgraduate study within higher education institutions and provides funding for museums, galleries and collections that are based in, or attached to, higher education institutions in England.

It supports research through its research grants scheme (for individuals or teams of researchers), research leave scheme, and resource enhancement scheme (for projects aiming to improve access to and use of research resources and materials), as well as its schemes for

postgraduate awards which provide funding for students undertaking masters-level courses and doctoral research in the arts and humanities.

The AHRC has a significant commitment to the provision of research e-resources, both on its own and in collaboration with other national bodies such as JISC; and its role is both direct, e.g. through funding support for the Arts and Humanities Data Service (see below), and indirect, through its resource enhancement scheme and other research grants which may lead to the production of resources. As with other funding bodies, it is concerned that research resources produced on grants are not subsequently lost through failures to preserve and offer adequate access to them. These resources may be both primary resources like data collections, and secondary resources like catalogues.

The AHRC has been substantially developing its activities, and adopting a proactive leadership role, in relation to ICT. Thus it has established a four-year ICT Programme, with the aim to encourage, support and enhance the use of ICT in all areas of arts and humanities research. The programme will provide support for the development and use of digital research and tools, and for work on ways in which ICT can be used to enhance and promote a more effective research process and to communicate research outcomes more widely and efficiently. It will also fund a range of projects or services that will complement those already funded by the AHRC and other national bodies. As part of the programme it will establish an ICT Methods Network intended, for example, to promulgate good practice in resource development and use.

Again, through the HUMBUL humanities hub (HHHWWW), which it co-funds with JISC, AHRC is participating in JISC's Resource Discovery Network, fostering subject hubs with catalogues and navigation tools.

More generally, AHRC is developing relationships with key resource stakeholders such as the British Library and the new Research Libraries Network; is seeking to learn from pertinent ICT initiatives outside HSS, like the e-science one; and is participating in a variety of ways, including expert seminars, in discussions on significant ICT issues including electronic publishing, and digital preservation and curation.

Thus one of the AHRC's long-term goals for ICT is in providing scholars with seamless access to a wide range of data of different types and in ensuring that scholars are able, through information and training, to exploit these.

The AHRC is clearly becoming materially engaged with e-research and e-resources; is anxious that existing e-resources like those held by AHDS should be enhanced, better known, and further exploited; and is seeking, amidst a miscellaneous set of bodies concerned with, and a heterogeneous mass of initiatives on, e-resources, to achieve more effective coordination and focus than has hitherto been the case.

However when many HSS researchers are not research grant holders, AHRC faces a significant challenge in seeking to raise HSS researcher consciousness about e-resources, whether in contributing to or making good use of major e-resources that AHRC sponsors. The AHRC's new ICT Methods Network is intended to tackle this problem, and AHRC and JISC have a collaborative Awareness and Training programme, though pervasive consciousness-raising may take time.

### e) The Economic and Social Research Council (ESRC)

ESRC spends some £82 million a year (in 2004/05) on support for research, including grants and its programmes and centres, and some £32 million on supporting postgraduate training.

ESRC has long played a major role in UK SS resource creation and use, especially for primary resources. It has encouraged dataset deposit from its funded projects but has also acquired others, and through the Economic and Social Data Service provides access and support for an extensive range of key economic and social data, both quantitative and qualitative, spanning many disciplines and themes; its datasets include multimedia materials. It has been actively engaged with preservation and curation issues, including those associated with data from major longitudinal studies, with rational policies about IP, charging, and also privacy. It co-funds with JISC the SOSIG Social Science Information Gateway (SOSWWW) as part of the Resource Discovery Network.

ESRC has just embarked on a substantial and ambitious five-year programme aimed at providing a first-call Information Centre which will provide a gateway for the social sciences, enabling researchers to identify resources over the whole SS area. This virtual collective resource would cover both primary and secondary resources, with the primary including both grant-associated publications and other information and data in the conventional sense (both quantitative and qualitative), and the resource would reach far beyond ESRC's own-funded materials to provide leads to resources world-wide. Development and operation is through a contracted commercial service provider working in partnership with ESRC. Access to resources for researchers is primarily through metadata, as conveniently obtainable, but the larger aim is to offer facilities for direct end-resource (e.g. full text) searching. The Centre is envisaged as catering for a much wider clientele than academic researchers, but would naturally have much to offer researchers. ESRC recognises that there is an issue in identifying potentially useful materials for a given subject area within a very wide assembly, when starting searches outside the Centre and not knowing of its existence.

ESRC is working with the other research councils to establish common approaches towards open access. This is an area of recent public interest. A joint position statement from the research councils is expected in spring 2005. In principle there is support for open access to research outputs, and it is likely that the councils will either request or require that grant holders deposit a copy of their articles to an open access repository, and may suggest a review of relevant IP matters. Views on what archives to use for depositing material could vary between the research councils, for instance some would be able to host material. It is recognised that strategies for choosing repository types (e.g. institutional versus subject) and for ensuring repository sustainability in the long term need far more investigation.

### f) Joint Information Systems Committee (JISC)

JISC (The Joint Information Systems Committee) is an agency of the UK Higher and Further Education councils. It has roles in supporting both teaching/learning and research. Its aim is to support HE and FE communities in everything related to ICT that is sensible to do at the national level. For present purposes we consider only its research-related roles, which are also primarily pertinent to the HE sector. Its budgets are ca. £62 million a year, of which ca. £32 million is for the JANET network.

JISC's operations are relevant to research at two levels. Its responsibility for JANET makes it the key e-infrastructure provider for HE institutions. This contribution to research support is enhanced by the guidance, tools and assistance it provides to computing services and their staff. However JISC also has significant other roles in relation to the support of and access to research resources.

JISC negotiates with e-content providers for access to e-content, sometimes on a nationwide basis, sometimes more specifically. JISC uses various models for the licensing of content. In some cases it simply negotiates terms at the national level, in others it licenses the content and

then sublicenses it to the individual institutions. In most cases the aim is to ensure cost recovery and thereby sustainability over the life of the deal. For journals it has used an agent to negotiate with ten key publishers or groups of publishers. In each case the deal covers all of the publisher's electronic journal output. Institutions were consulted to identify and decide the ten candidates. In the case of the Early English Books Online (EEBO) database, the content has been licensed for access in perpetuity by UK HE, with institutions paying a subscription for access and hosting costs, but not for the content itself. While UK researchers may exploit many other resources than the JISC-enabled ones, JISC does play an important part in helping to make key resources more accessible to researchers.

It is open to any institution or, indeed researcher through their institution, to propose resources for JISC to consider; JISC's criteria for resource adoption include long-term sustainability, and it adds value by, for example, defining and securing publisher agreements to a sensible, uniform licensing model.

However while JISC may make arrangements for access that may be applicable in principle throughout the UK, it does not follow that particular institutions choose or can afford to participate through committing their funds to licensing some resource. As a result the ability of individual researchers to access a specific resource can be very variable in practice. This is unsatisfactory: there is no good reason why Dr A in University X should be able to use resource R when his colleague Professor B in University Y cannot.

JISC has recently participated in some major digitisation projects, through one-off funding received from central government, but there are not likely to be immediate successors to this programme. It has addressed e-preservation issues and is also involved in the larger questions of digital curation, especially in connection with e-science resources but with wider implications. With the Core e-Science Programme, it has recently funded the setting up of a Digital Curation Centre based at Edinburgh University to carry out research and provide advice on best practice. JISC has developed particular expertise in handling image and multimedia resources, important for HSS as well as STM.

JISC has responsibilities for and co-funds a wide range of content services, including major resource centres and data services such as the UK Data Archive, the Arts and Humanities Data Service, national data hosting centres at MIMAS and EDINA, and also the subject portals making up the Resource Discovery Network, including the HUMBUL and SOSIG gateways for the humanities and social sciences respectively. Some of these services manage content resources that are significant in the longer term for HSS researchers, but it is possible that some of the other discovery services may be being superseded by, e.g., Web engines.

JISC's content services illustrate the problems that UK strategies for e-resource management have to tackle – namely sustainability in the longer term, the real value of deposited data (including deposit mandated or encouraged by funding agencies), and the nature of and responsibility for metadata and other supporting documentation, as well as the comparative importance, for a national body such as JISC, of concentrating on the provision of and access to secondary or to primary resources.

Thus one view is that while it may be appropriate for JISC to act as an enabler or manager of primary resources that are independently motivated and funded, it has at least as significant a role in improving secondary resources, or access to them as embodied, for example, in having a single point of seamless access to multiple library catalogues. This in turn leads to the crucial question of resource characterisation in relation to access paths (metadata etc.).

JISC is heavily involved in the UK's initiatives in e-science with the Grid. The e-science initiative came from STM as a response to the reality of research communities with interests in shared resources, processing tools and experiments, i.e. in research that needs and benefits from the scale of work and synergy that dynamic electronic operation allows. HSS is not in general calling for this kind of facility, but can in principle benefit in the longer run from what is learnt about how to sustain and do this type of research. HSS researchers may properly, rather than 'unfortunately', work alone rather than in close collaboration, though the value of shared data resources is evident from the range of studies that have already been done using the *British National Corpus* or *Penn Treebank* in linguistic research. JISC, in company with AHRC and ESRC, emphasises the importance of thinking in terms of e-research rather than 'just' e-science, and some subject areas, especially in the social sciences, could become leaders in HSS participation in use of the Grid. Some initial projects, for example in the text mining area, are indeed now beginning.

JISC supports a large number and wide range of projects across the spectrum from technology studies, e.g. in relation to middleware, to experimental services such as electronic theses management and the development of major new facilities such as the SUNCAT serials union catalogue. The technology studies are important in investigating, on behalf of the community, the many and challenging issues that arise in a rapidly changing ICT scene in connection with access to, and provision of, resources including, for example, authentication, interoperability, format viability and storage developments. In the absence of established technologies, a UK body such as JISC has a proper role in acting both as a participant in an area crossing national boundaries, and as a more specific interface for UK parties, in such activities. The service and similar experiments can throw light on the implications of service design and continuity requirements.

Thus repositories are clearly an important arena for JISC, evidenced by its support for and funding of the FAIR (Focus on Access to Institutional Repositories) programme (FAIWWW). As e-resources grow, the technical as well as political issues of repository design and organisation, for the long term, are becoming increasingly pressing. Ensuring e-resources will be viable over time involves far more than preservation in the narrow sense, and even curation: it requires adequate models for sustainability. JISC's remit and the know-how arising from its projects make it a significant contributor to the development of thinking and best practice here.

There are nevertheless two aspects of JISC's activities which raise questions about good ways forward on e-resources for the UK, and HSS in particular. One is that the combination of JISC's substantial budget and visible status make it all too natural, especially in technical matters, to devolve investigation and evaluation to JISC by routine default. While this is justifiable in general and often invaluable in particular situations, the pervasive assumption that anything that involves ICT is a primarily technical matter does not always lead to the degree of interaction with envisaged beneficiaries that is really desirable.

The second is that in relation to many specific, less exclusively technical projects, it is not always clear what their longer term value is. Thus while it may be appropriate to fund a project to ascertain whether a viable prototype for some resource can be found, it appears some projects are over hopeful: for example some resource services are established on the basis that it must be a good idea to have an X, but without sufficient analysis of outcome value for money.

Thus JISC operations raise important questions about the role of UK funding and organisations in creating and sustaining UK-based and oriented e-resource operations. E-resources are in virtual space, and researchers pull in whatever they can from wherever they can. Thus it is essential, when resources with a distinctively UK connection are proposed, to justify the UK

focus or locus. Digitisation of UK non-e resources may be one good reason. Making the results of UK-funded research more widely available may be another. But it may be far less clear, for example, whether there is a good case for supporting some subject-based Web site limited to UK materials or oriented primarily to a UK clientele. Collaboration with other parties or countries to take responsibility for particular remits as part of a general distribution of effort for resource support, internationally, is another matter.

### g) The UK Data Archive (UKDA), and Economic and Social Data Service (ESDS)

The UK Data Archive (UKDA) is a centre of expertise in data acquisition, preservation, dissemination and promotion, and is curator of the largest collection of digital data in the social sciences and humanities in the UK. It is funded by ESRC, JISC and the University of Essex, where UKDA is based. Founded in 1967, it now houses several thousand datasets of interest to researchers in all sectors and from many different disciplines.

UKDA is a lead partner of the Economic and Social Data Service (ESDS), which is a national data archiving and dissemination service that came into operation in January 2003, jointly-funded by ESRC and JISC. It is a distributed service, based on collaboration between UKDA and the Institute for Social and Economic Research (ISER) at the University of Essex, and MIMAS and the Cathie Marsh Centre for Census and Survey Research (CCSR) at the University of Manchester.

ESDS comprises five specialist data services that promote and encourage data usage in teaching and research. These are: ESDS Access and Preservation, hosted at the UKDA; ESDS Government, which provides access to the large-scale government surveys, such as the General Household Survey and the Labour Force Survey; ESDS Longitudinal, which provides access to and support for a number of major longitudinal data collections such as the British Cohort Study; ESDS International, which provides access to, and support for, a range of international datasets, and is run jointly by MIMAS and UKDA, and ESDS Qualidata, which provides access to and support for a range of social science qualitative datasets and is also led by UKDA. The overall direction and management for the ESDS is the responsibility of the UKDA.

UKDA also provides preservation services for other data organisations, supports the National Centre for e-Social Science (NCeSS) and facilitates international data exchange through agreements with other national archives. The UKDA hosts AHDS History and the Census Registration Service, facilitating access to the census data resources for UK higher and further education.

The social sciences material for which UKDA is responsible is organised by 'studies': these cover all the material relating to an investigation including, e.g., raw input, coded data, statistical analyses, explanatory documentation and reports. Individual studies may thus be very complex and substantial and comprise many datasets, for example where longitudinal surveys are concerned.

UKDA ensures there is standard-form metadata (so cataloguing is a major cost). Individual studies may have their own further internal discovery tools. The UKDA catalogue is an XML database, the content of which is visible to and discoverable via Google.

Government data is the most highly used part of the collection. UKDA's acquisition policy is driven by a mixture of user demand and by what is funded by the research councils and other funders including the British Academy. UKDA now takes in very little primary material in 'non-e' form. It does not deliver such non-e content to users, who need to visit Essex to see it.

UKDA does not charge UK academics for use of data, but reserves the rights to charge UK non-academic and international users for access.

The UKDA's focus on e-resources, and adoption of data from many sources, has meant the service places heavy emphasis on adequate preservation and curation. It has accepted that it need not hold, i.e. house, all major resources likely to be of interest to HSS researchers, but provides lightweight catalogue pointers to resources held by others, and is interested in and advises on maintenance. For the resources it does house, it has adopted a belt-and-braces preservation strategy by keeping all forms of a resource.

The UKDA is strongly committed to encouraging use of its resources and welcomes the fact that better access through ICT means that the same resource can be more widely exploited across disciplines, e.g. by historians, environmentalists or health researchers, as well as 'mainstream' SS researchers.

Overall, the UKDA plays an important part in supporting e-resources that are not only significant in and for the UK, but are valuable for researchers elsewhere. It also shows how any e-resource operation has to be continually developed, and provides a good model for the manner in which e-resource provision has to be packaged for effective and sustainable research support.

### h) Arts and Humanities Data Service (AHDS)

The Arts and Humanities Data Service is a UK national service funded by the JISC and AHRC to collect, preserve and promote the electronic resources which result from research, and also teaching to some extent, in the arts and humanities.

By preserving collections made in the arts and the humanities, the AHDS encourages research and educational use of its collections and makes information about them available through online catalogues.

The AHDS was established as a geographically distributed service comprising a managing Executive and a number of AHDS Centres on separate sites devoted to archaeology, history, literary, linguistic and other textual studies, the visual arts, and the performing arts. The Centres collect, preserve, catalogue, and distribute digital resources which are relevant to their subject areas, facilitate and promote good practice in their creation and use, and offer some user services.

The individual services are substantial enterprises in their own rights. AHDS Archaeology for example, hosted by the Archaeological Data Service (ADS) at the University of York, covers a large range of UK-referring resources from one dealing with glass beads from Anglo-Saxon graves through the Museum of London Archive to materials derived from the Channel Tunnel Rail Link. Individual resources have their own metadata and discovery tools, but there are AHDS Archaeology global search mechanisms that apply across the component resources; these can also be used for such records as the service has for non-UK resources. AHDS Archaeology is essentially an assembly of resources brought together, for greater potential value to the user and improved resource sustainability

The AHDS's federal structure reflects the varied origins and development of the services for the different subject areas. This means that there is a lack of consistency in resource treatment which is not really justified by intrinsic differences in the character of resources or user needs in the areas. The AHDS is working towards a more homogeneous approach, but as inspection of a component service like AHDS Archaeology shows, there is and always will be, great variety in resource origins as well as characteristics, and higher-level bodies like the AHDS cannot expect to do more than throw a light mantle of uniformity over the whole, though hopefully one of a sort that can enable users to find their way among many resources in a reasonably efficient and effective way. It is evident that any attempt to impose extensive or

stringent metadata requirements would be extremely costly and probably not especially useful: lighter weight paths and tools for resource access that apply across many different types of resource, and are cheaply developed and well-taken for users, are likely to be much more appropriate.

The current set of sub-services does not, by any means, cover all of the (non-performing) Arts and Humanities. It is policy that a UK-derived resource will be allocated somewhere, but this is not necessarily at all convenient for the individual researcher. There has also been the problem so far that while an individual resource, e.g. in philosophy, might be saved from complete invisibility by being assigned to the History Data Service, the latter had no commitment to providing much in the way of help to the philosophy resource seeker. The AHDS is, however, currently investigating needs for further subject coverage and how they may be met, with the intention of broadening its effective range.

## 2.3 University and similar libraries

The researcher survey in Section 3 provides gives a good picture of the current state of resource provision and accessibility, both e- and non-e, as experienced by academics in the UK's research intensive universities.

We were anxious however to also obtain information about the broader picture of provision in university libraries, including both older smaller universities and post-1992 universities. We therefore consulted a sample chosen to cover a range from a major historic copyright university library to post-1992 university libraries from different areas of the UK. We also consulted the National Library of Wales. The list of those contacted is given in Appendix 3.

We have divided the libraries into three groups: older universities, post-1992 universities, and copyright repositories, since there are historical differences of resource provision. However the information received shows there are similarities across the board. There was general consensus that the concerns identified in the policy review's terms of reference are real and important.

### a) Older libraries

Responses to our enquiry were received from Southampton, Exeter, Birmingham, Leicester and City Universities. All of them reported significant pressures on resource provision. There was general agreement that the ever-rising costs of STM journals were squeezing provision for HSS but that other factors were also involved, including overall low levels of budget settlement for institutions and their libraries, and the fact that some universities have always been, and remain, less wealthy than others. Non-e research resources are becoming less accessible in most institutions. Space is a problem for some, with more space needed for workstations and other equipment and priority for prime space given to high-use teaching materials rather than lower-use research material.

The purchase of books and monographs to support research has been under particular pressure, affecting HSS disciplines disproportionately. The libraries point out that few of them have been self-sufficient in the provision of HSS research monographs in the past, that there has always been dependence on inter-library loan for the borrowing of materials, and a generally accepted need for researchers to travel and use resources held by other institutions. The latter has been facilitated in recent years by agreed access schemes such as SCONUL Research Extra and by a stream of funding from HEFCE via the Research Support Libraries Programme (RSLP) towards the costs of providing library access to external researchers.

Libraries are finding it hard to maintain book purchasing for research rather than teaching needs, and have particular problems in justifying expenditure on resources where, as is often

the case in HSS, the number of researchers in a subject area does not reach critical mass, where the research, compared to STM, is relatively unstructured, and where the usage of individual items is relatively low. While there are many researchers in STM who work individually rather than in teams, and equally some HSS researchers engage in large projects, the fact that HSS researchers typically work alone or in small groups does have important consequences for much resource provision. But low local resource usage can be partially offset by e-operation.

Thus the libraries emphasise the value of e-resources, to which they are increasingly shifting, as more accessible and flexible than non-e ones and as having enhanced functionality. Resources such as Literature Online support both teaching and research needs, while large-scale full-text, searchable, primary texts enable a new kind of research to be carried out. EEBO, JSTOR and DNB are all mentioned as important resources of great value across many disciplines, particularly in HSS. The use of e-journals, both current and, where available, back runs, allows the freeing up of valuable space. Much of the shift is therefore driven by a positive interest in e-resource benefits, and not just by being able to reduce access unit costs. However it is also pointed out that while e-resources improve accessibility for the institution's own staff and students, licensing restrictions on many products prevent access being allowed to 'walk-in' visitors, including researchers from other institutions. Therefore while good progress has been made in recent years in providing collaborative access to physical materials, this does not apply at present to many resources which are only available electronically.

Cost is a major issue in relation to e-resources. Many large-scale e-resources valuable to HSS researchers are very expensive, while the number of researchers in an individual institution to whom they are of interest may be small. Even an institution as large as Birmingham University is unable at present to afford to provide access to the full text database *Times Online 1906-1980*.

Some resources such as EEBO and ECCO can only be afforded by individual institutions because a national deal has been done by JISC. JISC paid for the content on behalf of all UK higher education, so institutions pay a much lower annual access/hosting fee to use the material. Cost is a key barrier to institutions being able to make new electronic resources available, and it is likely that many institutions are already close to the limits of what they can afford to spend on e-resources. There is approval of the not-for-profit business model of JSTOR, and some see a clear international need for more support of large-scale digitisation projects, many of which will not be commercially profitable for a publisher working alone. The librarians also note, repeatedly, how cramping the IP position is, and how restrictive publishers' charges are.

While there is gratitude to JISC for its national licensing and funding activities in relation to e-resources, there are also some concerns about the proliferation of new e-resources and a lack of overall coordination and strategy, as seen in the funding of too many short-term projects and insufficient focus on long-term sustainability. There are also concerns about consistency and connectivity from a user point of view, with users faced with a multiplicity of access modes and opaque variations in detail that make it difficult for them to discover what there is, where it is, and how to get to it, and also make it difficult for librarians to help them. There is a lack of joined-up thinking, so resources that could be valuable are not as accessible as they should be and so are poorly utilised.

Better strategic thinking by the funding bodies, coordination and collaboration would make it easier to cope with technology changes, and to realise the benefits that e-resources promise. It is recognised that the technology is not robust enough for libraries to be able to develop very solid models for e-resource provision, and that the situation in e-publishing is fluid, so that strategies and plans have to be provisional. It is likely, however, that the difficulties faced by libraries are compounded by the larger lack of coordination and strategy.

The libraries emphasise the importance of need for appropriate training, of both users and library staff, as well as the need for institutional policy development that allows libraries to adapt their service models for more use of e-provision.

Future improvement lies in the hands of institutions, funders and researchers themselves. City University comments that "we need to better equip individuals to deal with how to access multiple resources and gain transferable skills so that they can adapt quickly to changed circumstances. But as institutions we need to ensure easy access mechanisms for our resources, joined up thinking in terms of what provision we are providing and pressure on funding agencies to think about sustainability of research and development into online resources".

The older university libraries are clearly engaged with e-resources; but while much good progress has been made, they are also clear that access and sustainability are the key needs, and that the current piecemeal provision and lack of coordination does not fully meet these needs.

### b) Post-1992 universities

Responses were received from the Universities of Wolverhampton, Westminster and Central Lancashire.

Libraries here have the same problems, and see the same opportunities, as those in the older university sector. But as they tend to have fewer researchers, it is harder to ensure that their researchers have access to the resources they need, both non-e and e-. E-licence charges can be unacceptably high for just a few users. Historically these libraries have not bought much HSS research monograph material, relying instead on interlibrary loan and on cooperative access arrangements with other libraries within their regions. Strong and valuable arrangements are reported to be in place in Lancashire, the West Midlands and London. The success of these arrangements for researchers may be undermined if the same level of access to material in other institutions that is possible for physical materials is not available for electronic resources, because of the licensing restrictions mentioned above. Institutions will need to develop patterns of cooperation for e-resources like those that they have established for non-e resources and it will be important that provisions in the JISC model licence to allow for use of electronic resources by visiting researchers and 'walk in' users are widely accepted by publishers and adopted by institutions.

The librarians' comments again emphasise how complex the resource position is. The arrival of e-resources have changed patterns of expenditure not only in such obvious ways as in a shift from non-e to e-journals; thus at Wolverhampton databases, as an e- rather than non-e resource of a kind that would not formerly have been a major item for a university library, are now a significant element in library expenditure; and they are important, along with e-journals, for HSS researchers. But substituting e-resources for non-e ones may not always be a net gain. Buying or subscribing to an e-resource may be far more expensive, for a small user community, than relying on interlibrary loan or document supply, especially since currency is often less vital for HSS researchers than for STM.

Budget constraints for new resources are a particular problem. Westminster chose not to subscribe to the JISC EEBO offering, even though it required only the payment of an access/hosting fee that was modest in relation to the scale of the resource, commenting that they could buy a significant number of e-journals for the same amount. Wolverhampton point out that "the historically lower level of price increase for HSS periodicals as opposed to STM periodicals has helped preserve access to journals, but may not do so in the future. However the inflation position is reversing with HSS price inflation now exceeding that in the sciences. On-going inflation of periodicals (print and e-) represents a real threat to HSS research. The proliferation of titles and their increasing specialisation has resulted in more fragmented demand, for what are often poorer quality publications".

Just as elsewhere, and as with STM users, HSS users like e-resources that can be accessed from their desks, especially discovery tools and databases and e-journals. However licensing and costs of e-resources are seen as having had a detrimental effect on provision of access to e-resources. While accepting the vital nature of copyright, there is concern that its application can often be used by publishers/institutions to restrict access. Licence costs for limited numbers of research users are often prohibitive. It is suggested that joint licensing through regional co-operatives or national licensing through a core collection could help alleviate this and, if nationally supported and nationally available, could help to level up access to resources across the sector.

The libraries are clear that institutions are having to cope with substantial changes in the form and hence provision of research resources, but they note that researchers themselves need to adapt, not only in the demands they make for resources but in the way, as creators, they contribute to them, e.g. in assigning copyright to publishers.

## c) Copyright repositories

### Cambridge University Library (CUL)

Cambridge University Library, along with the Bodleian Library at Oxford University, is one of the six libraries in the UK and Ireland entitled by statute to receive UK publications under legal deposit. Recent changes to legislation have extended deposit to include non-print publications although this will not begin formally for electronic publications until further statutory instruments have been agreed and come into force. In the meantime voluntary arrangements for the deposit of some types of electronic publication are in place.

CUL, as a major research library, has large historical holdings, some of which will probably never be digitised, though some individual archive collections are being digitised as special projects, and the library has begun to receive born-digital materials, both through the licensing of commercial electronic publications such as journals and databases and through items received by voluntary deposit. It has important component libraries, e.g. for law, and is complemented by many department and college libraries that may also hold valuable research material. Under financial pressure there has been some university-wide rationalisation of journal subscriptions. There are union e-catalogues for monographs and serials.

CUL illustrates the problems besetting major research libraries throughout the UK. Moreover, even though it has the advantage of legal deposit in building its holdings, this presents its own problems.

The Library has been under increasing financial pressure from rising STM journal subscriptions and an STM user community that demands both paper and e-versions of journals. It is committed to sustaining resource acquisition for HSS (both journals and monographs), and is keen to protect the relative proportions spent on HSS and STM materials (currently ca. 50%/50%) but the HSS community is also exerting its own pressure by wanting both non-e and e- versions of major reference works. Going electronic is thus not a simple gain even though more users may be able to obtain access.

However even this is not always the case: in recent years CUL has received some increased funding under the RSLP access funding scheme towards the additional costs of providing access to researchers from other institutions with poorer collections; however while such researchers have been able to have ready access to CUL's physical collections, they cannot be given access to equivalent e-resources where licences stipulate that these may only be used by members of the university.

The Library has always played an important role, nationally, through its extensive acquisition of foreign-language publications that are important for HSS and are disappearing elsewhere. But this role is becoming hard to sustain as desired under the combined pressures of an overall squeeze on library budgets, the ever-increasing costs of STM journals, and the fact that no extra funding has been provided for e-resources (e.g. to pay for ongoing access charges or the supplementary costs of e-access on top of the costs of the same resource in print).

The problematic consequences of going electronic are also evident with deposited e-materials: just as at the British Library, these can only be accessed just as if they were non-e, in the Library itself.

CUL has been involved in digital preservation activities for some time, hosting part of the JISC-funded CURL CEDARS project (CEDWWW), and it is currently hosting an institutional repository experiment on behalf of Cambridge University based on MIT's DSpace system. It is well aware of both the technical challenges and the sustainability issues involved.

In general, CUL's view is that the current pressures on research resources would benefit from a more coherent national strategy. This could help, for example, in negotiations with publishers, and in informing collaboration between key players.

Currently, the Library sees more relative value in improving its secondary e-resources, for example by completing retrospective conversion of and making e-accessible all its manuscript and card-based catalogues, than in undertaking piecemeal projects, unless independently funded, in providing more primary e-resources through digitisation.

### National Library of Wales (NLW)

The NLW is a legal deposit library, and the UK library with a special remit to collect print and non-print materials pertaining to Wales. It has a wide range of responsibilities and activities including, for example, hosting and displaying items from its substantial picture and image collection, and holding the Welsh national screen and sound archive. Unlike the Cambridge library, the NLW is independent of particular university conditions.

The Library has significant archives and has embarked on a programme to provide e-access to its archive material, initially at the high level of collection description. It is already engaged in the digitisation of primary materials, in the context of enhancing access to Welsh cultural resources for a wide range of audiences, and therefore paying particular attention to presentation. But though not focused on the needs of higher education and research, the materials include some that are of scholarly interest.

As a deposit library the NLW is now accepting e-materials as well as the non-e ones it has always taken. It takes these e-materials in whatever e-form they come in, and is currently treating these digital items in the same 'object' style as non-e ones, with the same anomalous and unsatisfactory implications for access that other deposit libraries have noted. But since there are formal as well as practical challenges about ensuring access to e-materials, the NLW is continuing, where there are parallel non-e and e-versions, to take deposit materials in non-e form.

Since the NLW is not also a university library is not under pressure to attempt to service HE systematically and to cover all subject areas. In practice its Welsh focus makes it more HSS- than STM-oriented, and implies, for example, that it only has some databases of interest to researchers. But it is exploring collaborative repository possibilities.

In general the library, like other major UK libraries, is grappling with the changing e-resource landscape. Here ICT is rewriting the realities of resource ownership, status and use, and producers are making more of the running than consumers. The familiar models for resource

selection, acquisition, storage, curation, preservation and servicing are becoming less and less appropriate. But there is far too little coordinated effort being put into developing solid new, long-term models for the future.

## 2.4 Museums and galleries

Libraries and archives may contain many non-text items of research interest. However it is important to consider museums and galleries as resource holders here for four reasons, even though we have excluded physical objects in general as resources. Thus museums and galleries may hold large image collections, e.g. the Imperial War Museum's photograph archive; they have had to grapple with the issues of catalogue strategies for extremely heterogeneous materials, where the scope and nature of metadata are major challenges; they engaged early in automated cataloguing: the Museum Documentation Association dates from the 1970s, for example; and they have provoked controversy, in relation to access to resources, about appropriate ways of exploiting IP, notably in relation to charging for copies of images.

As the Virtual Library museums pages (VLMWWW) show, many and varied museums are offering e-entry points, which may include e-catalogue access. The Imperial War Museum's Collections Online illustrates the increasingly sophisticated Web sites that are appearing in the museums and galleries world, including ones handling multimedia. But as with archives, museums and galleries range from the very large to the very small, and thus illustrate the challenges that resource owners that are not in the big league from the point of view of size of holdings and funding face in seeking to make their materials more accessible. We have taken the National Museums of Scotland, a federation of very different museums, as an example to illustrate museum resource issues.

### National Museums of Scotland (NMS)

The NMS is a federation of six museums of varying sizes and scopes. The combined catalogue for their collectively substantial holdings is automated, but at present available only in-house and primarily for museum staff, through a simple Web-based front end. The NMS is currently considering appropriate strategies for opening up the e-catalogue to non-staff users, as part of a wider 'online museum' development. The present version of the e-catalogue is relatively shallow so the issue, as elsewhere, is whether to make this available, despite its deficiencies, or to hold off for a catalogue with higher-quality entries. The first option would be easy to implement, but the second is curatorially more attractive. Either way, there is also a problem in that the catalogue metadata is heavily controlled, with thesaural indexing: opening up the catalogue to ordinary users would almost certainly require the development of search mechanisms which would not presuppose familiarity with specialist indexing tools, though this would not be a major exercise.

The current e-catalogue system is implemented using off-the-shelf collection management software. This has the advantage that longer-term sustainability issues (which are not trivial even though such a secondary resource is much smaller than many primary ones) may be contained by the software supplier's commercial interest in ensuring upwards compatibility over time. Using standard software also ensures interoperability over the component catalogues, and can help in linking the NMS's resources with those of other Scottish museums. The NMS is actively working to develop links with such museums, by approaches similar to those represented by A2A for archives though, as with archives, many museums still have only (limited and) paper catalogues.

The NMS also has image material including both historic photographs as primary museum objects in their own right and photographs of other objects. Some of this material has been digitised and is available, but independent of their own system, through SCRAN (SCRWWW).

The Museum operates a variable-rate charging policy for reproduction of images depending on intended use. The independence of the image materials is paralleled by that of the Museum's library materials: though the library's catalogue is electronic, it is not integrated with the object catalogue to form a single e-resource.

The NMS clearly illustrates the issues, for secondary resource holders, of shifting from non-e to e-operation; but it also illustrates the way in which researchers can expect a rapid increase in the quantity and range of e-resources, especially secondary e-resources, that they can reach.

## 2.5 Learned societies

Learned societies play an extremely important role in resource provision for HSS, in several different ways. Through their meetings and memberships they are informal research resources; their journals and other publications, such as monograph series, are often major resources in their fields, internationally as well as nationally; and they may have very important archives of manuscripts, photographs etc. Some, such as the Royal Geographical Society, are beginning to provide e-access to their holdings, at least at the catalogue level.

UK learned societies are typically not very large (of order at most a few thousand members), though professional societies (which combine being a learned society with ensuring professional accreditation) may be larger. UK STM professional societies like the Institution of Electrical Engineers (IEE) were very early committed to e-resources, through their abstract journals, though the major and very influential commitment to automation was by US organisations like the American Chemical Society. More recently, the US-based but international Association for Computing Machinery has been a leader in promoting relatively unrestricted use of papers from its own journals for research and teaching, and in creating an electronic archive of its back holdings. UK HSS societies have, however, generally adopted a rather cautious attitude to modernisation as manifested or enabled by ICT, but have been willing, for example, to place their journals in the electronic databases offered by Cambridge University Press or Blackwell, in response to author pressure, and the back runs of some of these journals (e.g. the journals of the Royal Geographical Society, the Economic History Society and the Royal Anthropological Institute) are accessible electronically via JSTOR.

There appears to be rather little interest in HSS learned societies in the trend towards open access, primarily because their subscription charges are lower and hence are under less pressure, but perhaps also because it may be difficult for a small or medium society to develop an appropriate business model for a radically new mode of operation which they may fear will undermine what appears to be a relatively steady income stream from conventional hard copy subscriptions from academic libraries.

We consulted some societies chosen to reflect very different subject areas, community sizes, and resource types. We were interested primarily in their activities as resource providers, but also in the information they supply about resource characteristics in different research areas.

### The Philological Society

The Society is an old one, one of four in the general linguistics area. It publishes a journal, and occasional monographs; but apart from its journal archive, does not have other research resources.

The journal is published by Blackwell, under the usual type of arrangement where Blackwell handle all the arrangements in return for most of the subscription revenue. The journal is published both on paper and electronically, as part of the Blackwell journal stable. Society members get the paper journal free as part of their subscription; other individuals and

institutions have several reasonably-priced paper/electronic options. The monographs are published only in paper form.

The Society has built up financial reserves over time, and thus is under no financial pressure to change the style of its operations. It could in principle benefit from its e-operation, both for current journal issues and the back archive, and the e-version of the journal could in principle benefit the wider research community, though it appears usage is not high.

The Society has an older member age-range compared with, say, the Linguistics Association of Great Britain, and appears to be operating in traditional mode. It is not seeking to take any initiatives that might influence e-provision, for example of resources in older or newer languages, where libraries are in general being subject to pressure on resource provision.

### The Royal Anthropological Institute (RAI)

The Institute is, like the Philological Society, old and not very large. It represents an area where there are many independent researchers.

Its primary publication is the *Journal of the Royal Anthropological Institute*. It also publishes the more popular *Anthropology Today* and the *Anthropology Index Online,* a bibliographic database which indexes all periodicals received by the Library of the Department of Ethnography of the British Museum (BM). The latter is effectively Britain's national anthropological library, housing the merged library collections of the BM Ethnography Department and of the RAI. The RAI contributes to staffing and support of the Library, which is owned and otherwise administered by the BM. The library is substantial (ca. 120,000 books and pamphlets and 4,000 journal titles, of which around 1,450 are current). The regional scope is worldwide and all aspects of anthropology are covered. The catalogue is online, but is only available onsite at the BM.

The journal is published by Blackwell, in standard style, free to Fellows of the Institute, and at modest subscription for others. The back archive is available in JSTOR.

*Anthropology Index Online* is well regarded and well used. The *Index* is licensed to and made accessible via the US Research Libraries Group online system. Access to the *Index* is also available direct on the RAI Web site for £280 p.a., but with more limited functionality. As a result of the subscription income *Anthropology Index Online* is a financial success.

The Institute's archives are significant, and illustrate some important issues for resource access. Primary materials in anthropology include still photographs and film/audio/video records, as well as field notes. These have often been rather casually treated by their collectors, and may be quite widely spread round museums etc. As with many such collections, the Institute's archives have been very lightly catalogued. Some materials are not catalogued at all: photographs are mainly uncatalogued, though there has been some scanning of photographs alongside cataloguing as part of a project over the last three years. None of the catalogues or finding aids is available online. However as full cataloguing is slow and expensive, automating even the minimal catalogue could be a community benefit, and the Institute could set an example here to demonstrate that the present good may be much more useful than the very remote best.

Much anthropological material is similar to that in other social sciences and is thus subject to ethical constraints. However anthropological material differs from much social science data in being small scale and detailed, and so is primarily qualitative rather than quantitative. Issues of confidentiality and ethical use are therefore extremely important. The Institute has been very careful about access to and use of its non-e archive resources. However as research becomes electronically based, and archives become e-archives, issues about proper attitudes and procedures in relation to new, much more accessible, archive materials are becoming more pressing.

### Economic History Society

This is a medium-sized society with a good range of interests from qualitative to quantitative work over a wide time range (including 'contemporary history'). It does not have a library, or significant subject area archives. Its main resource is its journal, the *Economic History Review*, which is a high status international journal with a substantial and profitable subscription base. The *Review's* standing means it is well established as a core journal, making it robust against libraries cancelling subscriptions. Members get it free and the institution cost is very modest. The journal is published, in both paper and e- form by Blackwell, under their usual type of model. It went e- fairly late, but can be had in e- form only. The back archive has been digitised and is available through JSTOR. The society has begun a fairly modest Web site.

However, though the Society's movement into e- has been moderate and conventional so far, it is becoming subject to increasing e-pressure. This is partly through the influence of the US Economic History Association, which has an active Web site, partly through the International History Association, which is proposing that the data on which articles are based should be made available for e-access.

This proposal has two aspects: that of mechanics, and that of intellectual property. Both are currently controversial for the Society. The mechanical aspects are not trivial, since making the materials public would mean satisfying all the usual requirements for sensible implementation, documentation etc. The Society has no 'in-house' e-apparatus and staff. On the mechanics side meeting the support 'publication' proposal would mean extra work for authors and some agreement with a suitable repository and database access service.

However the IP issue is felt to be much more important. It is not usually the case that all the research ore in a data resource has been mined for a single article, especially one published soon after the data has been created. There may be much more to be got out of the raw material. Thus publishing the dataset is simply making the raw material available to other, competitive academic miners.

The Society's members are generally unwilling to give up their exclusive access to the stuff of their academic status until they have fully exploited it. At the same time, they recognise that researchers generally benefit from access to one another's data, especially as more and varied findings are drawn from it for further use, and additional data sets may be built that extend it, e.g. with price series, by increasing the range of organisations, countries, etc. considered.

The trend towards data deposit, ensuring access, has been stimulated by funding agencies, and there are major data sets and series that are well established, internationally, as community resources. But the Society's individual researchers are still nervous about major policy changes associated with publication that alter their established modus vivendi, and there have been no policy decisions on this whole issue yet.

## 2.6 Other resource providers

In this section we have so far emphasised major UK players in providing or facilitating access to HSS research resources. These bodies range from those directly handling resources such as the British Library and university libraries to those, like the MLA, which have formal oversight of numerous resource holders or, like TNA with A2A, which encourage collaboration between holders.

In general these bodies have not been responsible for primary resource creation, whether of books, papers, images or databases. Their crucial role has been as providers of access to resources and their content. But they are of course in many cases responsible for very significant

secondary resource production. It is also evident, for example, from the long list of resource holders under MLA's wing, that there are very many UK resource holders, dealing with a huge range of resource types of vital importance to UK researchers.

The learned societies, on the other hand, through their journal publishing operations, are primary resource providers at an earlier point in the food chain, if not original content authors or creators, and the same applies to the commercial publishers. We consider publishing, as an early stage in primary resource provision in Section 4.

But there are also, increasingly many individuals or groups that are creating or producing resources and making them directly open, as e-resources, to researchers through Web sites and the like. While direct interchanges have always taken place through correspondence and papers 'with compliments', primary resource creation in e-form, with e-'posting', is now on such a large scale as to constitute a wholly new phenomenon. This distributed resource provision or, more generally, production is thus an important new opportunity for researchers to access resources.

More significantly in this, UK resources are merely part of a far wider, international construction of e-resources. Researchers have always operated internationally, but in what may be called a discontinuous way, by themselves having to pull together resources clearly marked by their separate origins. But with e-resources and modern discovery tools, it is far easier to roam the world.

We consider the general implications of ICT and internationalisation in Section 4. However there are also institutional aspects to note here.

## 2.7 Non-UK resource provision and access

As noted in Section 1, ICT is shifting the actual or perceived division between UK and non-UK resources. While some resources continue to be mediated by UK bodies, for example journals through libraries, ICT has brought other established resources directly to the user's desk, for example major non-UK library catalogues. More importantly, vast new resources that were formerly so inaccessible as to be virtually non-existent for many researchers have been placed within reach of anybody with a browser. This applies both to the visible Web, with its billions of pages, and the invisible Web, where access is controlled or charged for, but where material is far more reachable than before.

This development has wide implications for the funding and organisation of resource provision and access, and for bodies involved in this, with consequences for researchers beyond those they already see following from ICT itself. Non-UK operations can affect UK institutions in several ways:

1) through the actual provision of important e-resources that UK researchers wish or expect to be able to use;
2) through formal international agreements and collaborations, that (re)allocate support for resources between the UK and other countries;
3) through non-UK activities and investigations that provide examples for the UK, for instance in connection with resource preservation.

From the institutional point of view these imply changes in the patterns of financial commitment to individual resources or types of resource, moves to novel supply models and eventually, cumulatively, a radical restructuring of institutional operations, whether in the forms of resources encouraged and adopted or the distribution of money.

## Significant non-UK e-resources

We cannot attempt to review the many non-UK e-resources that are now available. But our researcher survey, and the number and variety of e-resources listed on PORTAL and on the Resource Discovery Network's HUMBUL and SOSIG gateways, show how important and wide-ranging they are. However even free ones have financial implications for institutions through needs for machines, printer paper and cartridges, while charges for access to other e-resources may be large. But simply to indicate how significant these resources now are, we can take as examples such major non-UK primary e-resources as *Gallica* (GALWWW), *Making of America* (MOAWWW), *Perseus* (PERWWW) or the *United Nations Common Database* (UNCWWW) or such secondary e-resources as OCLC's *WorldCat* (WORWWW) or *ISI Web of Science/Web of Knowledge* (ISIWWW)

## International collaboration

Collaborative schemes are important both for resources themselves, and for guidelines and standards for resource creation. The *Electronic Cultural Atlas Initiative* (ECAWWW) illustrates international collaboration in the provision of resources. Guidelines are illustrated by the European Union's *Lund Principles* (LUNWW) for the coordination of digitisation programmes across EU member states, and standards collaboration is illustrated by the *Dublin Core Metadata Initiative* (DCMWWW). At the organisational level there is collaboration between national libraries, such as in *The European Library* project (TELWWW), and between JISC and foreign agencies such as the National Science Foundation in the US and the SURF Foundation in the Netherlands. The need for collaboration with respect to data resources, illustrated for STM by the recent ICSU report *Scientific Data and Information* (SDIWWW), is also pertinent to HSS. Broader collaboration in the sphere of cultural heritage, for example as encouraged for Europe through *DigiCULT* (DIGWWW) has implications, too, for resource provision and access.

## Indicative experience

It may also be helpful to UK institutions to see how other countries undertake, or have responded to the problems of, e-resource provision.

It is natural, in particular, to see what lessons can be learnt from the US, since this has been both a leader in ICT generally and in e-resource provision since the days of Marc records and OCLC, through Web engines, to Google's recently-announced digitisation initiative.

In fact, the US's large size and federal structure has meant that the organisation and funding of e-resources has been a largely ad hoc matter. Some major bodies, like the Library of Congress and OCLC, have had very significant roles, primarily within a library framework, and organisations like the Coalition for Network Information have worked to develop distinctively e-oriented approaches to resources ranging over the spectrum from technology to politics and attitudes. But the main driver in the US has been STM as seen, for example, in the major repositories and the NSF initiative on cyberinfrastructure (RSEC03).

There is however a growing concern that the focus on e-STM may not necessarily benefit HSS, and the American Council of Learned Societies is currently investigating cyberinfrastructure needs and roles for the Arts and Humanities (CCHWWW). However while major university libraries and museums can see advantages, there is considerable grass-roots faculty resistance to what are perceived as the unwarranted claims for research gain from ICT and a continued dedication to the cause of the printed monograph. The value that STM gains from very large databases is viewed as not applicable to AH, even though, for example, linguists have learnt to value large corpora.

Overall in the US, resource organisation is more fragmented, so the US's technological experience is more significant for the UK than its organisational one.

Elsewhere, much can be learned from the experience of other organisations, such as the National Library of Australia in the harvesting and preservation of Web resources, and the Koninklijke Bibliotheek in the Netherlands in the storage and preservation of e-publications.

At the legislative and policy level the European Union plays an increasingly prominent role, with EU directives on copyright and on public sector information already, or about to be, implemented in UK law, and with the *eEurope 2005* (EUWWW) action plan aiming to stimulate secure services, applications and content based on a widely available broadband infrastructure, which is part of its Lisbon Strategy for economic, social and environmental renewal.

## 2.8 Conclusion

The evidence from the various bodies we have consulted, and others, shows a range of responses to the relative supply of, and needs for, non-e and e-resources; a great deal of activity in relation to e-resources and access to them: an awareness of important issues, especially in relation to e-resources in the long term, but only rather modest attempts to grapple with the problems involved; and an overall extremely heterogeneous patchwork of resource provision, which can only partly be attributed to the pace of ICT change and must also be attributed to a lack of effective policy coordination and development.

# Section 3: Researcher Survey

## 3.1 Introduction

Our survey of resource users was intended to gain information about

a) the resources used, both non-e and e; and both primary and secondary; and

b) the researchers' access to, and also help with, these resources.

We also sought information on

c) the researchers own involvement in resource creation; and on

d) their ICT technical capabilities and the support available to them;

and we invited general comments.

Time and resource limits, along with data protection restrictions, meant that our survey was limited to researchers connected in some way with the Academy. The sample was small and cannot be taken as fully representative of the UK HSS research community as a whole; it had a bias towards more senior academics and better-resourced institutions, and also towards humanities disciplines. We much regret that our survey could not be larger and more fully representative. But those surveyed are experts in their individual subject fields and resources within them. More importantly, we believe that the detailed qualitative information we got from the survey, taken together with our other data, does give a good picture of HSS researcher needs for resources and their modes of accessing and using these. The individual responses are a valuable bottom-up complement to the top-down institutional view presented in Section 2.

Within the limits of a 'postal' questionnaire we sought information in depth. We were gratified by the very detailed and constructive responses that we received, and are most grateful to those who replied for the trouble that they took. We have information for senior, mid-career and new postdoctoral researchers, and for an extremely large range of research topics, in all subject areas, illustrating a rich and varied research world.

As a whole, the responses show a proactive, and often enterprising, use of e-resources, even where much of a researcher's work is necessarily with non-e resources, and regardless of researcher age. Our survey clearly demonstrates that HSS researchers are far from being inhabitants of ivory towers poring over dusty charters. They are conducting their research with as much e-sophistication as STM workers, are developing HSS research in new ways, and are also, with ICT help, reaching across subject boundaries, both inside and outside HSS. Some of this ICT gain comes from using generic resources, notably email and Web engines. But much comes from HSS-specific resources, making it clear that access to, and the provision of, e-resources is vital for HSS research health.

This section summarises the survey process and presents the main results, following the questionnaire structure, with illustrative quotations from the returns. A fuller account is given on the British Academy's Web site at http://www.britac.ac.uk/reports/eresources.

## 3.2 The survey process

The survey was carried out between July and September 2004. The survey questionnaire was posted on the British Academy's Web site. Academy Fellows (FBA), Academy Small Grant holders (SRGH), and Academy Postdoctoral Fellows (PDRF) were contacted by email, where possible, and invited to complete it. Fellows without email addresses were also contacted by post with the questionnaire, and the Web site invited other UK-based researchers to complete it.

The specific questions asked and the responses to them are detailed in the following sections.

## 3.3 The survey results

The email message was sent to 1438 researchers (517 FBA, 489 SRGH, 432 PDRF), and the postal invitation to 263 FBAs. 260 completed questionnaires were received, representing a 15.83% response rate, out of 296 replies. There were 122 FBA responses, 77 SRGH, 36 PDRF and 25 others. In the detailed analysis which follows, all responses whether by email or post are combined, except where there is a striking difference between the email respondents and the whole sample that can be attributed to the fact that many postal respondents are retired.

### Institutions represented

Our responses were from institutions in all areas of the UK, though the largest sets were from Oxford (18.9%), Cambridge (13.1%), London (11.6%), and Scotland (11.6%), with for instance 3.9% from Yorkshire and 3.5 % from Wales.

### Positions represented

All levels of academic and analogous staff were included in the respondents, though professors, in post or emeritus, loomed rather large. Professors in post constituted 30%, senior lecturers/lecturers 25.8% and research fellows 12.7%.

### Subject areas represented

The respondents also came from all the major subject areas in HSS, with large groups from history (17.3%) and literature (11.9%), and others ranging from, e.g., sociology (2.7%) through politics (5.4%) to archaeology (8.9%).

## 3.4 Detailed responses

Unless otherwise indicated percentages are of the number of respondents giving that answer. Note that in the actual questionnaire, as given in Appendix 4, questions were amplified with illustrative examples, e.g. of resource types. We have simplified the question numbering here.

### Use of non–electronic resources

**Q1. What are the three most important types of non-electronic primary research resource you use?**

The major categories of non-e resource used were journals (80%), books/monographs (60%) and manuscripts (33%), but other classes like archival materials, editions of texts, early printed books, maps, newspapers and public/government documents were mentioned by at least 5% of respondents, and many and varied others, including interviews, music scores, inscriptions and treaties, figured at least once.

**Can you access these resources easily? If not, why not?**

65% of respondents said they could access their main resources easily, and a further 21% said they could access at least some easily. The problems with access to traditional, non-e, materials included both the more and the less familiar.

Problems described by researchers in access to traditional research materials included:

- Material not held or subscribed to by parent institutions
- Material held but not accessible because in use by others, not in fit condition for use, not borrowable or not yet catalogued
- Interlibrary loan costly, slow, items only usable in library
- Books/archives/other material required only accessible in other collections, either in UK or abroad
- Time constraints or funding problems for travel or for obtaining copies
- Organising problems re interviews/people
- Catalogues of owning institutions not online/accessible
- Some materials not retained or archived for long term access (e.g. foreign newspapers)
- Foreign fieldwork required
- Personal circumstances – health, mobility, location
- Access restrictions (e.g. Chinese mapping not accessible to foreigners)
- Difficult to know what exists (e.g. grey literature)
- Material poorly organised, difficult to use (e.g. publisher etc. archives), or not catalogued/indexed

The responses indicate the importance of journals to researchers across all HSS disciplines, together with the continuing importance for HSS research of the printed book and academic monograph and, in certain disciplines, of archival and manuscripts materials. The other striking finding is the sheer diversity of resource types used. The range of resources used is wider in AH disciplines, with SS researchers typically using a narrower range of resource types.

Availability and accessibility of traditional resources have always been and remains variable – printed materials have been historically better provided for in the copyright libraries, and in the libraries of the larger research intensive universities. However much material of research interest is held outside these institutions, in specialist collections in the UK, in archives, internationally, etc., and some research resources are not documentary in nature and therefore not held by institutions, such as archaeological sites, landscapes or the people studied by anthropologists or sociologists.

It is clear that future provision to support HSS research needs to take account of the 'hybrid' nature of the resources used (i.e. they will always be a mixture of physical and e-resources), and to ensure that developments to take advantage of the benefits offered by e-resources are not detrimental to the continuing need to discover and gain access to traditional physical resources.

> I have to travel to Germany to access many of the books and articles and all of the archive material that I require. (**Art history**)
>
> Getting time and funding to recruit participants for studies remains as difficult as ever. (**Human geography**)

(My university) does not have good primary holdings in my period (early C19th). Some source material only exists in London. (**Romantic literature and the history of science and medicine**)

Archives are widely scattered. Library holdings of journals and printed sources are patchy even in London. It all means much traveling and time wasted. (**Naval and maritime history**)

Library has large holdings but is 2 years in arrears cataloguing them, and often it lacks key German- and Czech-language books which I have to travel to those countries to find. (**Economic history**)

Problems can sometimes be caused by geographical location, limited hours of availability of private records, and material awaiting conservation (sometimes indefinitely) which is too fragile to be produced. (**Seventeenth-century English literature and manuscript studies**)

The list of journals taken by our University Library is reduced each year; this is certainly not peculiar to my particular interests (**Economics**)

Archives are in Italy; some published sources are not in the libraries; some journals have been discontinued by BL. (**Medieval history**)

A lot of what I want is in Baghdad… (**Cuneiform studies**)

## Availability of resources in electronic form

**Q2. Are any of these three important resource types also available, to a significant extent, in electronic form?**

The availability in e-form of the resources most frequently mentioned in Q1 was as follows (% is of the number of respondents who mentioned the specific type of resource in Q1):

| | *Yes* | *Some* | *No* |
|---|---|---|---|
| | % | % | % |
| Journals | 55.5 | 25.4 | 19.1 |
| Books/monographs | 6.4 | 15.4 | 78.2 |
| Manuscripts | 4.5 | 19.3 | 76.1 |
| Archival materials | 3.6 | 0 | 96.4 |
| Editions of texts and sources | 46.2 | 19.2 | 34.6 |
| Maps | 32.0 | 24.0 | 44.0 |
| Newspapers | 50.0 | 27.8 | 22.2 |
| Rare/early printed books | 16.7 | 44.4 | 38.9 |
| Public/government documents | 50.0 | 31.3 | 18.8 |

The picture coming through is of significant variability in the availability of different types of resource in e-form. As seen earlier, AH researchers use a very wide range of resource types, some of which are amenable to and available in e-form, but also many of which are not. SS researchers, on the other hand tend to use a more limited range of resource types, such as journals, statistics and other data, which are amenable to e-form and are often more readily available in that form.

Responses indicate that current English language journals in most disciplines are widely available electronically, as are data archives, many public documents and reports, current newspapers, legal materials and some maps. E-versions of some classical and literary texts and other documentary sources are available, and some books, especially early printed books and manuscripts and newspapers, are available in digitised form. Communication and sharing of information with colleagues in the researcher's discipline is also mainly electronic.

However it is clear that the greatest part of the book, manuscript and other archival material used by HSS researchers, particularly in humanities disciplines, remains unavailable in e-form and is likely to be so into the future.

**If so, can you access them easily? If not, why not?**

54% of respondents said that they could access what is available easily. (NB this figure was 63% in those contacted by email).

Access problems reported fell into the following categories:

- Resources not subscribed to by institution, e.g. Literature Online, Acta Sanctorum, or institution not member of service-providing body, e.g. RLIN
- Material not in e-form, or only partly, e.g. European humanities journals; 19th century books, archives
- Hard to discover what is available
- Difficulties of using, e.g. navigating through gateways to full-text versions of online journals.
- Access restricted or inconvenient, e.g. only available to members of institutions, or only usable in library
- Electronic records not online or accessible via Web, e.g. Archivo General de Indias in Seville, or only on CD-ROM, e.g. Patrologia Latina at Cambridge University Library
- Material not available in Mac format
- Online versions not complete or insufficiently detailed, e.g. manuscripts, images of art works.

Respondents are able to access material easily where it is freely available on the World Wide Web, or where their institution subscribes to resources and makes them readily available across their university networks. The combination of expensive licences and budget constraints causes problems of unavailability for researchers in smaller institutions, particularly where the number of researchers to whom a specialist resource is relevant is small and where its licence costs are high.

> Only a few 19th century Hebrew journals have been digitized. Most are available only in original or on microfilm in only a few libraries in the world. The same goes for Yiddish and Russian newspapers for my period. No archival materials or censuses relating to East European Jewish history have been digitized. (**Modern Jewish history**)
>
> An increasing number of rare books by C19th women writers are available via the Indiana University Victorian Woman Writer's Project and Bartleby etc. … Institutions such as the Harry Ransome Centre at Texas really should put their copious resources into scanning some samples (at least) of their collections of correspondence …. There are a number of sites which are seeking to preserve and make available culturally significant C19th and early C20th journals and periodicals. (**Nineteenth Century literature**)

No except for a few scattered samples, nor ever likely to be outside of a few popular classes. Much effort is put into scanning documents onto the web; this may be useful to give the public a taste, and give undergraduates some teaching materials, but for the serious scholar it is a waste of effort. He will need to use large quantities of interest to far too few people to justify the expense of scanning. Good catalogues and means of reference are what he really wants. (**Naval and maritime history**)

The main journals in the discipline (Hispanic Studies) are available online, but I work in a non-mainstream area (Peruvian literature and culture) and my main resources are small local journals which are not available in electronic format. (**Latin American literature**)

The reasons why the resources I want are not available must essentially be that they are very voluminous, desired by relatively few, would be very costly to reproduce and may be subject to copyright problems (**C18–C19 history**)

Few manuscripts that I need have been transcribed, let alone put into electronic form. (**Medieval history**)

Either no electronic version, or not subscribed to (usually the former). (**Early modern history**)

I use many well known scientific journals in my research and I don't see any reason why it's taking so long for nineteenth century runs to become available electronically. I also appreciate the cost of digitizing manuscripts but it would be very useful to have some of them available electronically. The most promising situation is with the textbooks. I regularly use Making of America, a website allowing searches on a vast number of non-fictional works published in the USA during the nineteenth century. This has proved invaluable in my research. (**History of science**)

Bibliographic control of reports has always been a problem. It remains so. They are accessible if one knows about them. (**Information science**)

**Q3. Where the same resource is available in both non-electronic and electronic form which do you prefer to use, and why?**

34% of respondents said they preferred to use the e-form, 38% said they preferred non-e forms, or the question was not applicable because e-forms were not available, and 28% said they preferred to have access to both or to use one or the other according to circumstances and purpose.

Reasons given for preferring to use e-versions included:

- Saves time and costs
- Is more convenient
- Is accessible from multiple locations
- Allows storage and manipulation
- Printable locally
- Constant availability – use by someone else does not make it inaccessible
- Searchability of content.

Reasons given for preferring to use non-e versions included:

- Easier to read and to handle
- Better for detailed consultation

- Material form is of significance
- Value in the collocation of material in libraries and archives, together with advice from expert staff
- Browsability
- Better for dealing with multiple items
- Fieldwork better face to face.

Many researchers said that it depended on both the resource type and the purpose for which it was being used, and their preference for using one format or the other depended on the circumstance, for instance they are happy to use journals in e-form because of the convenience and downloadability (though many still prefer to read a printout rather than on screen), but prefer to use books and other materials in physical form. While many respondents said that that they preferred to use manuscripts only in physical form, there were some who stated that the accessibility of digitised versions together with the ability to manipulate and enhance images brought major benefits for manuscripts available in e-form.

There was a marked difference in preference among those FBAs contacted by post (i.e. who did not have an email address). Of these 73% expressed a preference for use of non-e forms. This reflected that many of them (65%) were retired, in some cases did not use e-resources, and were content with the traditional methods of research to which they were long accustomed.

Electronic – quicker and more convenient than going to libraries. (**Sociology**)

Electronic material (video/DVD) tends to be more easily accessible than celluloid film. (**Film history**)

I prefer electronic resources because I can access them immediately without moving from my desk, and when working at home. (**Cognitive psychology**)

The problem with e-journals for art historians is that copyright issues often mean that images are not reproduced in e-journals, rendering the articles useless. (**Art history**)

Regarding books, when they are readily available I do confess to a Luddite fondness for the real thing. For journals, I find electronic archives extremely useful. (**American literature**)

This varies according to the precise nature of the object; very good scans of manuscript material might be preferable to looking directly at the object (since one could easily return, copy for study purposes etc); for printed material when one is scanning an entire run of a journal, it is still preferable to have the original copy to hand. The main point would be that the generation of electronic resources should not be seen as an excuse to dispose of or disperse collections of hard-copy originals; rather, electronic resources should be seen as a useful, but qualitatively different, addition. (**History and philosophy of science, technology and medicine**)

Non-electronic. Easier on the eyes and small details not always recorded electronically: "unimportant" articles, even adverts and notices can be important. (**Modern European art, thought and literature**)

I prefer to use non-electronic versions because: (1) my research requires the detailed consultation and handling of early printed material, noting features which often do not show up in electronic versions; (2) reading texts online is very tiring, and in these circumstances it is easy to miss important details. (**Seventeenth & eighteenth-century French literature of ideas**)

Non-e. Habit and aesthetic pleasure (**Medieval history**)

**Q4. Are you using e-resources that are wholly new, or sufficiently different from their non-e versions to offer quite novel research possibilities?**

65% responded 'No' or did not reply to this question. 35% replied 'Yes'.

Examples given of types of e-resources used included:

- Major reference works such as DNB, OED and Grove
- Large-scale compilations of digitised books such as EEBO and ECCO
- Full text databases, e.g. Literature Online, Acta Sanctorum, Patrologia Latina
- E-journals
- Digitised back runs of journals, e.g. JSTOR
- E-print archives
- Corpora of texts, inscriptions etc.
- Concordances
- Data archives and statistical datasets
- Geographic information such as GIS, aerial photographs, digital mapping, satellite imagery, site gazetteers and digital terrain models
- Law reports and other legal materials
- Online versions of newspapers and broadcast news archives
- Image, sound and video materials.

Also identified as important in relation to novel research possibilities were tools and capabilities for discovery, searching of content, downloading, storing, manipulation, analysis, and 3-D modelling.

Responses indicate that for many HSS researchers either the resources that they use are not yet available in e-form, or where they are, they are not yet sufficiently different from their physical counterparts to change the nature of the research or to offer novel research possibilities. However a substantial minority of respondents confirm that such new or different e-resources do exist and have indeed offered new research possibilities.

I am beginning to explore using 3-D modelling of buildings and computer replications of lighting effects. (**Byzantine art history**)

Yes, interactive survey data, newspaper archives world wide. (**Sociology, Anthropology**)

Yes – digital versions of government documents allow one to perform your own analysis on them (e.g. coding voluminous documents for subsequent quantitative analysis. (**Public policy and administration**).

Electronic data sources are essentially new in their richness and scale. Easy access to journals turns quantity into quality. (**Economic and social history**)

The A2A site has opened up a wealth of searchable catalogues for archives a cross the country. This has made locating interesting material much more convenient, although many detailed catalogues still have to be consulted at the NRA in the PRO. (**Early modern history**)

Yes – Neuro-imaging databases such as the one at Dartmouth in US. (**Psychology**)

Not yet, but would like to have some! (**Classical archaeology**)

No – I have been manipulating data in ways impossible before computers, but my sources are as yet traditional. (**Naval and maritime history**)

e-data archives allow large scale research that was never previously possible. (This is) increasing the amount of historical research done by social scientists who are not 'historians'. (**Economic geography**)

No, but there is potential in my field, prints and drawings. (**Prints history**)

Manuscripts – some of the manuscripts which I study have been digitalized and they are wonderful because they have historical annotations, translations, and other written comments not found in the original format. (**Modern Languages – Spanish and Portuguese**)

If a reading is unclear in the original mss, you can 'do' things to the electronic image to help make that reading clearer. This is a major advantage. (**Medieval Welsh poetry**)

I don't think I understand the question. Using electronic resources has made all sorts of things possible that didn't used to be possible. But I have had access to a computer since 1982 and barely remember what it was like to do research without electronic resources. Listing, alphabetising, searching, looking at library catalogues, tagging, writing, editing, spellchecking, incorporating images, drawing graphs.... EVERYTHING is different from how it used to be. (**Linguistics**)

I use video-clips and animated short movies (to be played on a computer) developed by the Language and Cognition Group at the Max Planck Institute in Nijmegen to elicit linguistic data from native speakers of Quechua. These tools make the elicitation of certain types of data much more efficient than without such tools. (**Linguistics**)

Yes, detailed images (of three-dimensional objects as well as maps and manuscripts) allow one to zoom into details it is otherwise hard to see. Such resource enhancing e-resources are potentially available for materials I research, but are not really employed by museums yet. (**Classics, especially Ancient art**)

The possibility of contrast, playing with size and more than one person looking at once at the screen means that the exercise of decipherment has changed (**Ancient history – papyrology**)

**Q5. What other primary resources (both electronic and non-electronic) do you sometimes use?**

Respondents gave a long and diverse list of resources used. Almost all of these were mentioned by other respondents elsewhere in the survey. Among additional resources mentioned only here were: online genealogies; tombstones and other monuments; material culture data; guidebooks; human movement databases; oral history; time-diaries; focus groups; experimental data; scientific and microscopic analyses; transcriptions of oral testimony; film posters, publicity documentation and censorship files; professional handbooks; opinion surveys; online compendia of historical dates.

The responses confirmed the variety of resources used by and of importance to HSS researchers.

## Tools, software, research mechanisms

**Q6. Do you make significant use of data processing programs or tools?**

68% of respondents said they did not use such programs or tools, or did not reply; 32% replied 'Yes'. The main tools mentioned were:

- Statistical software, especially SPSS
- Qualitative data analysis software, e.g. Atlas /Ti, NVIVO
- Geographic information systems (GIS)
- Image manipulation and enhancement tools
- Concordance software
- Personal bibliographic software, e.g. Endnote
- Spreadsheets, especially Excel
- Desktop database management, e.g. Access.

In general it appeared that data-processing programs were more commonly used in SS disciplines than in AH, particularly for the processing of statistical information. The use of personal desktop software was mentioned here by some, but arose more frequently in response to Q8 on whether and how research had changed through the use of e-materials and e-capabilities.

> I do large amounts of statistical modelling; I use computer simulation tools to analyze econometric methods; I have designed and use automatic modeling methods; I use mathematical word processing and checking tools; I have own electronic bibliographic databases; and indexing systems for authors and subjects in my books. (**Econometrics**)
>
> I use a database called Shoebox specifically for the management of anthropological and linguistic data, and plan to use software packages developed by the Max Planck Institute for Psycholinguistics in Nijmegen for transcription of audio and video files and the management of metadata respectively. (**Linguistics**)

**Q7. What traditional resource discovery tools, i.e. secondary resources, do you rely on most?**

The major traditional resource discovery tools mentioned were bibliographies (67%), library catalogues (58%), and abstracting and indexing services (26%). The importance of information provided by colleagues, both within the home institution and elsewhere, was mentioned by 46%. Other tools mentioned included footnotes, reviews, discussion lists, publishers' catalogues, and browsing in or talking to expert staff in libraries, archives, etc.

Since the greater part of resources used by HSS researchers in many disciplines will remain in non-e form, these traditional discovery mechanisms will remain a vital and fundamental part of the research process, and an essential part of the national and international infrastructure needed to support it.

> Over the years my reliance on the above has declined significantly, to be replaced by electronic search engines. (**Geography**)
>
> Colleagues are certainly not 'secondary resources'! (**Archaeology**)

**Are there e-tools for discovery that are particularly important to you?**

Those mentioned most frequently were Google and other search engines (47%); online catalogues (18%); A&I services (11%); Web of Science/Knowledge (10%); others mentioned included alerting services, portals, and listservs, as well as specific services such as JSTOR.

Lexis/NEXIS and BIDS. 20% replied either 'No', or that there were not really any e-tools for discovery that were particularly important to them.

Again, responses from the FBAs contacted by post had an impact on the overall figures. Of these FBAs, 42% said that there were not really any e-tools that were particularly important to them, reflecting that many of them are retired and are not significant users of e-resources or methods.

The responses showed search engines, particularly Google, frequently attracting use as the first resort approach for resource discovery, ahead of more traditional tools such as online catalogues, abstracting and indexing services, or library or specialist services such as subject portals

> Yes – ARCHWAY and HEIRNET/HEIRPORT are invaluable, and of course Google. The Archaeology Data Service provides a fantastic service. (**Archaeology**)
>
> The most important research tool for me is Google, probably: for getting linguistic examples; for finding on-line material, and for finding out (presumed) facts. (**Linguistics**)

**Do you have any difficulties in accessing or using these e-tools?**

59% replied that they did not have difficulties; 22% said they did not use such tools or did not answer, while 19% reported some difficulties.

Difficulties reported in accessing or using e-tools included:

- Resources not subscribed to by institutions, e.g. MLA, Historical Abstracts
- Lack of necessary or up to date software, e.g. Acrobat, Mac browser
- Difficulties with or inadequacies of specific systems or services e.g. Lexis/NEXIS. BL catalogue
- Difficulty of knowing/finding what available, e.g. music resources
- Lack of confidence or expertise, e.g. because of unfamiliarity, age or unavailability of training
- Resources not accessible off campus
- Technical problems, e.g. network mapping, non-Roman scripts
- Personal difficulties, e.g. RSI
- Quality/reliability issues, e.g. of Web sites.

In general, those who use e-tools for discovery were confident in their ability to use them effectively, although there were adverse comments about the quality and usability of some online library catalogues. Some major resource discovery tools remain available only on a subscription or membership basis, e.g. Web of Science; OCLC; the various databases hosted by the US Research Libraries Group on its RLIN system; the Hand Press Book Database hosted by the Consortium of European Research Libraries (CERL). Researchers whose institutions do not subscribe to the resources, or are not members of the organisations concerned, have difficulty in gaining access to these important discovery tools.

The difficulty is that there are far more musical resources out there than I, or anybody else I know, knows about. Too much effort is put into creating new tools as against auditing what exists and making them available. (**Musicology**)

I use Google, metacrawler and other usual search engines – we all know the politics of search engines. It would help if the European and other Academies created THEIR OWN SEARCH ENGINE and PORTAL so that we do not have to use Google to get academic information. (**Communication and media**)

You get to use what is available, learning new sources can be difficult and rather random. (**Sociology, Anthropology**)

Many resources I have seen in the lists of AHRB successful grants have been difficult to access. (**English Literature 1660-1830**)

Portals generally a disappointment. (**Naval and maritime history**)

Those difficulties common to my generation (b 1930) of handling electronic materials. (**Ancient Greek literature and science**)

The search options on electronic library catalogues do not always seem adequate to handle works with multiple editions and complex textual histories, nor to search for early printed books with variant spellings of key words in the title. (**Seventeenth-century English literature and manuscript studies**)

Sometimes systems in foreign libraries are not transparent, due to language difficulties. (**Typography and graphic communication, history of**)

Portals/websites are always hard to get to know: each works differently and one has to learn proper terminology for use on particular site/project. (**Classics, especially Ancient art**)

**Q8. Has the way you do your research already changed, or is it changing, through new e-materials, programs, e-discovery tools, or basic e-capabilities?**

68% of respondents said their research had changed; 32% said that it had not changed at all, had not really changed, or did not answer the question.

Responses to this question showed most clearly the impact of ICT on active HSS researchers in recent years, through improved discovery and access to a much wider range of resources, through the importance of email for communication and in facilitating UK and international collaboration, through personal 'office' capabilities, and through access to and use of digital content. Many personal instances were given of how research had been changed or helped through e-tools and capabilities. Examples fell into the following categories:

- Discovery of a much wider range of material through search engines and portals
- Ability to identify and locate specific resources through online catalogues and Web sites of libraries, archives and other holders of material
- Access to many more resources, on a worldwide basis
- Improved convenience and speed, saving preparation time, avoiding wasted journeys, and allowing instant access to content
- Ability to work when and where the researcher wishes, whether at their institution, at home or when travelling through affordability of home computing and broadband access, and through the portability of digital devices and convergence of technologies

for text, image, sound and communication (particularly useful for fieldwork and data collection in archaeology and the social sciences)

- Ability to capture, create, store, search, manipulate and enhance data content using standardised personal 'office' tools such as word-processing, databases, spreadsheets and bibliographic software
- Facilitation through email and discussion lists of communication and collaboration with colleagues nationally and internationally, including on large scale and interdisciplinary projects, on organisation of conferences etc.
- Facilitation of the peer review, publication and dissemination process
- Improvement of the presentation of results and of conference papers through the use of presentation software.

Many examples were given of the e-resources that are transforming research, overlapping with and confirming those identified in Q4, such as Web sites; content databases; digitised materials; journals in e-form; large scale reference works; report literature; official documents; GIS; digital maps; online economic statistical data. Also mentioned as important were the capabilities to search, process and manipulate content and data, and software that allows for large-scale data processing and analysis that would simply not have been possible by traditional methods.

A number of younger researchers commented that they have always had these tools, and therefore the way they do research has not changed. At the same time many FBAs contacted by post retired before the use of ICT and e-resources became so widespread, so even if remaining research active, they have not found that the way they carry out research has changed significantly through the availability of e-tools and resources.

While it is now easy to take for granted the availability of the World Wide Web, the pervasiveness of email, and the power and affordability of personal computing capabilities, all of these are relatively recent phenomena within the career span of the older generation of scholars, and the impact and benefit these developments have had in assisting the research process, improving personal productivity, and facilitating the production and dissemination of research outputs should not be underestimated.

I have had access to more resources than ever before, particularly in other fields of study. My literature searches are much more efficient now that I need to spend less time hunting through journals in the library and photocopying them. (**Sport psychology**)

Email allows me to contact quickly colleagues in my area all over the world; Google etc. allows me to do preliminary, fact-finding research from my office. (**Renaissance literature and culture**)

Fast and easy electronic access to journals is crucial for update in the subject area. We need more electronic subscription from the university, as it is extremely expensive to subscribe at individual basis. (**Bioethics, medical ethics**)

I now make extensive use of online library catalogues worldwide to locate early editions of printed books. This makes it possible to carry out preliminary research prior to a research visit. (**Seventeenth & eighteenth-century French literature of ideas and history of the book**)

Very much so. Email, searchable databases, and particularly electronic text processing have made my life as a researcher far easier! … the problem nowadays is not so much access to information as being able to process extremely large quantities of complex information and to compare and assess the validity and utility of sources. (**American literature**)

Survey and focus group analysis has always been electronic. So the big CHANGE has been literature/info searching by electronic methods… That has immensely increased the scope & reduced the time needed for embedding my own empirical research results in "theory" and "comparative context". It sounds trivial but it really is a case of quantity (or speed) generating quality. (**Politics**)

The changes in publishing process. All material is submitted electronically and all exchanges of material for publication (other than proof reading books) are now electronic. (**Philosophy**)

Databases allow quantitative processing of large quantities of data. Search engines and dictionaries associated with large archives enable finding and identifying of cuneiform tablets. Digital photos substantially reduce time away on research trips. (**Cuneiform studies/Assyriology**).

I am young enough that my research has only been conducted with the use of these resources. (**Political science**)

I am part of the generation that 'grew up' with email and Google. I cannot imagine life without it. My research begins and ends at my computer screen. However, books are still the most important materials for a number of reasons and these, even if available on line, will never be replaced by digital versions. (**Communication and media**).

Web portals make a huge difference, but being able to communicate with others by e-mail is very important. I co-edited a book with a colleague in Illinois whom I have never met! (**Developmental psychology**)

Yes. Access to primary data is much easier than it was five years ago – e.g., official statements, government data, local & national newspapers from Russia/Germany, etc. are now available within days rather than weeks or months of publication. (**Politics**)

Yes – enormously – systematic searches of much larger corpora of data have become possible. (**Anglo Saxon literature**)

My research has been revolutionized by the digitization of 19th century newspapers. (**Modern Jewish history**)

It has not changed drastically. Paradoxically, one of the best uses of e-resources is for finding and ordering library books and journals. I use email for research purposes. Overall, email is as much a curse as blessing. (**International relations**)

I am compiling digital data (spatial data, image files, databases) as part of my current research. The availability of e-tools and materials is changing how I work at the strategic level (by enabling choices for the future storage of different kinds of data), as well as the everyday (provision of more of own resources in electronic form, as well as access to e-resources). (**Classical archaeology – east Mediterranean and south-east Europe**)

Yes, absolutely – it's become much easier to find and identify primary and secondary material through the use of portals. It's also becoming easier actually to search primary texts online with the recent publication of the Eighteenth Century Collections online. (**English**)

Yes – through scope for large scale analysis and search of historical and current documents. Statistical records that would previously have taken years to search and can now be searched in minutes and large scale databases constructed. (**Economic geography**)

Absolutely transformed. One can find out relevant articles, reports and other materials world wide within literally minutes; ability to search library catalogues means I make more use of books and of a wider selection of books since I can easily find what's relevant in departments other than my own. I can plan my use of library materials ahead of time, order or reserve books. (**Museum studies**)

Ability to deal with quantities of tabloid & numerical data, especially statistical analysis of historical sources. (**Medieval landscape history**)

Yes – EEBO is transforming by making pre-1700 printed material so easily available; more and more material on web that is useful; the possibility of a virtual research environment for the humanities also offers the prospect of further significant change. (**Early modern British history**)

Since the 1970s my research has been revolutionized by use of computer corpora of text and spoken transcriptions, & associated software. Like most academics, my way of doing research has changed enormously through the use of e-mail and word processing software these days I engage in team research, including international team collaborations. I am poor at statistics, but now elementary statistics are essential to my research. (**English linguistics**)

## Resource creation

**Q9. Are you yourself an e-resource contributor or creator through the provision of data or of literature?**

51% of respondents said that they were e-resource contributors or creators; 49% said they were not or did not answer the question.

The following categories of e-resources created or contributed to were mentioned:

- Articles published in online journals, and/or membership of editorial boards
- E-books/online dissertations e.g. through Oxford Scholarship Online
- Papers etc. on personal, departmental or institutional Web site
- Content of project or external specialist Web site
- Production of digital catalogues or indexes, e.g. Persian manuscripts
- Contributions to online databases or corpora, e.g. Fontes Anglo-Saxonici, DNB
- Creation and deposit of files/databases available via UKDA or AHDS
- Transcriptions/editions of texts and manuscripts
- Organisation of discussion lists
- Creation of software/tools.

I work with the IADB (Integrated Archaeological Database) on the Silchester Roman Town Life project which stores all the information on the excavation (plans, context cards and finds). This is a crucial tool in the preparation of the written report. (**Roman archaeology**).

(With a colleague) I will be putting our transcriptions of mss of an Old Norse poem on the web shortly, on the international skaldic project website. We will also try to make available digitised manuscript images. (**Medieval literature**)

We have already created two databases which have been made available over the web by the History Data Service and have two more in preparation which are likely to be disseminated by the same means. (**History of cartography**)

I e-publish my own articles and some chapters on my own website. I tried hard to get Oxford Colleges and faculties to publish their member's thoughts, with complete unsuccess. (**Philosophy**)

OUP has put two of my books online. (**European History**)

Am in the process of producing a computerized corpus of inscriptions. (**Greek and Roman history**)

Yes – most of my time is taken up in managing the creation of a database of the print collection in the British Museum – a collection that has never been catalogued before (as indeed has no other large print collections). (**Prints history**)

Yes, my dissertation was published on the Dissertation Abstracts International. (**Modern languages – Spanish and Portuguese**)

Yes, all the articles and books that I've published so far are available in electronic format on the net. (**Archaeology**)

Yes. I maintain a website on Singapore English (including selective annotated bibliography). (**Linguistics**)

Yes. I was part of the team behind Virtual Norfolk, which transcribes manuscript and early printed material from the Norfolk Record Office and Norfolk Studies Libraries, and makes the material available on the web. (**Early modern British history**)

From 1970 up to the present day, I have engaged in the creation and annotation of computer corpora for linguistic research. These resources are made available to other researchers internationally. (**English linguistics**)

**Q10. Do you have your own Web pages?**

51% of respondents said that they had personal or departmental Web pages or were currently planning these; 49% said they were not or did not answer the question.

## Technical capabilities and support

**Q11. How do you normally work electronically?**

63% of respondents used a desktop computer in their office, 39% had a desktop computer at home, and 43% used a laptop computer. 8% mentioned using library computers. 8% said they did not use a computer or did not reply to the question.

Many of the respondents had access to more than one machine, with a desktop PC provided by the university in their office, and a personally-owned desktop machine or laptop (often both) for use at home or while travelling. The significant number answering 'none' or not replying reflected the fact that a number of the FBAs contacted by post, many retired, did not use e-resources or a personal computer.

Desktop computers at home and at the university (**Art history**)

Own laptop; I take it everywhere, and would never go back to a desktop system. I also use faculty and library computers when necessary (**Music**)

Computing service user machine in my office (older than Moses and prone to any number of recurring faults) and my own laptop (owned by me and older than my office computer). Unfortunately the two are not compatible and I also do not have Internet access at home. (**Anthropology**)

**Are you able to get technical IT help?**

55% received technical help from an institutional support service, 30% from departmental or faculty support, 13% from colleagues and 8% from family or others, including suppliers. 7% said only limited help was available to them, while 6% said help was not available or was inadequate. In some cases both institutional and departmental support was available.

> Yes – University and School IT support for hardware and generic software needs; support for my e-resource creation project provided by collaboration with colleagues at another university. (**Medieval English literature**)
>
> Our needs are challenging and generally in advance for Humanities. Computing Services have been helpful. (**Classics/Classical art**)
>
> Main problem is keeping virus and firewall software up to date. (**Economic geography**)
>
> Very little support form university due to IT staff being overworked. (**Anglo Saxon literature**)
>
> There is an excellent training system …but staff IT support is very unsatisfactory. Once a PC has a problem it can take weeks to be sorted. (**Viking studies**)
>
> No – I teach myself, I search the internet for help and ask my partner. Our university technical support are no better than me! (**Communication and media**)

**Are you able to get expert guidance on availability and use of e-resources?**

43% replied that expert guidance on e-resources was available from the institution's library services, 18% from institutional support services, and 20% from colleagues. 5% said limited support was available, 12% that support was not available, and 8% were not sure or did not need help.

> Subject-specific information would be exceedingly useful! (**American literature**)
>
> Not really, but there is far too much information of poor quality available. The task is to find useful information, not just more of it. (**Political science**)
>
> Not entirely. I am not sure how to ask the right questions. Most information comes informally from colleagues (**English literature**)
>
> More limited – library staff need training and development – (suspect there is an age/generation effect) – also training for librarians is probably out of date (**Economic geography**)
>
> No – There is still so much to be learned on e-resources that I think most of the university services and colleagues in the UK can´t keep up. I learn mostly through many of my Silicon Valley friends in California. (**Modern languages – Spanish and Portuguese**)

Most respondents appeared to have access to technical support either at institutional or departmental level, sometimes both, though there were some complaints about shortage of staff resources, speed of response etc. Guidance is mainly provided through library or IT support services, or by colleagues. In general it appears technical support is better provided for than expert guidance.

## Gaps in provision

**Q12. What resources or tools would you most like to have electronically that you don't have already?**

Resources or tools mentioned as desirable fell into the following categories:

### Personal equipment

Access to a decent scanner. (**Early Modern history**)

Broadband access at home (**Sociology**)

### Specific software

As a corpus linguist, I would like to have a really good set of tools for manipulating XML-coded text data. (**English linguistics**)

### Improved help/support

More institutional support, especially advice on specialist software involving non-European scripts. (**Middle Eastern languages**)

### Access to existing e-resources not provided locally or accessible remotely

The Consortium of European Research Libraries union catalogue of printed books to 1850. (**Bibliography**)

Acta sanctorum. (**Medieval history**)

### Specific new primary resources

Full runs of historical journals. (**Naval and maritime history**)

Better digital map bases, increased access to primary archaeological data increased access to primary archaeological data. (**Archaeology**)

Database of corporate environmental and sustainable development reports. (**Accounting for sustainable development**)

### Specific new secondary resources

The complete run of catalogues in the National Register of Archives (PRO) for archives across the country. (**Early Modern history**)

Complete museum catalogues online. (**Anthropology**)

### Q13. Other comments

Comments covered topics in the following categories:

#### The questionnaire

> I am an active and up to date social scientist, but still think the questionnaire makes unwarranted assumptions about the centrality of IT and electronic resources to the work we do. (**Russian politics**)

#### Concerns about quality

> Some way of knowing the quality of Web documents. A document you find through Google may be scholarly (whether or not peer-reviewed), or it may be unsubstantiated fantasy. Which are which? (**Phonetics**)

#### Difficulties of creating e-resources

> Providing searchable text for 16/17C works in the vernacular is very labour intensive: OCR is not really worth attempting; a normalised text is required as well as a transcription because of the idiosyncrasies of spelling. (**Emblem studies**)

#### Optimism about e-resources

> My work has been aided enormously by electronic provision especially since I am retired and my university library is not too good in my field. (**Political economy**)
>
> My field, medieval history, will continue to be transformed by the availability of texts on-line. (**Medieval history**)

#### Pessimism about e-resources

> There are probably no more than 20 people in my field (including graduate students) nationwide, so can't imagine that my/our needs are likely to take priority. (**Cuneiform studies**)

#### Personal difficulties with e-resources

> The general position of the older generation active scholar whose experience has not prepared them for the computer age perhaps deserves a little attention! (**Medieval landscape history**)
>
> It will be obvious that I am not at all advanced in the use of e-resources, and for this reason I fear that this return is not of much use to the Study. But I find that I make more use of such resources every year, as they become available. My demand is supply-led. (**Medieval history**)

### Access inequality

While there are some excellent electronic resources now available, it depends completely on one's home institution as to whether you are able to access them. LION is an incredibly useful tool, for example, but I have been unable to persuade my department to buy it. Spending decisions are often made by considering the areas in which a cluster of academics work. (**Romantic literature**)

As an independent scholar, I am concerned that I do not have the equal access to electronic resources which I do to print and manuscript material. Without being part of an academic network, I cannot access many scholarly electronic resources on my own computer, but need to be physically present in a subscribing library. (**17th Century English literature**)

### Resource constraints/Sustainability

There seems to me to be a funding problem with digital record creation in that there are two elements to the cost – creation/implementation and on-going maintenance. The latter is potentially a significant overhead, the size of which I'm not sure the research community has yet appreciated, and I wonder if adequate provision has been made on many projects. The risk here is a plethora of one-off initiatives that rapidly become dated, and as a consequence may become inaccessible as technology moves on. (**20th Century political history**)

With 25 years of effort of more than 5 terabytes we find that Humanities lacks the facilities to help which would be widely available in Sciences. Our storage problems illustrate this well: no funds to help – left to fund our own back-up. Not really very satisfactory. Please may we have 'science-style' support and profile? (**Classics**)

### Scholarly communication

Academics produce the papers at the public's expense; publishers should not make profits by restricting access to the research output. (**Sociology**)

### Continuing importance of physical resources and mechanisms

Just to reiterate that the provision of electronic resources is certainly to be welcomed; but that it should be understood that these resources must be seen as supplementary to – not replacements for – existing materials of essential use to the historian. There does seem to be a tendency to believe that all materials can simply be made available electronically – which is absolutely not the case for manuscripts, archives, old or rare books etc. It's important that the funds going into digitization are spent well – and not to the detriment of conventional library functions (which seems to be happening). (**History and philosophy of science, technology and medicine**)

**IPR/Copyright**

> It is maddening that copyright constraints, the failure of the music industry to establish workable licensing procedures, and the disregard of most licensing schemes for recorded or printed music, prevent the Web dissemination of resources of no commercial value. This isn't a specifically internet problem – it's the same problem that is creating huge problems for publication in areas like popular music studies and art history. (**Musicology**)

## 3.5 Overview of the survey information

Those who replied to our questionnaire amplified their answers with a mass of detail and comment. We are very grateful for the time and trouble that many took to respond.

The most striking features of the responses as a whole are:

First, that many HSS researchers are making not just effective but enthusiastic use of e-resources, even though important materials are not electronically available. The idea that STM researchers are up to date and HSS ones are stuck in the non-electronic age has no substance in reality. Senior, even retired, researchers are also often as electronically active as junior ones.

Though HSS researchers may experience problems with e-resources that can in fact be easily overcome, e.g. difficulties with screen management, and some may genuinely find that they do not need or gain from e-resources, it is evident that HSS researchers are out there, taking a very determined line about using e-facilities and making their way in research with them. They are proactively exploiting the new opportunities for research that e-resources offer. The negative and backward-looking HSS academic, reluctant to proceed beyond the 1940s typewriter or even a fountain pen, is on the way out.

At the same time, researchers are very aware that the gains from easy e-pickings, for example through using Web engines, have a price in the need for quality control.

The second point is that HSS researchers are making use of a huge range of e-resources, both primary and secondary, as the many unexpected examples in their responses show. The survey, though only indirect as evidence of what e-resources there are, shows how varied they are. HSS researchers, one way or another, are successfully accessing materials that they value, with the frequently-mentioned advantage that they can easily reach remote resources.

This does not imply that HSS researchers do not also rely heavily on non-e resources. This is primarily because these resources are not yet digital (and may be unlikely to become so), but secondarily because non-e versions may be more convenient e.g. for reading, or because digital materials, e.g. scanned images, may not be of sufficient quality for some scholarly purposes. But the responses repeatedly emphasised the added value gained with e-materials by being able to search them conveniently, as well as manipulate them in other ways.

# Section 4: Factors and Themes

## 4.1 Introduction

The two previous sections show how varied HSS researchers are in their needs for resources and modes of work with them, and how varied resource suppliers are. The sections illustrate the extremely wide range of resource types used, as well as of individual resources, and make it clear how many providers and mediators are involved in some way or other in supporting the UK HSS research community.

But there are common factors cutting across the detail that are important for their influence on e-resources and are common to them. These factors may also have indirect consequences for non-e resources, whether through competition, e.g. by pre-empting librarians' time, or through enhancement, e.g. by offering online catalogue access. We consider these factors, or themes, in this section, taking into account not only their current impact but also potential future influence on resources and access to them. The factors can be grouped as primarily technical, covering ICT advances, internationalisation, the Grid, resource developments, access mechanisms, metadata and digital libraries; and as organisational factors including publication and dissemination, publishing, intellectual property, legal and ethical matters, charging regimes, open access publishing, repositories, and preservation and sustainability.

This section reviews the present situation. In Section 5 we identify particular issues arising from the present situation on which action is recommended.

## 4.2 Information and communication technology advances

It is well known that ICT is becoming more powerful, and more pervasive, at a striking rate. This has already had a major impact on the provision of, and access to, research resources, whether through the types and volumes of material that can be held or through the convenience and speed with which these resources can be reached. Web engines that are routinely invoked from standard home computers simply illustrate the general point: it is evident that many HSS researchers are already operating in this style.

Current ICT trends, namely cheaper processors, more memory, greater connectivity, higher bandwidth, wider applications, can be expected to continue, enhancing the infrastructure for research. The particular implications relevant to e-resources for HSS to be noted are as follows.

### Processors

Researchers can expect to have ready access to, if not to own themselves, very powerful workstations with as much power as they are likely to need for any routine work, significant memory, good quality displays, facilities for handling CDs etc. These machines will come along with appropriate and relatively easy to use software for document handling, mainstream database management, and internet access with email, Web browser etc., as completely standard. Ancillary devices like printers are now extremely cheap (even if ink cartridges are not). Laptops are continuing to fall in price (and size and weight) and can thus be expected to be a material convenience for the mobile researcher. Displays are continually improving, making text material more refined and easier to read and allowing better graphic and image presentation. Better displays also have a useful side effect in increasing the number of

concurrent windows that can be exploited, which may be helpful for data comparison purposes.

Portable computers, whether laptop or handheld, are becoming increasingly common, with obvious benefits for researcher convenience: carrying your world round with you and being able to connect to other worlds from anywhere are major gains from ICT developments. Being able to bring such powerful machines to the place where resources are used adds to the advantages that computers offer for note taking or consulting other pertinent sources of information. This applies to research in general, but smaller, more portable, and more robust computers, digital cameras and other special devices with computational capabilities are already of value for more specialised research uses, for example for onsite archaeological recording. With better facilities for, e.g., handling running commentary or automatic location logging, mobile computing can be expected to become an increasingly useful tool for observation and data capture in many areas of social study from child language use to social behaviour in urban space.

### Memory

Vast quantities of memory are available at very low cost. Researchers can thus expect not merely to have plenty of disk space to support their computing at their institution; they can also expect to be able to take advantage of large repositories and databases elsewhere, whether as places to which to supply their own resources for other researchers, or from which to draw resources.

### Connectivity

Connectivity is particularly important for researchers. The development of the Internet and the World Wide Web, along with email and search engines, have revolutionised access to research resources. This includes not only access to discovery resources like catalogues or directories, such as the Academy's PORTAL, but access to end-resources covering many types of data and document, including image and multimedia material, often on a very large scale. The impact of the World Wide Web, and of Web search engines, on access to resources has been profound, and can be expected to continue to have far-reaching effects on the researcher's ability to locate and explore research resources. Though citation is a well-established scholarly concept, the URL has given it a new convenience and power. The Web's downside, the mass of miscellaneous stuff it contains and lack of quality control, require more user assessment effort, but this is more than compensated for by the wide range of useful material. Email is also an important resource, as a means of sustaining scholarly communication.

### Bandwidth

Connectivity even with slow telephone wires is better than nothing, but increasing bandwidth is making it easier not just to send and receive email but to download large data files or image sets, to play video etc., and to do remote computing. The ability to work with high-quality images without being obliged to go to special machines in special places is already an advantage for research, and HSS researchers will benefit, like everyone else, from being able to make connections virtually anywhere and to shift material with increasing speed and decreasing effort. Affordable broadband for the home as well as the office is becoming increasingly important and valuable for the dedicated researcher.

### Speed

More processor power, memory capacity, connection links, and bandwidth are together increasing, to a striking extent, the speed with which computing activities can be carried out, i.e. reducing the time between issuing a command and getting the result. This is very obvious in the way that Web engines can deliver returns for arbitrary searches from vast data files in

effectively zero time. But it applies to many other forms of research activity from data analysis through document formatting to email interaction. There are downsides, with the pressure for instant action promoting stress, RSI, and too-hasty papers, but the overall benefits of more speed are considerable.

## Displays

Display technology has improved very rapidly. Higher resolution screens help legibility and naturalness, whether through finer detail that supports, e.g., many font types and point sizes; through more coherent visual presentation that allows, e.g., smoother line diagrams; or through closer colour separation that enables, e.g., better colour variation and gradation. Screen technology still needs substantial development to obtain really high-quality images, whether still or moving; but it can be expected to gain from development pressures that the multimedia industry exerts.

## Security

As computers become more connected and computing becomes more pervasive, security becomes increasingly important. This is evident in the spread of encryption for routine purposes by ordinary users. But security as a whole has far more significant and comprehensive implications; it is also far more a matter of software engineering in general, and of user care, than of hidden applications of cryptographers' expertise that are miraculously able to ensure safe computing for everyone. Security as a broad term covers both database protection against improper access and damage (say by using robust authentication) and communication protection against improper scrutiny and corruption (say by using solid protocols). It is essential as a means of countering malice, negligence and ignorance, and applies both to system operations in general and to user concerns about confidentiality and privacy.

From the HSS point of view, quite apart from security as it applies to any computer system users, security is important for the integrity, i.e. validity and legitimacy, of the databases they use, including the databases' temporal and derivational status. For those responsible for resources, it is also crucial for access control and for digital rights management, i.e. for the business of permissions and charges.

Security cannot be guaranteed, ultimately because the humans always involved in systems are not guaranteed reliable. Security is an active area of computing research, but as new security safeguards are developed, new threats appear. In general technology developments make it easier to manage resources, for example by backup against failure, and to manage access, for instance by multi-level regimes. But the more sophisticated tools now available are also hard to understand, and HSS researchers, like everyone else, remain ultimately responsible for their own security.

## Applications

There are numerous application packages ranging from basic document production ones, through tools for manipulating text corpora or making statistical analyses of archaeological finds, to advanced visualisation routines. Modern electronic document systems are playing an important role in improving access to publications, whether through deposit in repositories, posting on Web pages, or email attachment. More specialised application tools have not only made primary resources more readily accessible, e.g. as text or image files, but have supported data interpretation that in turn provides new research resources. The main challenges for individual researchers are finding and learning how to use packages and updating this knowledge, and in keeping personal resources up to date in format. Sustainability is a major issue for community resources, but it is also a non-trivial problem for the researcher in relation to their increasingly large quantities of personal material.

### Current research on content access and management

HSS researchers can expect to benefit (like everyone else) from the present very active research on information retrieval and related information management tasks, e.g. information extraction. The continuing development of modern methods of text topic and content access should make all text resources much more accessible than before. Recent research has also made significant progress with speech transcription, so spoken material can be searched with text-based techniques. Methods of retrieving from image files that do not require heavy human indexing are also under active investigation, and there is other pertinent work on retrieval from handwritten materials and music. The OCR technology that may be applied to scanned (printed) image files as a component of digitisation is also important: even scanning, which is now very cheap, is a significant e-enhancement of research resources.

HSS researchers can also expect to benefit from current research into general methods of interpreting really large datasets, for example by clustering or other machine learning techniques. Much of this research has been stimulated by STM needs, but is also applicable to HSS data: thus machine learning has been successfully applied to linguistic data, and automated image analysis methods developed for medicine have been applied in archaeology.

Current research under the umbrella of the Semantic Web is also relevant here. While the scope and value of ontologies, especially general as opposed to domain-specific ones, are matters for argument, other work, notably on what are essentially programming languages for document data, like XML and RDF, is potentially very important for higher-level resource management, for example in conjunction with Dublin Core, just as character-handling protocols like Unicode are at the bottom level.

## 4.3 Internationalisation

ICT has radically changed the space and time constraints on reaching and using resources, and thus has made research resources far more effectively global resources for research as an international activity. This is not to assert that this logical access is matched by administrative access, but ICT is driving a radical change in researchers' perceptions of their resource world.

Scholars have always crossed boundaries in search of their materials. But ICT has made research international in a wholly new and dramatic way. It is now possible to access resources, especially primary ones and not just secondary ones, from the scholar's own workstation, regardless of where a resource's formal owner or any physical correlate (e.g. the book that has been digitised) resides, and of where the digital version itself is stored. The researcher can access e-materials from anywhere, and can also supply materials for anyone anywhere else. In particular, where researchers working with non-e resources, and looking outside their local context formerly tended to look within national boundaries, they now treat local, national and international e-resources equally. Moreover, even with primary resources still in non-e form, the internationalisation of secondary e-resources gives the researcher a much better chance of locating the most convenient or useful instance of a primary resource. Access is also effectively instantaneous, with further profound effects on the efficiency with which resources can be assessed or used.

This internationalisation of resources is important for providers as well as users. It can lead to very different choices and outcomes for resource creation or reuse. Since it could be difficult, formerly, to discover whether some resource existed, inadvertent duplication could occur. It is easier now to promote complementarity in resource creation. This is important not only for what might be described as de novo resource creation, but for improved resource formation by synthesis from multiple sources. The OCLC union catalogue was an early case of such resource amplification made possible by ICT.

## 4.4 The Grid and e-research

The Grid and e-science initiatives in the UK (ESCWWW) started from and are clearly important for STM. The Grid is the ICT infrastructure designed to support very large-scale experimental data manipulation, where material is spread over many machines, and vast calculations, for example massive parallel computations over many linked machines. However e-science is envisaged as more than larger-scale e-*based* science. The presumption is that the significantly enhanced ICT power will create new scientific possibilities, i.e. that more quantity will become more quality. It has already been recognised that quantitatively-based areas in SS should be able to take advantage of the Grid and benefit from the lure and lessons of e-science, and indeed the broader term 'e-research' is now being substituted for the original 'e-science'.

There is no reason to suppose that areas of HSS research that are already e-based, and essentially quantitative, such as econometrics or speech analysis, could not gain from the opportunities to scale up offered by the Grid. This assumes that the methodological underpinnings of quantitative approaches in these areas are sound and would continue to be so when scaled up. There is also reason to believe that more modest e-based research operations could benefit from more ICT power even when this does not involve any very sophisticated quantitative data processing, for example the ability to find examples of particular linguistic phenomena in very large text files. The underlying Grid technology is also an appropriate base on which to build connected multimedia databases. But none of these cases, even the first, necessarily implies a step change in the qualitative character of the research, though this would clearly be good if it occurred.

It is evident that research that is essentially and necessarily qualitative, as much HSS research is, is outside the scope of calculative e-science, though it may benefit from a larger or otherwise richer e-base; and also that HSS attempts to climb onto the e-science bandwagon by pursuing the quantitative paradigm could be a mistake. But HSS researchers should take an interest in the current Grid and e-science initiative and seek to ensure that their particular concerns as genuine potential participants are taken into account. The AHRC already recognises the initiative's wider e-research importance.

## 4.5 Resource developments

Many resource types come in both non-e and e-form. But ICT (like printing before it) has had significant effects in creating resources that are so different in their practical and operational consequences from whatever preceded them that they lead not only to new research lines and outcomes, but also to new modes of access. These novel resources include, for example, linked Web pages, which are sufficiently different from traditional citations and reference lists as to constitute novel assemblies of material, or dynamic databases.

Importantly, ICT makes immediately available resources that are strikingly larger than any that could earlier be handled by individual researchers: this quantitative effect of ICT has had a qualitative impact on resources. ICT is also rapidly enlarging the scope and supply of derived resources, through the ease with which materials can be edited, modified, restructured, reassembled, etc.

This applies not only to individual databases. The facility with which items can be assembled and organised with ICT means that that it is now possible to have resources (Web sites, digital libraries) that are far more complex and varied than many in the past. While many collections and libraries are in fact far richer in their variety of materials than is normally assumed, other resources relevant to the researcher are frequently just of one type – a catalogue, an abstracts journal, for example. As important Web sites illustrate, many resources of different types hitherto to be found in different types of place, or in the hands of different types of institution,

can now all be found as one multifaceted resource, e.g. the International Dunhuang Project (IDPWWW), which brings together digitised images of, and information on, manuscripts, paintings and artefacts from Dunhuang and other Silk Road sites now held in institutions in the UK, Europe and Asia.

One particularly conspicuous development that ICT enables is closer integration between secondary and primary resources. While the classical physical library combined catalogue and books in one place, researchers have long been familiar with secondary resources like bibliographies and abstract journals that are physically disjoint from the resources they cover. This also applies when catalogues are electronic but their primary resources are not. On the other hand, where primary as well as secondary resources are in e-form, proceeding from the secondary to corresponding primary resource can be much more immediate: the Web illustrates this vividly.

All of these developments reflect the fact that ICT has led to virtual resources, i.e. resources that the researcher sees as single (though possibly complex) resources regardless of their physical distribution. Large databases are obvious examples, but the same applies to union catalogues or directories, which not only subsume multiple component catalogues but can also refer to dispersed non-e resources. The effect of ICT has been to create a far more seamless web of resources than researchers had before.

From the researcher's point of view, this unity in diversity is one of the major gains from ICT. The other is that resources, especially secondary resources, are far more accessible than before, in the simple sense that what was remote is now near: it is as easy to use a resource that belongs somewhere on the other side of the world as it is to use one that belongs in the next office.

## 4.6 Access mechanisms

The main consequence of ICT, well illustrated by Web engines, has been to improve the means of logical access. It has done this in two ways: first, through a wide range of content search mechanisms, and second, through the increasing availability of natural language word, and character string, matching as a common, readily-understood, means of searching. Thus on the one hand there are more different kinds of search hooks, e.g. URLs as well as author names, and on the other a single type of search hook that can be applied across multiple representations of a primary resource, e.g. names whether as authors or cited authors.

In general in the past, a particular primary resource or collection of resources had its own particular secondary access apparatus. An individual library had its own catalogue with various metadata fields, embodied in a particular way that constrained the user's logical access to the primary data, perhaps even more than the range of metadata keys in principle allowed. Thus an author-organised catalogue offered no direct access to the content keys represented by titles. A content classification scheme might only be accessible through the hierarchic structure, not by an index on class names. This specialisation of secondary resources, and close coupling of primary and secondary resources could apply even where common protocols, like AACR, were used for the metadata, and also where common means of content characterisation, like UDC or LC, were used.

This approach was continued in early automation, though it became much easier to search on the various metadata keys provided, either as alternatives or in combination, such as subject label and/or author. However the bibliographic and abstracting services were gradually, but also very fundamentally changed, and made much more powerful as discovery tools, by the spread of natural-language word searching to appropriate fields like titles, keywords, and abstracts. This was critical for users who were often not familiar with specialised indexing languages and classification schemes. But the search logic was typically not very flexible, as illustrated by conventional Boolean matching.

More recently, the appearance of full-text searching has revolutionised content access in two crucial ways. Metadata can now be searched in a much more flexible way, using best-match techniques developed for text, and these techniques are themselves being applied to take advantage of the rapidly increasing amounts of full text in both the open (directly accessible) and the hidden (access controlled) Webs. Improvements in character handling have also made it increasingly feasible to search over many different notations and alphabets. At the same time, as Web engines demonstrate, it is possible to exploit URLs as powerful content-grounded search devices.

One effect of this extension of content-based access mechanisms has been to shift the boundary between metadata and data, or between secondary and primary resources: Web pages do not necessarily fall neatly into either the secondary or primary camps but, by being accessible to the same search mechanisms as metadata, extend the user's reach to resources beyond those conventionally labelled discovery resources. The same effect is evident in the way that e-journal databases offer search that ranges easily across tables of contents and articles and not only for individual issues or volumes, but whole series.

## Web engine effects

The Web engines have radically changed the situation for users: they treat everything on the Web in the same way, so access mechanisms are the same regardless of the type of material and wide variety of origins: the analogue would be in being able to ignore all the specific entry systems for different library catalogues and simply search homogeneously across all of them. The whole point of Web searching is to get to any resource, primary or secondary, minute individual or mightily collective, as represented by a Web page, in one single way.

The Web engines also illustrate, in a striking way, the more general impacts of ICT on access to resources by making it much easier to move from one item to another, whether secondary or primary. All of the resource browsing, inspection, checking and other processes that are part of the researcher's larger access process are made much easier, with e-resources, by the ability to click through from one item to another, to have focused display and highlighting, to show multiple items in parallel windows, to scroll and so forth.

The implications of all this for research providers are profound. We are already seeing what may be called 'user googlisation': people expect to be able to search in the way that many Web engines, as exemplified by Google, allow, and to get system responses in the same way. That is, users expect resource access to be in Web-engine style, even though they normally do not know how the back-end technology works, certainly in detail. They expect to be able to use the ordinary language with which they are familiar, to have the system range over the largest possible pool of resources, and to receive conveniently ordered and readily interpreted output. This is true even for researchers who also complain about Web engine deficiencies in response relevance and quality, and are willing to work with other discovery tools, like specialised classification schemes, within the environment of specific subject resource sites. Even with the latter, users expect to be able to switch between specific and generic access modes.

The most obvious implication is for secondary resource providers. Many catalogues, for example, come with their own particular packaging, even if the items in the catalogue conform to 'industry-wide' protocols. Thus Cambridge University Library's Newton catalogue, covering the university's libraries, uses the generic Voyager system but packages this in a particular way for users. From a logical access point of view, it thus differs from other secondary resources that deal not only with the same type of material, but use the same form of item characterisation, and also refer to the same primary items like specific books. Thus the Cambridge catalogue, for the user, is not the same as the British Library catalogue.

Such differences might not matter so much when both the libraries themselves were quite separate but also held their catalogues only in their own locations. It is much less acceptable for

users, at the same single point of e-access, to have to operate within distinct secondary resource subsystems, for no very obvious reason to do with effective content access. The same situation affects other types of resource providers, for example those offering e-access to journals, where both secondary and primary resources are involved. Thus ScienceDirect offers its own user interface which, though designed to capture some of the functionality with which any journal reader is familiar, like table of contents browsing, also embodies a particular view of search query that is not necessarily that of other journal e-sites. Such sites increasingly offer full text searching, like open Web engines, but nevertheless vary substantially in often quite unobvious ways, for example in how simple natural-language queries are interpreted.

The issue for resource providers, especially secondary resource providers, is thus how far they should treat their e-resource as simply an automated version of their non-e one and ignore any logical access styles for which they have not so far catered.

But other implications in the transition from non-e to e- ICT developments mean that research resources have to present new faces to the world. The presumption that, in general, the serious researcher will know where the pertinent primary resources are, largely through informal means, or is willing to engage in slow traditional processes of going to likely places and delving further there, is no longer reasonable. One major advantage that ICT offers is the opportunity to discover new resources directly, or to check more efficiently on the detailed nature of resources of which the researcher may already by aware. This implies that e-resources need to be e-visible, and sufficiently so to attract attention from a distance.

Thus the user now expects that any resource will make itself sufficiently visible through its Web pages to attract the attention of Web engines and appear in the engine search output. The particular problem with secondary resources under this access model is that individual catalogue entries are too impoverished to attract engine attention, while whole catalogues are too disparate. Secondary resources therefore require new attraction mechanisms in the form of appropriately substantive descriptive metadata. But the need to ensure visibility to new access mechanisms also applies to primary resources particularly, but not only, where primary e-resources are embedded in complex sites.

These ICT developments have also had the effect of encouraging direct user access to resources, rather than access mediated by professionals like librarians. Users indeed assume that they know how to find their way around, and frequently ignore specialised tools like complex subject indexes, even if they do use their own familiar technicalese. However the fact that users expect natural and flexible access to resources, and think they know how to use their own language to achieve this, does not mean they search effectively. Many end-users could benefit from thinking more carefully about what they are doing when searching. Thus while researchers may indeed know their subjects, it does not always follow they know best how to search for relevant material. The Web experience has developed search skills generally, but the fact that Web searching is so easy can also only too easily mask incompetence.

## 4.7 Metadata

Access mechanisms depend on search keys, or hooks, that may also be, or be viewed as, metadata, i.e. second-order data about, or describing, first-order objects: thus text labels attached to images are metadata. Similarly, it is easy to see labels on books taken from some classification scheme like the UDC as metadata, because the labels are not parts of the books themselves. In practice, the situation is much more complicated, and ICT developments are having a significant impact on the forms and uses of metadata.

The notion of metadata has itself become more prominent through ICT because of the importance that computer operations attach to labels, e.g. file names or creation dates. As this

implies, metadata may have no direct or close relationship to object content. But as metadata is used for identifying and manipulating objects, it has become natural to treat much conventional secondary resource information as metadata, i.e. as having the functional role of metadata, regardless of whether anything with such a function is also part of the primary object. The use of titles for searching illustrates a borderline case, while that of text words illustrates a complete blurring of the division between meta and object data as between data that is applied to, or drawn from, the object.

Thus one consequence of ICT has been to extend the notion of metadata in functional terms: for resource access it becomes anything that is made available by a system to enable access, and can be exploited by a user to gain access.

At the same time, ICT developments have further stimulated work on metadata in two ways. The fact that deposits and resources of all kinds can become large, complex, heterogeneous and dynamic objects, can be derived or compounded from multiple sources, and subject to many forms and layers of use, has naturally led to an increase in the potentially useful metadata associated with objects, for example, whole histories of who has accessed a file and when. Such metadata is automatically, and hence painlessly, acquired. But equally, the proliferation of e-resources, and especially primary resources, has led to demands for content-oriented metadata that relies on ontologies, as illustrated by the Semantic Web movement.

Some of this metadata work is largely replicating ideas familiar to the library and information world. However the more general extension of metadata types, and enrichment of the metadata associated with objects or collections of objects, is important both for researcher access to resources and for resource providers. For researchers, more metadata per resource item or set constitutes more hooks to reach primary resources, so it is less necessary to understand particular hook types and less constraining to be obliged to use certain types. For providers, it is less necessary to choose types and to instantiate these in a proper and reliable way. The gain from more metadata types is well shown by the earlier appearance of document abstracts as metadata, which provided not just more hooks but alternative types of hooks so that, for example users did not have to understand subject tags and indexers did not have to ensure they were absolutely right.

Richer metadata, more intimately bound to primary objects, has other implications for resource use. The metadata itself may be valuable when treated as primary data. But equally, there is no need to assemble the metadata for a set of primary objects as an autonomous secondary object in the style of a non-e catalogue. It can exist as a logical, but distributed object, open to many angles of approach. The downside to this may be a lack of consistency, the upside a realistic reflection of source variety and the passage of time.

The advantages of flexibility that follow from a larger range of metadata types apply generally, regardless of whether the metadata is seen as oriented towards a more or less specialised clientele. However in practice it may be appropriate to make rather different assumptions about the role and hence the nature of metadata. Thus with specialised databases, sets, and collections, and especially where the secondary resources provided by metadata are close to their primary objects, as with fully electronic resource sites, it is reasonable to use carefully controlled and complex metadata, e.g. indexing thesauri, ontologies, and the like. Though researchers may not always fully understand these, they can in general be expected to know how to use them. However when a secondary resource covers a very large set of heterogeneous primary resources, and especially where a secondary resource is disjunct from its primary one, as with an e-catalogue to a non-e library, it is much less sensible to offer constraining and/or sophisticated metadata. It is also desirable to consider the function of particular metadata types, e.g. article page numbers may not be much help for searching.

ICT interest in metadata has also had a different, but very useful, outcome, in technical work on metadata formalisms, like SGML and XML, that are clearly valuable for resource providers. These formalisms are potentially of particular significance in the especially challenging area of integration across catalogues, and other descriptive materials about resources, of very different types, notably across text and image and multimedia generally.

The important general point, however, is that in the ICT context, resource originators and providers need to rethink their metadata. They should not assume that metadata types traditionally supplied for non-e resources are just fine for the e-world too.

## 4.8 Digital libraries

Digital libraries are currently more a declaration of intent, or perhaps an expression of desire, than an actuality, and what a digital library might be is a matter of argument. At the conservative end of the spectrum a digital library is no more than a conventional library whose materials happen to be held and used in digital form.

At the other end of the spectrum, the Web, along with some access engine to it, constitutes a digital library. Another, rather different version is that represented by a special area or community database or set of associated databases. It is clearly possible, while operating within an essentially conventional framework, to enhance libraries digitally so they come to hold virtual books, e.g. a mediaeval manuscript reconstituted by assembling images of its physically scattered pages, or become virtual libraries over and linking multiple physically or administratively distinct components, e.g. digital versions of all books in Welsh published before 1700. Indeed, digital libraries are often seen as covering materials in different media, therefore so, for example, a music digital library could include recordings, scores, and books on music.

Put more positively, a digital library (cf Borgman FGGI01) is seen as having contents and capabilities, beyond those of non-e libraries, that are enabled by ICT. A digital library has richer content – say through a greater range of materials including specifically digital ones like video, and more functionality – say through multiple content presentation modalities, than either non-e libraries or current e-resource systems like database or information retrieval systems. So though non-e libraries may in principle offer the same functionality as a digital library, in practice they may fail against real-time constraints. Both databases and document or text retrieval systems are limited in resource range or functionality or both. Thus the notion of a digital library is associated with having a large range of materials in easy to manipulate formats, and an equally large range of access and manipulation tools. There is also an assumption that an effective digital library offers very user-friendly interaction and object and content presentation modes. Equally, but informally, a collection of e-resources has to be fairly substantial to rate designation as a library.

Such a view of what constitutes a digital library is not incompatible with a specialised e-site for a particular research community being a digital library; but it is also not incompatible with the Web, as it is developing, being a digital library. However it is sometimes also assumed that a digital library has to be more than a collection of material with some manipulation tools. For something to qualify as a digital library there has to be quality control, over the material it contains and how this is characterised, of the same sort that has normally been applied to non-digital libraries. That is, materials will be selected as worthy of being in the (digital) library, and catalogued in the same sort of way, and to the same sort of standard, as materials in non-digital libraries. On this basis the World Wide Web, as instantiated via some search engines, is not a digital library, though many people use it as if it were.

It is clearly appropriate that special data collections for special purposes should satisfy whatever entry, and description and access, constraints that the purposes justify. It is also

probably reasonable to assume that the digital library, at least in the nearer future, is more likely to be a conventional library in modern dress (or perhaps just with a few new garments or accessories) than a creature of the quite other sort de facto represented by the Web. But it is not clear that investment in e-resources can only be justified by maintaining the degree and type of quality control on materials and their characterisation that traditional research libraries have maintained.

## 4.9 Publication and dissemination

For HSS, publication and dissemination have traditionally been conflated as a means of making an author's ideas and findings known; and formal publication has served both as the means of dissemination and, with the development of learned and academic book presses and journals, as a mark of scholarly quality and peer approval. But first the copier, and then ICT generally, have made it possible to separate publication and dissemination. Much of the current argument about electronic publishing and open access journals has focused, though often in a confused and technically illiterate way, on whether the mere fact of electronic publishing, or some version of it as embodied in open access proposals, necessarily subverts the requirement for quality and approval. For example, it might be supposed that when the page limits imposed by physical constraints of journal size can be abandoned, more papers will be published, diluting journal quality.

Maintaining document integrity and authenticity calls for proper technical attention, but there is no necessary connection between electronic publication and a lack of content quality control. There may be a practical connection, with inconvenience both for writer and reader, when writers post versions of papers on personal Web sites, for instance, that are not identical with the formal published version. But writers have a clear interest in there being an authoritative and locatable formally-published version, and readers equally have a clear interest in checking the precise authority status of anything they take from anywhere but the formal publication source. There is no difference in principle here from the past. The same applies to databases.

The technical complexities of maintaining the authenticity and integrity of e-published items, whether of the more conventional kind like journal papers or databases and collections, are not trivial, and whether protection is against accidental damage or deliberate tampering. Maintaining and propagating authenticity and integrity is a problem not just because of the ease with which e-items can be copied, but because of the ease with which they can be modified, reformed, dis- and re-assembled etc. It is much harder for the ordinary user to recognise that an e-document has been tampered with than when they look at a paper government document with lines blacked out, or a manuscript diary with pages torn out.

It is thus important to distinguish the impact of e-publication on what may be described as peer-review type quality from its impact on what can be labelled object authenticity. Both are important, and ultimately rely on common ICT implementation mechanisms, but require different provider and user responses. In particular, content quality control is the same for non-e and e-resources. This applies, moreover, not only to resources that, e.g., embody poor scholarship or research, and to those that are biased, e.g. may be deemed propaganda, but to fakes and frauds as well.

However, publication, in the simple sense of making something publicly available, is so easy with ICT – compare putting up a Web page with photocopying a paper and sending it round in envelopes – that the issue of e-material status and quality is felt as a general and problematic issue for researchers. It appears not merely in the form of a complaint about the dross that Web engines return, but about the difficulty of checking and validating materials and their providers. The form of open-edit content checking embodied in the Wikipedia

encyclopedia is unlikely to reach many HSS resources. However there is no substitute for caveat emptor: the researcher has to inspect, and judge, e-resources in just the same way that they do non-e ones.

More generally, and perhaps more importantly for research overall, ICT has made publication, in the sense of dissemination of materials, so much more easy as to have a considerable impact on research. Thus quite apart from formal publication in traditional venues (whether non-e or e-) and in new, recognised 'semi-formal' modes, for example deposit in community-endorsed data archives and databases, it is becoming much more common to exploit ICT to publish informally. This can be achieved by putting working papers on Web sites or by circulating them to distribution lists or colleagues etc. Such publication can of course carry IPR with it, and publications of this kind may also be cited. But the important effect is to make publication-as-dissemination increasingly valuable, even though publication-as-'proper'-publication remains vital. Proper publication should of course also be seen much more as making a resource available than as some academic bean-counting act.

At the same time, ICT has encouraged not only a general enhancement of traditional formal publication by other styles of publication. It has also encouraged a specific type of enhancement. This is between the main published object and its backing supports, e.g. an archaeological excavation report and a large file of site and artefact photographs, where only the former is taken as a formal publication. ICT has made it possible to make backing supports for top-level publications far more widely available than before, i.e. to encourage dissemination. Indeed granting bodies, like the Academy, have increasingly encouraged, if not required, electronic dissemination for supporting materials through, e.g., posting on a Web site, with appropriate metadata indicating relationships to formal publications. Such backing supports are often extremely valuable research resources, and deserve encouragement, though they can also raise the issues of preservation and sustainability in a particularly tricky form, because these supporting resources may not be maintained in the same, hopefully robust, way that the main publications are.

## 4.10 Publishing

ICT has had far-reaching effects on publishing, especially journal publishing. HSS researchers draw on a wide, international supply of publications, and many, including journals, can be expected to remain non-e, or without parallel non-e and e-publication, for some time at least. Official publications, especially in English and European Community languages, are already published, and deposited, in e-form: in some cases along with non-e versions, in others exclusively in e-form. However significant, and perhaps more far-reaching, change is coming through commercial publishers, especially with serials.

### Monographs

Monographs have remained non-e for a range of reasons ranging from reader preference for hard copy to HSS faculty dedication to 'real' monograph publication as an academic status marker. Monograph numbers have remained surprisingly high given that prices have risen substantially and library budgets have been under severe stress. The widely-predicted collapse of scholarly publishing has not yet occurred and continues to lie ahead just around the corner. But a crisis does seem to be very near. In the 1960s and 1970s, far fewer monographs were published than now, with routine global sales of 1500 or more. But these sales levels were not sustained, and a declining sales step-curve has been evident throughout the past quarter-century, with a vicious circle of declining sales driving higher prices driving declining sales. Individual publishers have responded by issuing more and more individual titles, but with lower expectations of each. Global sales can now be as low as 250 or 300 in some fields. At some point in the 1990s, the UK academy ceased to be a self-sustaining monographic community: the

subjects that have survived and/or thrived in this context have been those (like economics or linguistics or classics) with international appeal.

Thus even if major academic publishers can sustain some support of less popular titles across their lists, they cannot sustain large numbers of low-sales titles, however highly esteemed, in specialist fields. Moreover though authors now undertake the bulk, if not all, of the copy preparation in many cases, publishers clearly have large real costs both in handling hard copy distribution and, though to what extent they really contribute is not completely clear, in terms of marketing, as well as, usually, copy-editing costs.

The main barriers to primarily e-publishing for highly specialist monographs appear to be the general conservatism of much of the scholarly community, along with the fact that many specialist libraries are unable to handle digital materials effectively. Much of the conservatism, as with journals, comes from the entrenched belief that quality control of the scholarly object itself somehow evaporates with e-publication, even if the whole reviewing and editing process is electronically done; and, further, that this implies that the academic reputation that comes with monograph publication will be undermined. However as e-journal publication with proper peer control continues to grow, the argument that quality control will be lost will become less convincing. Doctoral theses, which have an explicit quality control apparatus, are widely disseminated electronically in some areas, and can be printed at a very much lower cost than that of a hard-copy monograph or other comparable book. Improvements in such aspects as digital image management mean that illustrations are less likely to suffer from this approach than hitherto. There will of course remain many good reasons for widely-used and important books to continue to appear in non-e form, including for convenient library purposes, but perhaps through print-on-demand. Equally, as with journals, monograph e-publication requires a proper repository apparatus.

## Journals

The impact of ICT on journals has been far more important for publishers, in two ways.

The first has been e-publication itself, where STM has led the way with parallel non-e and e-versions of journals; e-only journal publishing has only recently begin to take off, and primarily from particular energetic interest groups or through initiatives like the Public Library of Science (PLoS). Some STM e-only publication activities, e.g. BioMed Central, are important, but the proportion of journals published as e-only is very small and the startup and continuation costs, as illustrated with PLoS, are far from trivial. Continuation costs in particular tend to be underestimated by the small 'amateur' startups now beginning to appear.

Parallel e-publication has been primarily a natural extension of e-production, which is now the norm for major publishers, and is thus also a natural extension of traditional journal publishing practice. But ICT has been a stimulant both through software developments supporting document management and presentation, and through the spread of computing access. Moreover, though parallel e-publishing may have been seen as a natural extension of non-e publishing in both technical and operational terms, it has still required significant resources.

Appropriate business models for parallel e-journal publishing are still under development. While publishers have had a vital role in developing journal markets, there is a widespread perception that publishers charge too highly for access to e-forms when these are additional to non-e ones: this is on the assumption that all the primary costs of production are covered by the non-e form and the marginal costs of extra e-publishing are low or even negligible. Researchers argue (as also with monographs) that they now undertake all the real work of producing electronic copy, while the editing and refereeing peer review process is low cost in real terms or even free. Publishers claim, consistently, that authors and readers significantly underestimate the actual cost of journal production, including, for example all the copy-editing, assembly, indexing, and other processes

that apply to e-journals as much as non-e ones; they underestimate, even more, the costs of longer-term journal maintenance, reference backup etc, that apply beyond the current issue or volume and are vital to a journal's well-being. The latter applies whether or not the publisher hosts the repository through which readers access an e-journal.

Finally, publishers can very reasonably argue that if e-publication increases a journal readership, even if this is at marginal cost in addition to supporting non-e publication, they should get some return from the greater exposure. Indeed, the reader's ability to download and print off particular articles of interest is a direct competitor for the non-e version of a journal that may be costly to produce and requires a reasonable sales volume to be viable, or may need subsidy from the e-sales; though it is not clear whether e-publication has become sufficiently established to be used to subsidise non-e publication.

Publishers are beginning to experiment with e-only supplements to regular journal publication, i.e. as an intrinsic element in the publication itself rather than as a separate subject domain requirement mandated on authors and supported by some subject field apparatus like a data repository. Offloading the backup requirements, through e.g. URL citations, to someone else clearly has logistical advantages for publishers as well as, in general, advantages for researchers in a field. Even though e-publication is becoming increasingly common, it is to a large extent replicating the non-e ring-fenced model of what scholarly journal publication means and what a journal is. The presumption is that proper journals have editorial quality control, authorised version control, and presentational quality control. Many researchers believe that going wholly e- will inevitably subvert these important constraints on journals, whether for logistical or psychological reasons, so both researchers and publishers believe they have a vested interest in maintaining the current logical model of journal publication, and reasons for caution when faced with at least some proposals for open access repositories.

ICT is nevertheless having a significant second effect on journal publishing through the development of publisher repository sites, where some publishers, notably Reed Elsevier, have embarked on major enterprises in which they have invested large sums. These publisher sites, even in their more modest forms, represent a new type of publisher activity. Though they may have started as online versions of sales catalogues and targeted advertising, they have become far more like periodicals reading rooms for browsing current journal issues and checking the latest articles; in the most ambitious cases, they are rapidly becoming whole library periodicals sections. In a concurrent development, these sites have become more oriented to reader needs rather than librarian concerns. Thus, making available current issue information for a whole stable of disparate journals that happen to belong to a single publisher is not necessarily of much use to an individual researcher, however helpful it may be to a university librarian. Sites like ScienceDirect have therefore concentrated on extending their range to include back issues and, increasingly, have made access deals for access to materials from other publishers, e.g. INSPEC and PsycINFO abstracts databases, that increase their total holdings as an electronic serials library.

Quite apart from a subject community's concern about the long-term sustainability of commercial publisher-owned rather than institution-owned repositories, there are clearly major challenges for the publishers themselves in formulating appropriate business models for such explicitly long-term operations.

Running such repositories also calls for new capacities in publishers beyond those of managing large databases. The initial models of user access to journal repositories were extremely conventional, with traditional modes of searching by title, author, etc., along with title page displays and so forth. But repository owners are being forced to keep up with researchers' experience of Web search engines and to provide for flexible, full-text searching. There are also increasing pressures to improve Web engine visibility for the repositories, since no researcher wants to have to visit many separate publisher sites in order to cover the journals in a particular field.

## 4.11 Intellectual property[3]

Intellectual property (IP) includes copyrights, patents and database rights. Copyright is vital for HSS researchers, in affecting access to and use of research resources, as well as in protecting resource creators. Patents are significant for STM, but much less so for HSS, though an economic data analysis procedure might be patented for instance, and there are US patents in text content processing. IP in databases, whether it applies to specific component data or to the ensemble as a whole, and regardless of whether it takes the form of copyright or database rights, is also increasingly important for HSS research.

In the UK, copyright normally lasts for 70 years, less in some cases. It covers intellectual property in text, images and sound, and applies both to content as expressed, e.g. in certain words, and to their presentation, e.g. in the published layout of a poem. While this Review excludes the performing arts and hence property in performance, scholarly research in the multimedia age can easily involve such things as musical recordings. Copyright also covers property in databases, reflecting the intellectual investment in constructing the database, even if its entries are public facts not subject to copyright; the lesser form of investment in databases reflecting effort without creativity is protected by database rights.

Copyright may extend far beyond what is conventionally thought of as characteristic manifestations of creative intellectual effort, as illustrated by a novel. Thus the Ordnance Survey has copyright in the National Grid for mapping. There may be several different copyrights held in the same object and by different holders, e.g. author, editor, etc., in a book. It is important to recognise that while a primary literary source, say, may be out of copyright, and it may be supposed that many objects of scholarly interest are not in themselves subject to copyright, e.g. Chaucer's *Canterbury Tales*, particular versions of such objects can be subject to copyright. HSS researchers have, of course, also an interest in copyright ownership as the authors, editors, etc., of resources used by other researchers.

ICT developments have led to great changes in the operational context within which IP exists, and are putting great pressure on the existing system of rights. These developments are particularly important because of the way they interact with the two components of IPR: moral rights and economic rights. IPR regimes in the past have balanced moral and economic rights in particular ways, and the practical consequences of these ICT developments appear to be shifting the balance again.

The ICT developments have had both direct and indirect effects. The direct effect of ICT has been to question previous notions of what an (IP-relevant) object is, what a copy of it is, and what a use of that object is. Alternatively, these developments can be seen as questioning accepted interpretations of what constitutes IP in editing, presentation etc. For example, while reprinting a book in a different font and omitting some footnotes might in the past have been interpreted as breach of copyright, does digitising a paper and changing the page layout (purely automatically) genuinely embody some new intellectual input? The invention of the photocopier, as early 'ICT', was seen as requiring some adjustment of how copyright, fair use, etc., were interpreted. But both the reproductive and modificatory powers of mature ICT can be seen as far more substantial threats to IPR.

The natural reaction to these ICT developments by at least some commercial owners of IP has been to project, and as far as possible to incorporate in law, the most conservative possible interpretation of what constitutes new IP. Thus a simple change of medium may be claimed to generate new IP, though in practice the rationale is not the new IP itself but the economic value that exploiting that new form creates. This is obvious in the case of music.

[3] The British Academy is undertaking a study of intellectual property. Our comments here are specifically geared to e-resources and access to them.

There are clearly problems for resource users if, e.g., digitising an out-of-copyright photograph is taken as creating new IP (and thus potentially indefinite IP for the photograph owner). But equally, there are problems for resource providers if they cannot claim any rights in the form of a resource which is taken, by those who use it, as having (intellectual) content that they can exploit without any obligation.

## Modernisation

Unfortunately, the present situation is highly fluid with, on the one side, a thrust to legislate in a way that maximises the rights of IP owners, and to back this up with aggressive digital rights management regimes (DRM). On the other side, there are at least some who believe that anything as readily available as e-material is there for the taking. The situation is extremely complex, and is made more complex by rapid ICT development.

This complexity and the challenges that any attempt to develop IP to fit the electronic age present are well illustrated by the Royal Society Report (KSOE03). There is clearly some inertia as well as some reluctance, beyond IP owners' self interest, to think through the implications of the radical changes that ICT is bringing. It is less effort to preserve, as far as possible, the interpretations of rights that were appropriate for the older technology. For example, with a paper publication, copying may be done by writing out again, or by using a copier, to generate a new paper instance. Scanning to create an electronic version gives a form of the object that may never be used to produce a paper one. The current copyright regime means that when scan-on-demand is used by the British Library to produce an electronic copy of, say, a journal article for a user, the internal source of the copy sent to the user is not preserved, as an internal master electronic source for future copies for other users, so as to save the paper sources from wear and tear by scanning again: this would subtly change the particular technological interpretation of the personal copy model developed in an earlier age, regardless of the fact that it is now technological nonsense.

The STM academic and research world is increasingly pressing for IPR modernisation to fit the electronic world; and there are some shifts towards recognition of the need to refresh accepted notions, e.g. of fair use in a way that accommodates technology changes. This is illustrated by, e.g. journal publishers who explicitly allow electronic reproduction and Web site posting of papers for academic and research purposes. But HSS researchers as both resource users and resource providers would, like other researchers, benefit from an IP review that recognises public value as well as commercial gain.

The need for such a review is made more pressing by the other major recent development, internationalisation. ICT implies that where an IP-relevant object 'is', i.e. where it is formally located for identification, authentication and use-authorisation purposes, is far less clear than previously. ICT equally implies that scholarly research is much less restricted than before by national boundaries. Thus legislative developments in the US, motivated by corporate interests and embodying aggressive attitudes to copyright ownership and protection, are having very wide-reaching consequences in limiting research access to resources for anyone other than the extremely wealthy. Unfortunately these changes are being replicated in Europe. There are significant moves within the STM community to counter the weight of commercial interests, e.g. in journal publication, but while these are also relevant to HSS, HSS is affected by commercial pressures in other areas, e.g. news or culture, which are of less immediate importance for STM. HSS thus needs to tackle some important issues in IP from its own standpoint, but in as wide coalitions as possible.

Thus while HSS researchers, as e-resource users, may be anxious to recognise resource originators' moral rights, they are not convinced that the economic rights claimed actually reflect these. They also have an interest, as e-resource providers, in being able to establish copyright themselves, though it may be far from easy to maintain copyright effectively

especially, for example, for a personal Web site without the benefit of any larger institutional support as supplied by a regular publisher, and without, as yet, adequate technology support from DRM. It is thus important, given the ease with which resources can be assembled, that HSS researchers maintain proper procedures and standards when exploiting work by others for such purposes. The same applies to digitisation projects. As with conventional publishing, investments in e-resources should have their legitimate returns, but there are also public interests, both direct and indirect, in fair access to and use of resources.

HSS researchers as resource creators and producers have other IP interests. But as our concern is with resource access, we have not considered these.

## 4.12 Legal and ethical matters

Access to and use of resources is affected by legal, semi-legal and ethical constraints other than those referring to IP. These include the UK Data Protection Act as the most obvious case, but also the Human Rights Act with respect to privacy, the Freedom of Information Act and also, more variously, the sets of principles, rules, conditions etc., that may be imposed on data definition, collection, retention etc. by such bodies as institutional or domain ethics committees. These constraints can be international as well as national.

While some constraints, for example those imposed by ethics committees, might be assumed to bear more heavily on (access to) STM resources, e.g. in the medical and biological area, they also apply significantly in HSS research, for example to sociological survey data. Even where there are no formal bodies with legal or semi-legal roles, ethical considerations can apply, for instance to the use of anthropological data, perhaps over long periods of time.

Such legal, semi-legal and ethical constraints in general apply equally to non-e and e-resources, but may have quite different operational impact, for example in the context of the relatively far greater potential for reuse with e-resources than non-e ones.

The constraints are important and valuable. It is essential to note, however, that as such legal and semi-legal constraints become more established and ramified, especially over time, they (can) have an increasing impact on resources, in a variety of ways. They may do this directly, by modifying their form (e.g. aggregated data only), their retention (e.g. up to time limits), their clientele (e.g. people with status S). They may also do it indirectly, through the expansion of the liability and litigation culture: if keeping data D just might, somehow, some time, somewhere, get you into trouble, then dump D. Resource providers in institutions may, in particular, be obliged by their administrators to adopt hyper-prudential strategies, as is increasingly the case in the US. This is potentially an area where more guidelines addressing bona fide research use, designed to take the potentialities of e-resources into account, might be justified.

## 4.13 Charging regimes

As with non-e resources, charges to users for e-resources, whether directly or vicariously via, e.g. institutional libraries, range all the way from completely free to extremely high. Much of the angst associated with library resource provision, both non-e and e, has been fuelled by the fact or belief that some resources are absolutely extremely expensive, or are far too costly given their small number of users. This has been a particular concern in relation to journal subscriptions, but also scholarly monographs.

Many researchers have the advantage that they do not bear, or contribute to, the cost of the resources they use for their research, since this falls on their department or university, or on a national body. This has been the norm with non-e resources, and free individual access to research resources has almost been regarded as a natural right. Even where there are charges to

individuals for copies, e.g. of a paper or archival document (Boston Spa, National Archives), these are often more in the nature of a handling charge than a true price that reflects the amortisation of capital costs. For example, page copying will cost the same regardless of whether a journal is heavily or lightly used. Larger resources, like whole databases or archives, have also been essentially free at the point of use to individual researchers, as have libraries, though using them for research could well involve incidental costs like travel.

ICT has radically changed the situation, because it is technically possible for users to have their own copy not only of some small resource like an individual paper, but of a large database, conveniently to hand. This copying capability is, of course, exploiting the IP of the resource, and there is reason to argue that multiplying the copies should multiply the charges that reward intellectual effort quite apart from any operational costs of copying. Thus quite apart from electronic analogues for personal subscriptions to non-e resources, the changes to resource form that ICT is bringing are affecting the role of the institution as an intermediary that bears some of the financial costs of access to resources.

In particular, there is, or there is a perception that there is, a growing divergence between the real costs of resource provision, and specifically the costs of e-provision, and the charges that are made for access. Thus, for example, if the e-form of a journal is largely a byproduct of non-e version production, why should it be so expensive, even if having the e-version on tap for any member of the subscribing institution in practice means that it has more users, or more readings, than one paper copy in a specific library does? Equally, individual researchers, whether acting autonomously or through an institution, may feel that charges for a copy of a specific e-object like an image or audio clip bear no proper relationship to either their production cost or their IP value.

In considering charging for e-resources, it is useful to distinguish four regimes:

Case 1: no charge, no conditions

Case 2: no charge, some conditions

Case 3: moderate charge (also conditions)

Case 4: high charge (also conditions)

All of these four cases have applied to non-e resources but their incidence appears to be different for e-resources.

Case 1: many e-resources are absolutely free without any conditions, to anyone with access to a computer. This includes not only many secondary resources (discovery tools) like catalogues, but also primary tools including not only material available on personal Web sites, but that in some repositories. The only conditions are implicit, namely the assumption that users will subscribe to fair use behaviour with respect to citation, and will not attempt commercial exploitation.

Case 2: conditions here are typically in the form of a requirement for user registration and some explicit signature on a proper use form, or the clicking of an 'I accept the conditions' button.

Case 3: this covers, on the one hand, single per-copy payments as charged e.g. by the National Archives, and on the other, the many variations on the theme of a modest subscription carrying with it access to primary resources at either no charge or with only small handling charge. Very large corpora can be shipped at neglible cost for CD-Roms.

Case 4: the obvious examples here are, on the one hand, high charges for images and, on the other, high journal subscriptions, even if the user never goes near the paper version taken by their library and invokes only the accompanying electronic version.

In general the pattern with cases 3 and 4 is that subscription costs are borne on behalf of users by institutions, but individuals may be faced with direct costs for, say, images.

It appears to be the case that, whether as a result of individual researchers' enthusiasm for personal Web sites, or as a result of considered economic policy, or simply as a result of ICT's tendency to leakage, many resources are moving towards case 1, or at least spawn case 1 offshoots: publishers' sites that list current tables of contents, abstracts, and indexes are an example, since these are so much more readily accessible and usable than any hard copy analogues that they can be seen as novel resource effectively existing only in e-form.

The fact that appropriate business models for e-publication are still in the process of evolution, while the ability to police e-resource usage is problematic, means that the relation between charging regimes for non-e and for e-resources is variable. The relationship between non-e and e- versions of the same resource is often perceived to be arbitrary, and charges for e-usage not sufficiently justified by functional differences, and especially by additional functionality in the e-version. Thus being able to search a major publisher's entire stable of journals electronically is not perceived as an advantage when they are nearly all outside the user's own subject area.

However the charging regimes for non-e resources, and particularly rising journal subscriptions, are also perceived as unjustified given that so much of the production effort, from refereeing to paper formatting, is undertaken for free by the academic community, and the spread of ICT has also reduced ancillary costs like postal advertising.

It is not evident that leaving everything to the unfettered operation of free markets, or to the outcome of a guerilla war waged by individual researchers armed with subversive ICT weapons against institutions and organisations, is satisfactory. The public bodies that support research need to be more proactive. These points apply in general, though additional arguments may apply for public sector information (cf Weiss BCCP02).

VAT on e-publications but not non-e ones is also not rationally motivated from the research resource point of view, and compounds the unsatisfactory charging picture.

## 4.14 Open access publishing

Open access publishing (OAP), which is seen as primarily or exclusively electronic publishing, is becoming an issue for the STM community, which has been taking the lead and proposing OAP as a new model for journal publication. Thus open access (OA) to research outputs, notably journal papers, is being aggressively advocated in STM (HCSC04; WTPS04). There has so far been little pressure for OA from HSS. But the issues are important for HSS partly because the same publishers may be involved, partly because changes to funding mechanisms for research publication (from reader pays to writer pays) would be likely to spread across all funding agencies and institutions regardless of subject area.

In assessing the possible implications for HSS it is essential to separate distinct factors that are often confused, for example associating an OA remodelling of conventional journal publication with having institutional repositories. Publications have to be somewhere to be accessed, but this does not require institutional repositories. Again, OA applies in principle to non-e publication, but it applies much more obviously to e-publication because ICT allows more disaggregation and redistribution of production and access functions as well as easier access to materials. Thus we focus on e-OAP here.

OA need not be confined to journal articles. It can apply to books or, more importantly, to any type of material (e.g. a database) or form of publication (e.g. a technical report, a program), and formally retain whatever status and value such things are recognised to have. OA can also refer

to other materials made available to the reader which are not published in the conventional sense, and the Web has made it easy to post a writer's drafts etc.

In all cases, whether items are OA is independent of their copyright standing and formal ownership. Thus from both the writer's and the reader's point of view, the formal and pragmatic status of a document are the same whether it is OA or not, e.g. this comes from a reputable journal, these are Joe's latest half-baked ramblings. Similarly the security issues about whether the document to hand is the 'real' one, are fundamentally the same regardless of whether the document is OA or not.

OA, most simply, means that the reader can read an (e-) document without them or their institution paying for it. (This is quite separate from access in the sense of being able to find the document, i.e. logical access.) On the presumption that a paper is a 'proper' journal paper, this would strictly mean that the paper is in the final published and formatted form, but open access is often interpreted to cover pre-prints of the refereed paper, which may not include final proof changes and be in a different format. The term 'e-print' used in the context of OA and repositories is a generic one covering many types of document, and many e-print repositories are hospitable with respect to material type. However OAP discussions are primarily concerned with journal publication, which is especially important for STM, and where OAP proposals directly affect publishers' financial interests. We therefore concentrate on journal OAP here.

### Journal papers

OA is independent of what the host journal is like, e.g. whether it is purely electronic or hybrid e- and paper, how it is owned or run, and whether costs are high or low. Thus there might be purely electronic journals designed to reduce costs to subscribers without their offering OA to the published papers. Similarly, the quality control mechanisms (peer review etc.) that are regarded as vital for research work status are independent of OA, and relate to it only contingently and no more than in many other situations, e.g. ordinary Web pages, in the requirement that an electronic object is what it purports to be and not something else. Again, OA is quite compatible with the existence of publisher resources and services, e.g. large archive databases with a search apparatus, since these add a form of value that is not readily achieved by reader 'self-help' on an assembly of individual OA papers.

However journal OAP has focused on two specific models. One, where a journal (in its e-form) is OA to its readers; and the other, where the journal is not OA per se, but where authors have the right (under various conditions) to reproduce their articles in e-form, e.g. on their own Web pages: the articles could be refereed pre-prints or proper reprints, as would be more appropriate for deposit in an authorised repository. The first model, *open journal*, is very radical, and making the more modest second one, *open posting*, effective is therefore the immediate focus of attention. Open posting does not necessarily imply any author payment.

The OA movement has been stimulated partly by developments associated with ICT. These have included not only a rapid spread of self-help OAP through personal Web sites but, more importantly, the rise of e-print repositories supported by government funding, e.g. arXiv (ARXWWW), or major professional societies, e.g. the CoRR (CRRWWW). But the movement has also been stimulated by the perceived or actual trend in conventional journal publishing of escalating costs without correspondingly increasing value.

On the commercial publishing front, major STM publishers are rapidly developing their e-operations and are investing heavily in quite sophisticated and comprehensive Web sites offering access to large databases including digitised back issue archives. They are also allowing ready access to metadata and limited current material, though access to the full resources is restricted to subscribers or pay-per-view users. But while such publishers are offering new databases and search services, and are also becoming, under pressure, somewhat

more accommodating than in the past in matters like authors' uses of their own publications, all of this added value is not accepted as a sufficient offset to very high institutional subscriptions and the lack of flexibility for libraries that bundling multiple subscriptions in block deals involves.

Publishers seem to be reluctant to allow cheaper e-access-only subscriptions, without any paper, as opposed to paper+e versions; and even where e-only subscription is possible, the cost is still high. There is a very widely-held feeling in the STM community that since author material is prepared electronically, and most of the serious editorial functions including peer review are not paid, there is no real cost justification for the size of publishers' subscription levels. Researchers, institutions, and funders see these levels as impediments to the flow of ideas.

The House of Commons Select Committee and RCUK have endorsed OAP under the open posting model on the grounds (a) that the results of publicly-funded research should be made publicly available and (b) that with OAP more researchers will reach these results than under the restricted commercial regimes, though the Committee is cautious about author payment. The Wellcome Trust is setting an example in taking steps to require OA deposit of funded papers. The proposed Public Sector Information Act may have some future impact on deposit and access to deposits, but exactly how is not yet clear.

## Implications

It is not clear what the financial impact of OAP under the open posting model will be, given that it is strictly parasitic on the existing journal publishing apparatus. But some impact is to be expected. However OAP under the open journal model, as illustrated by the recently-established Public Library of Science, is a more direct attack, since it is being promoted as a way of reducing overall costs and reducing the large publishers' grip on the STM journal markets.

Thus the main focus in discussions of open journal OAP for STM is in the consequences of the alternative business model that it involves for researchers. It shifts charges from readers through their library subscription charges to authors who bear the publication charges. This clearly has many logistical implications, most obviously in STM in reallocating institutional library budgets to research funding agencies for inclusion in grants. But it would bear differently on HSS researchers, since many of these are not grant supported and thus do not have research incomes that could be simply and properly applied to author publication charges. In principle OAP, by reducing overall costs, would help to maintain limited-readership and specialist journals: the costs for authors would be more directly related to the 'pure' costs of production and less at the mercy of a small subscription base. There are many specialised STM journals, but such a development could also help HSS. However, the more modest open posting version of OAP also has business model implications, through the complicated interaction between publisher operations and posting, which will have long-term effects on how *all* papers, whether derived from funded research or not, are treated. Even in STM not all papers come from funded research, while in HSS some at least are from funded projects.

OAP in its widest sense, covering any type of e-print, and ranging from DIY on a personal Web site to replication under warrant of a commercially published article, raises many general issues, notably of content quality control, presentation quality control, and document authenticity and integrity control. But OAP in itself does not necessarily imply any loss of control. The nature and degree of quality control depends entirely on the particular form of OAP adopted, though with e-objects new and different methods are required to maintain document authenticity and integrity. Thus though OAP may mean that (subject to legislative restrictions) anyone can publish anything on their Web sites, OAP does not imply there will be no proper journals. There are equally no reasons why particular copyright regimes should not be maintained, with whatever distribution of rights between, say, authors and journal publishers, is acceptable.

The principled arguments for journal OAP, under either model, apply to HSS as much as STM researchers. HSS and STM researchers have the same interests in control issues. HSS researchers can therefore benefit from STM community initiatives. But the HSS community could gain from being more directly involved, and not leaving OAP matters to national bodies.

So far, there has been little interest in, or pressure for, OA from the HSS community. This is not through conservatism, but primarily because the funding agencies' investment in HSS research is much smaller in absolute terms than that in STM and many individual grants are extremely small; because, while associated publication is welcomed, publication is not seen as a major, or at least immediate, element of research outcomes; because subscription costs for journals are much lower than in STM, so the pressure for alternative mechanisms for publication is much less; and because journal papers in many areas play a much less significant role as research records than monographs do.

The subject range in HSS is very large, so some areas will be much closer to the STM model for publication than others. However while many in HSS seem to see OA as nothing either actually or potentially to do with them, it is important that HSS bodies and individuals consider the implications of OA developments in STM for HSS. Anything that significantly shifts patterns of process and distributions of costs in one part of the academic world is likely to affect others, and is as likely to do this with negative as with positive effects. Thus it would be foolish to believe that, if journal OAP takes off in STM, it will reduce STM journal subscription costs so that there is more money available to purchase HSS journals or monographs. It is also essential to consider such developments outside as well as within the UK, since the UK is not an island as far as academic publication is concerned.

OAP for HSS thus justifies more investigation, both in its own right and as a response to rising library costs that are squeezing HSS purchases. Even if OAP for STM reduces pressure on libraries, and thus might in principle give HSS some relief, it does not follow that this relief will emerge in practice. It would also be sensible to investigate OAP for books, and specifically monographs, since these are important for HSS. (Textbooks are a different matter and are not considered here.) The OAP movement has focused on journals, and the presumption is that books are not candidates for the OAP treatment. But doctoral theses, and massive technical reports, have been routine e-objects in STM for some time, and there is no reason why monographs should not equally be given the electronic treatment. Though the monograph cost structure is different from the journal one, since it is more common to pay book reviewers than journal paper reviewers, and for book publishers to contribute much more, directly, to the outcome product quality, the former is no real barrier to OAP, and the latter is a decreasing one. High quality monographs are already being published where all of the text preparation, to final pages, has been done by authors. However what author charges might be for monographs is not at all clear, and deserves investigation.

## 4.15 Repositories

Any library or collection of documents, whether non-e or e, is in the general sense a repository. However the term is currently becoming established to refer to deposits of electronic papers, or e-prints, belonging to a research community, subject area community, or institution such as a University. In some subject areas, for example physics, e-print repositories are long established and have superseded, for researchers' immediate access to the latest results, the actual journals where the papers are eventually formally published, cf ArXiv. Such repositories depend on deposit criteria, management policies, and mechanisms for updating bibliographic details, file search and, perhaps, new accession notification. They may be run more or less selectively and rigorously, e.g. some may accept technical reports that will not be otherwise published, others may only accept regular journal preprints. The actual computational support may be supplied and managed by government or other service suppliers, for example.

We focus on the technical and logistical aspects of repositories here. It is evident that any major repository effort will have significant long-term funding requirements, especially when the real costs of preservation, curation and sustainability are taken into account.

The open access publishing movement naturally calls for repository support for good practical reasons, though there is no reason in principle why documents should not be held on individual authors' sites. It is particularly important to separate the idea of OA from that of where the OA paper 'lives', i.e. is logically located, and even more from where the 'master' copy is physically located. The House of Commons Select Committee advocates institutional repositories. There are in fact many options for location, including the author's own repository (Web page), an institutional repository, a research community or subject area one, a publisher repository, perhaps a national repository, etc.

There are some recognised criteria for choice. The most important is whether the repository has guaranteed support and medium or long-term life (as a user-accessible store, independent of archive backup), that both writers and readers would endorse. For authors in particular, being able to hand over preservation and sustainability to someone else is very attractive; but the same is true of, for example, smaller learned societies or other bodies that publish journals and do not have the resources or expertise to undertake their own repository operations. From the reader's point of view too, being able to rely on repositories for usable versions of publications, like current-format ones of older papers, is clearly also an advantage.

From this point of view, whether repositories are institutional, e.g. university based, or subject based and run by autonomous agencies, or are publisher based, may not matter as long as the repository is robust, though the choice may not be available for all types of publication. For instance a commercial publisher might not be willing to host working papers but an author's institution might.

However there are other rationales for repositories than those of making research resources available to other researchers and, in particular, reasons that administrators might endorse that are of limited or no interest to end-users, e.g. that institutional repositories are bureaucratically convenient, or enhance an institution's public image. Thus universities have begun to consider repositories as a way both of facilitating access to their research outputs, including metadata information, and of improving an institution's public image (cf the MIT/Cambridge DSpace project). The EPIC Report (DMAM04) also associates repositories with administrative convenience, e.g. in compiling performance indicators.

The other important criterion, from the researcher's point of view, for choosing one type of repository rather than another, is publication visibility. In principle, in the world of Web engines, having a paper on a well-organised personal Web site is as effective as having the paper in some other repository. The function of the repository is to manage preservation and sustainability, with content access provided from the author's site by suitable referencing or, as the paper becomes known, by other mentions and routes from the Web like, for example, citation indexes.

The arguments for subject-based and publisher repositories are of the one-stop kind: researchers can find not only particular papers but sets of related papers within a single (actual or logical) site. Publishers, in particular, may argue that they facilitate access to the materials they host by providing an extensive metadata apparatus which is a valuable means of access to the publications covered. However the same applies to any repository treated as an e-library and run by a library or archive body along the same indexing lines as for non-e materials.

Clearly, where repositories are for particular subject areas, and their advantages for authors are manifest, researchers are unlikely to complain about any logistical effort required to deposit their materials. However the logistical effort required for deposit varies, with some repositories

more stringent than others about permitted formats, etc. It is possible that university repositories will develop in the UK, in response to a variety of administrative pressures. They may be seen (especially by the administering powers) to make researchers work more widely known, but they would appear to be far less likely to do this effectively than subject-area repositories and, if operated in too demanding a style with respect to formats, bibliographic details, etc., to be seen as burdensome rather than advantageous for research authors. Equally, while subject-area repositories offer obvious advantages to researchers seeking information, it is not clear how comparatively useful institutional repositories will actually turn out to be.

If commercial publishers, as agents, run repositories under an OA system, authors are again unlikely to complain about any effort involved in placing materials there; the work will already have been done as part of the publication process itself. National repositories may not be open to direct personal deposit, but may have an important role, through inter-institution agreements, in overall repository provision. However these repositories, like the others, have to address end-user access.

## Visibility

Thus a great deal depends on the search scope offered by repository sites, for example whether access is only to metadata or to full source text. In current discussions, the questions of access visibility for readers are too often ignored. This would not matter if the repository was treated only as a piece of backroom mechanism, designed to help authors preserve their materials while leaving the author to deal with the question of how his materials are found. But many discussions seem to be based on the view that the repository is itself the primary access point to its contents, i.e. runs itself as an autonomous resource with its own cataloguing and discovery apparatus. From the researcher's point of view, a repository needs to be easily found and offer rewards (beyond the specific document) when reached.

Easy finding is critical when searching is not to recover a known item, but to find items of potential interest. Individual Web sites that are visible to Web engines satisfy both needs, though the latter perhaps only to a limited extent. Institutional sites are likely to be unsatisfactory on both counts from a subject search point of view: the seeker after virology papers has no interest in a university's papers in classics. Whether such disadvantages, for user access, of individual institutional repositories may be mitigated by some sort of metadata harvesting for a single, national, higher-level access point depends on the form of the metadata. Metadata designed to suit other institutional requirements may not do much to help access, even if the user only has to work with one unified secondary resource.

Publisher repositories may suffer similarly, since they may cover subject-unrelated journals. Subject-area repositories may be most satisfactory for useful content, though they may not be directly visible to Web engines. Thus the limitations of different types of site are only overcome when these sites are treated not only as independent resource holders, but ensure that their contents are appropriately signalled to such discovery tools as Web engines.

It is therefore essential, in advocating repositories, to provide a convincing answer to the questions: how does the reader-user find the repository, how does the reader-user find their way round the repository, and how much resource value does the repository offer the reader-user? Much resource provision appears to be approached from the supply end, with a natural reference to the world that has resource experience, namely that of libraries, and hence to carry with it assumptions about forms of metadata and modes of searching that are familiar to librarians and not to end users. Users are approaching search from a very different point of view and are developing very different strategies and methods as they become familiar with Web engines. However, while the nature of what is most usefully supplied may change, there is a clear need, with repositories, for effective materials curation. This may not be done in-house, but it has to be done somewhere along the line.

There are of course many other more technical issues that need to be addressed for any type of repository. Thus for any formally published item, the precise relationship between the finally-published form of the material in a repository and the form actually in the repository can vary, and is a critical issue from three points of view. First, ensuring propriety in relation to copyright; second, ensuring that the status of the repository form is clear (e.g. this is the form of a paper accepted by a journal but not guaranteed to incorporate final publisher's version proof corrections); and third, ensuring that where an item is a precursor of an otherwise published document, the relevant publication details, pointers etc. are attached to the repository version. More generally, where materials with different types of publication status can figure in a repository, their status should be clear. In the fluid world of e-documents, richer protocols about labelling and indexing, etc., are likely to be required than in many ordinary libraries. From this point of view, repositories can be expected to have much in common with data sites, and working Web sites generally.

It would thus clearly be desirable, if an institutional (or other nationally-driven) repository movement gets under way, for there to be a uniform policy about repository modi operandi for access and search. Sharing practice on preservation is also highly desirable, and there are obvious issues about sustainability.

## 4.16 Preservation and sustainability

Digital forms of resource were initially seen as a splendid way of reducing resource physical volume and the requirement for costly shelf space. It was also often assumed that once the initial digitisation costs had been borne, further costs were negligible. However it became apparent quite soon that digital media could degrade, and obsolesce, far more rapidly and comprehensively than had been expected, so using digital means to preserve document and data resources was less of a one-way bet than had been expected. Even if resource migration from one medium to another could be largely achieved automatically, there would still be a non-trivial migration cost to be met over time, partly through having to replace the hardware and partly through labour and overhead costs.

It also became apparent that, as ICT became more complex, there would be a corresponding and significant requirement for support software migration. Modern documents (of any kind, whether text, data, or images) normally go through multiple process stages, and the objects associated with any one stage may be far from suitable candidates as the sole items for preservation. However preserving the whole sequence does not necessarily meet later needs, for example for modification or updating, where new processing, and hence the software to do this, is required.

All but the youngest researchers are now familiar with the fact that even some quite recent machine-readable document can need tedious housekeeping attention before it can be printed under their current computer regime. The implications of digital preservation as a much bigger deal than originally envisaged are now a major problem for anyone proposing, committing to, or actually owning, large digital resources. It is a problem, therefore, not only for libraries but for, among others, repository owners, e-journal publishers, and data archive holders.

Thus with digital materials, preservation means far more than simply conserving bits and ensuring that the physical medium supporting the bits is playable/readable, as might be implied by migration from magnetic tape to magnetic disk. At the lowest level, software is needed to interpret bit strings as characters, but technological progress has meant that many more stages are typically involved in conversion from raw bits, or simple ebcdic/ascii characters, to a text that a human can treat as they would a printed page. Modern document systems involve at least two major steps: source processing from input in some document formatting language to create a formatted version, and version processing to create the actual

presentation output for paper, screen etc. In practice, there may be several intermediate steps. All of these steps require software, which is typically complex and substantial. Such application software also relies on support from generic operating system software.

Preservation for usability therefore implies not just the preservation of bits, or even of characters, but the maintenance of, or new provision for creating, subsequent forms. The same applies, in a generic way, to images, so even holding text documents only as images does not remove the requirement for management software. In general, we can say that effective preservation depends on access not only at the bottom level, but at intermediate process levels, and at user readability level. In addition, software itself is constantly changing, so using old materials in later, new, computing environments requires proper upwards compatibility. The same issues of principle apply to Web resources, though the details may be significantly more complicated because of dynamic linking, for instance.

A good deal of attention is now being paid to the preservation problem. But as the practical and technical issues are far from well-understood, it is difficult to develop sensible business models and hence rational political strategies, whether for major institutions or on a national scale. It is tempting to assume that the ICT industry will provide the tools for anything that really matters, i.e. will act for sound commercial reasons (so Microsoft will always support Word documents, say), and therefore that there is no need for the academic/scholarly resource community to attempt to deal with the preservation technology side itself. But as observation of the music industry suggests, this may not be a responsible strategy for national institutions with public roles as the guardians of resource. It is reasonable to assume that collaboration to preserve is appropriate, but in the absence of cost models, especially for the longer term, decisions on actual collaborations are naturally hard to take.

## Sustainability

It is thus important to recognise that the major issue is not the preservation of e-materials, in a narrow sense, nor even of curation as a larger notion involving, e.g., cataloguing, but the much larger one of e-resource *sustainability*. A commitment to digital material involves a long-term strategy not merely for preservation in the simplest sense, typically taken as meaning keeping the bits, or for curation, which might be taken to mean at least knowing where the bits are, but for the sustainability of what is effectively a resource maintenance and access *service*. This is true even if a belt and braces approach to archive conservation has also generated a unique paper print-out stored in an underground bunker somewhere.

Sustainability is not merely an issue for organisations that own their digital material. It is also an issue where organisations rely on others for the preservation function, and especially with reliance on commercial publishers. In general, go-it-alone is unlikely to be a viable option for individual organisations like particular universities and their libraries, when all the ICT requirements are taken into account. Any organisation that does commit to full responsibility for sustained preservation will either also commit to providing all the software for this, at substantial cost, or, by relying on the market to provide, will be taking a gamble. It is tempting to assume that today's popular document software has such a large customer base that it will be renewed into the indefinite future, but this is a large assumption for the kinds of resource of concern here.

Over centuries, non-e resources have acquired sustainability models, whether these are represented by government funding of major libraries, university commitments, regional authority support, or individual volunteers helping in local museums; and these have been complemented by the provision models represented by commercial publishers, learned societies or scholars. These organisational models have been backed up by skills models for librarians, for example.

It has been natural to seek to extend these traditional sustainability models to e-resources, but this is imposing stress on them. More importantly, quite apart from whether this extension is appropriate, the models are independently under stress through financial pressures like paper journal subscription costs.

There are currently a number of investigations under way into business models for specific types of e-resource, e.g. OA journals. There are also many particular projects, for example, the Cambridge/MIT project on institutional repositories, or ones funded by JISC under its FAIR programme, though these sometimes fall uneasily between technical feasibility pilot projects and initial stages of new infrastructure or service installations, in both cases without the development of adequate business models for the longer term. The British Library is also engaged with long-term preservation using Digital Object Management technology, for an envisaged national repository.

Larger sustainability issues are thus beginning to be addressed, particularly in the context of resource support and development over the long term, but much more work needs doing on this. There is plenty of existing experience, for instance in university computing services, on how to run a service which may involve long-term data resources. There is also plenty of commercial experience about the issues of long-term resource maintenance and migration in businesses like banks or insurance companies, that have to sustain their resources come what may. But there is nothing like enough experience of how to sustain long-term e-resources and their accompanying services in the academic and public sector, and not enough commitment to longer time horizons than those for the typical research project, i.e. three years. The ease with which multiple copies can be made shows that ICT, as well as presenting preservation problems, offers a novel preservation strategy. Thus the LOCKSS (Lots of Copies Keep Stuff Safe) Program (LOCWWW) is designed to ensure persistent access by collaborative harvesting, storage and auditing of journal material on a distributed multi-institutional basis. However it is clear that such strategies have to be considered in the light of larger models for life-cycle collection management, embracing all aspects of resource acquisition and treatment over the very long term. Lavoie's analysis (IPDM03), for example, illustrates the need to develop adequate economic models for preservation, not just technical ones, and the challenges this presents.

The US, though in general further ahead with the provision and use of e-resources, has only begun to grapple with the issues of sustainability, especially within the library world that is still the main context for HSS research, as opposed to the STM resource world where funding and government agency commitment can in general be taken as a continuing reality. For the UK, the issues are ones that have to be addressed nationally, with clear leadership. It might be supposed, from the point of view of HSS research, that with anything coded for use on the Web it will be possible to rely on external support for continued usability. It might also be inferred that if large digital objects are made, secondary resources like catalogues may be less demanding, and easier to sustain, than primary resources. Further, one important indirect consequence of making resources electronic should be to increase the user community for resources by making them more conveniently accessible, and hence to increase the pressure and amplify the justification for sustaining them. (This is also one argument for not taking an over-narrow view of IP.) Again, it might be assumed that any application software, and hence electronic format, with a very large user population, is bound to be maintained indefinitely in an upwards compatible manner.

But all of these are mere assumptions without adequate supporting arguments. Much more needs to be done to investigate the range of possible options, without too much constraint from presumptions such as that the future sustainability of e-resources should be some natural continuation of current models of non-e preservation, for example, in their requirements for cataloguing standards and detail.

The Internet Archive (INTWWW) is important not merely in its own right as an archive of what has now become a gigantic e-resource, but as an experiment, of clear interest to both resource providers and users, in e-archiving in a radically dynamic world. It faces, in particularly acute form and in a way that non-e archiving does not, the problem of recording changing state. Dumping every nanosecond is impracticable, but what is a reasonable snapshot interval? Again, as users of the living Web already know, the connectivity structure supplied by hyperlinks is temporally fragile, indicating one form of challenge for searching the archive. URLs are in any case often opaque content keys for searching, so the way Web engine-style searching works out for such cumulatively vast archive is of great interest.

## 4.17 Conclusion

It is clear that the factors we have considered in this section are related in many and complex ways.

We can assume that the external world of ICT will continue to develop, and in new ways which will have unexpected and perhaps large effects on HSS research and its resource providers and users, just as our survey in Section 3 shows us that past developments in ICT have already had.

It is evident that the provision of resources, especially in the shape of formal publication and its concomitants, is undergoing substantial change, with likely consequences for HSS even if HSS is currently less directly affected than STM.

It is also evident that new opportunities for resource hosting and access are appearing, with potentially large redistributive effects between different responsible agents, in relation both to the hosting and access operations themselves and their funding.

The arrival of e-resources presents important new challenges for resource preservation and sustainability, both technical and economic. More generally, e-operations are raising questions about existing models of IPR; and they are also leading, through new modes of resource access, to questions about the value, in the new ICT world, of existing models of resource content characterisation and organisation.

In the next, and final, section of this Review we highlight what we see as the specific, important issues, for the UK, of resource provision and access for UK HSS researchers, and offer our recommendations on them.

# Section 5: Issues and Recommendations

## 5.1 Introduction

The previous sections show:

1) that HSS researchers are already making extensive use of e-resources;
2) that there are many and varied relationships between non-e and e-resources, both within and across subject areas;
3) that historic patterns of resource provision and access are changing, affecting relations between resource users and resource suppliers and among suppliers;
4) that common factors affecting resources across subjects and types, like intellectual property regimes, are being transformed by ICT to both positive and negative effect.

Overall, a great deal is happening, miscellaneously, and much will continue to happen as a by-product of ICT, regardless of formal UK bodies and their decisions. There are, however, important matters that our Review suggests can, and thus should, be pursued as UK policies for HSS resource provision and access.

In this section we summarise the salient issues, and make Specific recommendations for action, under the following heads: resource types; access; open access; technical support; repositories; intellectual property rights; organisation and funding.

We recognise that valuable work may already be under way in areas where we make recommendations. But we believe, even so, that more needs to be done.

We emphasise that many of the recommendations under individual headings are dependent on, or interlock, with those under other heads. These recommendations should also be taken together, and are therefore summarised in an integrated way as Overall combined recommendations in the final part of this section.

We have not prioritised our Specific recommendations as these are addressed to many different bodies. However we flag some Overall recommendations as more important than others, though we regard all as important.

We have tagged our recommendations with their primary targets; but there are other parties to whom they are relevant. We have used abbreviations in the tagging as follows:

"Funding councils" means HEFCE, SHEFCE, HEFCW and DELNI.
"Funding bodies" means these Funding councils, AHRC, ESRC, and other relevant Government departments, Lottery agencies, etc.

At the end of the section we list all the targets with those recommendations addressed to them. This final table also spells out acronyms used for targets.

We provide back-references to the most important preceding sections relating to our Specific recommendations. But we emphasise that these are not the sole basis for the recommendations, which are based on our Review as a whole.

## 5.2 Resource types

We have broadly grouped resources as primary and secondary, and considered researchers' reliance on non-e and e-resources.

The issues here are:

- relative provision for non-e and e-resources;
- relative distribution of e-resource provision for primary and secondary resources.

HSS researchers, far more than STM researchers, have to live in a hybrid resource world, with a complex mix of non-e and e-resources, primary and secondary. Their information needs extend far back in time, and range across many languages. They are often international but also closely tied to specific geographic locations. Unpublished materials (archives etc.) are very important, and in many subjects books are as important as journals among published resources. In many cases the user community for a resource is small, and though making a resource more accessible can enlarge its user community, may still remain relatively small. Finally, the total set of resources of interest to HSS researchers – the deposits of ages – is very large.

In all these vital respects, HSS resources, and their users' interests, differ from STM ones. STM resources are also rich and varied, and STM researchers' interests are manifold. But relatively, in STM there is more emphasis than in HSS on e-resources as a natural consequence of STM's emphasis on currency (or recency), on cumulative data exploitation and experiment, on English as the de facto language of research interchange, and on journals. Thus though the number of STM journals in all languages is large, there are far more HSS books; and though the number of STM resources rising from strictly personal or within-group private status to at least semi-public status through some Web site presence is rapidly growing, there are far more HSS resources overall.

ICT offers a range of mechanisms for resource support and access, but also powerful common mechanisms that can be applied across resource classes and types, e.g. Web search engines. It is, moreover, evident that ICT facilitates resource access and in doing so, enlarges the user community for a resource. Both users and providers gain, from a cost/benefit point of view.

This analysis clearly implies that provision of resources, and especially support for new e-resource provision, whether through creation or access, has to focus on the best value for money. This applies

a) to the allocation of effort and money between non-e and e- resource provision, in particular of primary resources; and

b) to the allocation of effort and money between primary and secondary resources.

In both cases, it applies specifically to what should be provided in the UK because the materials involved originate in the UK. We consider resources of interest to UK HSS researchers but originating elsewhere under Access, subsection 5.3, below.

### Non-e versus e- provision

It must not be assumed that non-e primary resource provision for HSS can be allowed to wither because e-isation will come along. There are many and varied non-e resources that are vital to HSS researchers that are unlikely to be converted to e-form, certainly not systematically.

It must also not be assumed that provision mechanisms appropriate to STM, especially in relation to the balance between non-e and e-resources, will automatically meet HSS researchers' needs.

Thus in relation to the various resource types, and especially those types of importance to HSS researchers, our views are as follows.

## Monographs

(See 2.2; 2.3; 3 Q1; 4.10)

It is unrealistic to expect total monograph digitisation, especially for in-copyright material, or the early appearance of e-monographs on a large scale; monograph resources must therefore continue to be provided, i.e. physically available, in UK libraries. However as, realistically, the money available for monograph purchase is under pressure, more steps need to be taken to ensure that their presence and location is known: see under primary versus secondary provision, below.

**REC 1: that organised, national provision of non-e monographs be maintained, with effective discovery tools for their location.**

(CURL/BL/RLN)

## Journals

(See 3 Q1-3, 4.10)

UK-origin journals are increasingly available, for current issues, in both non-e and e-versions. However back runs of journals are important for HSS researchers, and digitisation makes this type of resource far more available to researchers. Moreover, since HSS researchers are less oriented to contemporary content than STM researchers, digitising back runs, especially of key journals, is relatively good value. Thus high priority should be given to identifying widely used, heavily replicated journals, and also ones that are key for small fields, and digitising back runs under a JSTOR-type model. A coordinated approach to content selection is essential.

**REC 2: that digitisation of key journal back runs be promoted, based on a national strategic collection policy guided by user needs.**

(CURL/BL/RLN/publishers/learned societies)

Since there are many parties involved with journal resources, and also intimate relations between material in, and responsibilities for, the older and newer portions of journals, effective partnerships are needed to enable and support back run digitisation. There are many possible models for organising and funding both the digitisation process itself and for supporting the resulting resources. Currently, digitisation is happening ad hoc, with variable and often unsatisfactory consequences both for the form of the resource itself and for access to it. It is important now to foster collaboration between rights owners, resource managers, service providers, and researcher and subject area representatives, through the development of guidelines and best practice models. These guidelines should cover:

a) technical features for appropriate resource forms;

b) organisational and funding strategies;

c) protocols for access conditions.

a)–c) are considered further under later subsections. Our concern here is to emphasise the value of this type of e-resource and to encourage its provision.

**REC 3: that national leadership and collaboration be encouraged in creating journal back run e-resources through coherent and consistent digitisation practice.**

(RLN/BL)

## Archives (all types of manuscript or very rare materials, text and image)

(See 2.2; 3 Q1, 2, 7, 12)

The total volume of UK archive material is large. Digitisation has been ad hoc and patchy, with the focus on items of potentially large use or of particular interest to some funding agency and user community. It is unrealistic to propose that systematic digitisation, especially on a large scale, be undertaken, even though digitisation often brings new and larger archive use with it. However where digitisation is envisaged, this should as far as possible be well motivated in content choice.

More importantly, since digitisation is often done ad hoc, and apparently often with little consideration for how any specific resource will fit into the larger e-resource world, it is essential to establish and promulgate good practice about digitisation and e-resource form, so as to get a proper return on the investment made in digitisation. By comparison with STM, where ICT is more familiar so digitisation strategies and protocols are well established, smaller HSS organisations or individual researchers are at risk of poor DIY. (The same applies to the design of born-digital resources, see below.) There is thus a need for guidance both on the business of digitisation itself and on the design of e-resources so that they properly enter the larger e-resource world, i.e. so that they are visible, connectable, and usable.

**REC 4: that guidelines and practice models be developed and promoted for archive digitisation that support good forms for e-resources and facilitate e-access for archive material.**

(TNA/National Council on Archives)

## Databases

(See 2.2; 3 Q9)

Databases are increasingly being created 'de novo' in HSS, e.g. as a product of surveys; however databases are also being created that are drawn from archive material, but without being simply archive digitisation. While many researchers are familiar with the requirements for sensible database design, implementation, and documentation, funding agencies also encourage good practice, and the UKDA is a recognised agency for this, HSS researchers are often not sufficiently aware of appropriate methods. Moreover, while in the past reliance on university mainframes may have stimulated design appraisal and sound implementation through advice from computing professionals, the spread of personal computing has led to more DIY and, sometimes, naive uses of commercial database software. The spread of digitisation, say of selected learned society materials, is also involving organisations in e-operations for which they may have little prior experience or technical support.

Proposals, at least in some areas, that data backing up journal articles must be made available, is a further form of e-resource provision that may place a considerable burden on individuals or groups that are not necessarily well qualified to sustain it. It is easy to recommend, and at its simplest to implement, Web sites with resources somewhere on them, but this does not necessarily make these resources computationally well-founded and readily accessible to other researchers, or cater appropriately for longer term preservation and sustainability. These points are particularly important for the many HSS researchers who are not constrained in resource

provision by funding agency conditions, but who at the same time miss out on whatever helpful support meeting such conditions may carry with it from those responsible to the funding agency for holding the resource.

It is important to recognise that HSS researchers are still learning fast and are in many cases highly professional about e-resources, and not to engage in overkill on resource design and implementation. But there is a need for policies and strategies that reduce the risk of unuseful, or disappearing, e-resources. Other issues like access, repositories and intellectual property, for example, are relevant here, and are considered further later. But we think it important to raise awareness about the implications of e-database creation for HSS researchers, independent of whether these are enforced, for some, by funding agencies.

**REC 5: that guidelines and practice models for e-database creation be developed and disseminated.**

(AHRC/ESRC/JISC)

**REC 6: that such organisations as UKDA, AHDS and JISC be encouraged in leading on good practice on forming e-resources, especially databases.**

(UKDA/AHDS/JISC)

## Primary versus secondary resource provision

(See 3 Q7, 8; 4.5-4.7)

The preceding points focused on primary resource provision. Secondary resources, especially traditional discovery tools like catalogues, have a key part in resource access. Modern search engines for e-material appear to have, or have actually, reduced the need for such tools because they offer direct access to primary resource material, especially text. Thus while the search apparatus itself is a discovery tool, this need not rely on any subject-special hooks like topic or category labels. At the same time, engines can be applied, again without formal subject intermediation, to catalogues as text objects.

The points made above under non-e versus e-provision imply that as much existing HSS resource material is unlikely to become digital, catalogues to reach it really matter for discovery and access. Converting existing non-e catalogues to e-form would have substantial advantages in making information about primary resources readily, and conveniently, available, with clear benefits to researchers. E-catalogues can enable a researcher to get a better idea of the range of primary resources that exists, as well as detail about the form and location of particular resources. While printed catalogues are available in many libraries, there are numerous resources which do not have even printed catalogues, or these are themselves not widely available (for example because they were published long ago).

Moreover, while specialised resource catalogues may not always have large user communities, in general a catalogue as a whole will have far more users than any individual resource mentioned in it. Again, while even catalogues can be very substantial, digitising them is much lower in cost-per-item than digitising primary resources.

The arguments for catalogue digitisation are thus very strong. Similar arguments may be made for other secondary resources, like bibliographies, but catalogues are the most important secondary resource for researchers. This has been recognised in the many catalogue digitisation projects in the past, and is also implicit in current e-catalogue creation projects.

However we believe that HSS research would be materially assisted by further catalogue digitisation and, moreover, that in any competition for funds between primary and secondary digitisation, the advantages of secondary resource digitisation should be very carefully

considered. Primary resource digitisation for all types of material – text, data, image, etc. – is clearly important. But the cost-benefits ratios of secondary resource digitisation (or digital creation) are relatively favourable since a secondary resource with potential value to a substantial user community can be established as a whole for the same money as perhaps only a small part of a primary resource; and even a whole primary resource may have a quite limited user community.

There are further important cost-benefit aspects. As noted, once a catalogue has been digitised, it can be exploited (provided it is sensibly implemented) by Web search tools. These can be applied directly by researchers with subject knowledge but without specialist cataloguer's knowledge. Further, the flexibility of electronic search means that there are multiple ways of accessing the same catalogue entry, so high-quality cataloguing is less vital. This may be anathema to librarians, but researchers may gain far more from having ready access to a far from perfect e-catalogue than from having much less ready access to a high-class card catalogue. It is much easier to revise and improve catalogue data than it is to revise primary resource data.

Low quality digitisation projects (other than quick scanning which can be taken at face value without higher expectations) may, for instance, make text material of less value to the user than the original through poor keying, editing etc.; but this matters much less with catalogue material. Quick and dirty digitisation for catalogues would have particular value in making at least some information available about many smaller archive or museum collections, for example. Finally, it may be more easy to get collaborative support for catalogue and other secondary resource digitisation than for many 'pure' primary resources. The A2A project is a valuable exemplar here. A great deal of catalogue conversion has already been done or is under way. But there is still a great deal to do: a recent CURL estimate is of some 14 million items requiring retrospective conversion or cataloguing in its libraries alone (SOMR04). We thus believe that positive encouragement should be given to digitisation for secondary resources, notably collection catalogues, with suitable guidance on sensible, simple, e-formats that will make the catalogues easy to access and use, and to combine with other catalogues.

We recognise that catalogue conversion and similar secondary resource digitisation may not be a glamorous candidate for special funding, though particular, especially collaborative, schemes may be able to make a good case (as with A2A and the Heritage Lottery Fund and the Invest to Save Budget). But our view is that this form of e-resource provision should have a high priority, as a clear service enhancement with a good cost-benefit ratio, within institutions' and organisations' internal funding arrangements.

**REC 7: that secondary resource digitisation, or digital creation, be actively encouraged.**
(collection-owning institutions/Funding bodies)

**REC 8: that guidelines for digital secondary resources oriented towards users and e-search tools, not conventional library practice, be established and promulgated.**
(JISC/RLN)

## 5.3 Access

The issues with access concern physical, logical, and administrative access, and specifically barriers to access.

With non-e resources, whether unique or multiple-copy, there are obvious constraints on physical access. With e-resources, constraints on physical access are largely, if not wholly,

removed, and the opportunities for logical access to both primary and secondary resources are much enhanced, even for primary resources that remain non-e. Administrative access conditions have also changed under ICT especially, though not only, in relation to intellectual property.

## Physical access

(See 2.2; 2.3, 3 Q1)

Physical access to non-e primary resources, particularly monographs and journals in their as-published (paper) forms, is becoming more difficult as library budgets are being squeezed. It is evident that HSS non-e resource provision is under pressure primarily from competition from STM, but also from competition with teaching materials and, to some extent, from internal competition with HSS e-resources, as well as (though not very severely) from increases in non-e item purchase costs. In particular areas, notably non-English languages, the pressure is severe and the distribution of primary published materials across the UK is being reduced to fewer locations. In addition, many of these resources, most obviously monographs, are unlikely to become available in e-form, certainly not soon and perhaps never. Many primary resources required by HSS researchers do not originate in the UK and their producer countries may not be well placed to convert to e-forms. All these points apply across HSS as a whole, but are more severely felt, in general, in the humanities than the social sciences.

It is also evident that this contraction in physical distribution is not being sufficiently offset – albeit with a loss of convenience – by mechanisms like interlibrary loan, since there are problems with these services. Interlibrary loan, and the British Library's document supply service (which is of special value for journal articles), are suffering in a number of ways which are collectively damaging resource access. Thus the core interlibrary loan and document supply operation run by the British Library is suffering because the STM community's switch to e-journals risks reducing demand to too low a level to sustain an economic service, and potentially more efficient collaborative schemes for loans are not yet visible. Again, the stock of materials for loan inside the library system as a whole is being undermined, for several reasons. Some materials are not now being purchased; and some non-e materials are being replaced by e-materials which cannot be loaned as the originals were or, alternatively, copied. At the same time, rationing may be occurring within institutions, whether because of the cost of lending operations or through lack of staff to maintain them. More convenient practical operations need encouraging, e.g. by investigating and trialling end-user ordering for interlibrary loans.

Physical access to non-e archive type resources, i.e. unique or rare materials, is less affected, given that access to such resources has always been non-trivial for most researchers. The main problem here, accepting that many such resources are never likely to be digitised, is that funding pressures may mean there is less staff support for material supply to users.

Thus on physical access to non-e resources, the implications are that libraries need to stand firm, and to be encouraged to stand firm, on their budgets; and that much better national coordination for access is required. STM researchers should not be allowed simply to assume their library will give them whatever they want, as too often seems to be the case, and in a manner insulting to the generally less well grant-funded HSS researcher.

**REC 9: that HSS institutions and libraries should seek to sustain budgets to prevent HSS resource attrition by rampant STM resource suppliers and users.**

(universities/major libraries)

**REC 10: that major libraries, especially though not only the copyright repositories, explicitly recognise and collectively develop their role, and be supported by funding bodies, in underpinning non-e primary resources.**

(major libraries/Funding bodies)

**REC 11: that major libraries coordinate their collection of foreign non-e research materials, and take the lead in organising and promoting interlibrary loan of such materials.**

(major libraries/CURL/RLN)

**REC 12: that the national interlibrary loan and document supply systems for ensuring physical access to non-e-resources be properly sustained, and made less bureaucratic and more user focused.**

(BL/CURL)

Physical access to e-resources is a matter of researchers' access to the computing facilities needed for this. In general, researchers are now quite well provided for by their departments, computing centres and the like and also, increasingly, have their own laptops. However, physical access remains sub-optimal for some individual researchers, and HSS departments need to be more proactive in ensuring adequate ICT physical provision for researchers.

### Logical access

(See 2.2-2.4; 3 Q2, Q7; 4.6; 4.7)

While logical access to non-e primary, and also secondary, resources is being rapidly improved by digitisation, coverage is very patchy. In the previous subsection we recommended digitisation of existing secondary resources as a high-priority objective in e-resource provision; we also emphasise it here specifically as a means of improving access to resources.

From this point of view it is especially important to make catalogues etc., not only e-available, but readily e-available through being easy to reach through Web engines and other Web site access mechanisms, and easy to use through simple, transparent search tools etc. As the volume of secondary e-resources grows, it is essential to make it easy for users to access multiple resources without heavy logistical overheads, and to provide easy to understand, and apply, means of search. Since multiple e-catalogues can be used from the same machine entry point, variations in organisation, style, format, etc., are less tolerable than when catalogues are physically separated and can reasonably be viewed as quite independent of one another. But the correct strategy is not to seek, unrealistically, to impose heavy uniformity on combined or federated catalogues, but to achieve flexibility by allowing for multiple alternative search paths. Thus we emphasise the need to view catalogue design for e-use from the point of view of ensuring effective access, by the way catalogue elements are characterised and organised, for general-purpose search engines, without requiring advance, specialised knowledge of what catalogues (or collections) there are.

At the same time, many e-catalogue interfaces do not provide enough information about how search mechanisms work, for example whether exact case or word matching is required, which matters when systems do not offer the interpretive flexibility that users naturally apply themselves in searching. Improving logical access to research resources implies therefore both that more secondary resources be made electronic and that sensible discovery tools for operating on them are supplied.

Within the UK, improving logical e-access has important implications for particular resources and responsible organisations.

**REC 13: that high priority be given to specific national-level initiatives for secondary e-resource provision and subsequent service support, notably COPAC, SUNCAT, and the RSLG recommendations on a national union catalogue.**

(JISC/RLN/CURL)

**REC 14: that archives holders should continue the buildup of A2A.**
(TNA/archives holders)

**REC 15: that museums e-catalogue provision be continued and enhanced.**
(MLA/museums)

**REC 16: that in creating these national-level catalogue resources, particular attention be paid in design and implementation to the changing environment that ICT developments, notably the Web, imply for effective ways of discovering and reaching these resources, and to the ways of manipulating them that users are most likely to adopt.**
(JISC/RLN/TNA/MLA/BL)

## Administrative access

(see 2.2; 2.3; 3 Q2)

Many important issues here will be considered later, under Intellectual property.

There do not appear to be major new issues in administrative barriers to non-e resource access. The more salient issues concern administrative access to e-resources.

With respect to primary e-resources, there are currently significant difficulties for many researchers in discontinuities between concentration of provision and lack of domain membership (i.e. not being affiliated to the licence-holding institution). With non-e resources, an acceptable administrative compromise for those who are not members of the holding institution has been achieved both by mechanisms like interlibrary loan and by the on-site 'walk-in' access model, for users from comparable institutions and other bona fide researchers. With e-resources, an analogue even of this limited mode of use may be prohibited by the explicit requirement for domain user identity, or given the narrowest interpretation by restricting e-access to specific, on-site library terminals. This may be deemed correct from an IP point of view, however contrary to the ICT spirit. But we believe that with licensed e-materials, a suitable form of walk-in provision could be devised for site licenses that would be very helpful to researchers without undermining IP.

Again, with digital deposit materials, limiting access to these to specific on-site terminals, even for domain members, i.e. replicating the formal non-e model, seems unnecessarily restrictive, and something a little more flexible would be reasonable.

**REC 17: that walk-in use of e-resources should be permitted by site licences.**
(JISC/Licensing societies)

**REC 18: that appropriate models and protocols for using e-deposit materials be developed.**
(Joint Committee on Legal Deposit/BL)

In general we believe that secondary e-resources, especially those with (some or full) public funding, should be freely accessible for free, though charging at reasonable levels, e.g. for cost recovery, may be appropriate to sustain resources. This is largely the case in practice, but should be explicitly adopted as policy. It is not incompatible with the provider, who may have made intellectual contributions, retaining intellectual property.

**REC 19: that secondary e-resources to which public funds have been applied should be readily accessible, for reading and search.**

(Funding bodies/JISC)

## 5.4 Open access

(See 2.2; 4.9, 4.10; 4.14)

As OA is a prominent contemporary issue, we consider it specifically here, though it has not so far been deemed a matter of concern, let alone urgency, for HSS researchers.

OA refers both broadly to free access to any sort of e-print and, more specifically, to access to journal articles (in refereed preprint or reprint form). For journals OA may imply that the whole journal is run on open access principles, or only that individual articles, perhaps subject to some choice criterion, are made OA-available through posting to a Web site or repository. The House of Commons Select Committee Report adopted a qualified position on the former, 'open journal' OA, but was strongly in favour of the latter, 'open posting' OA. There is some reluctance, on the part of Government, to disturb existing publishing models, but powerful bodies, like RCUK and the Wellcome Trust, are supporting open posting; more importantly, ICT developments will force changing business models for publication as a whole. Open posting OA along with conventional journal publication is growing, but ad hoc, partly through self-archiving and partly through, e.g., subject repositories; and pressures not merely for OA posting but for full OA publishing are likely to grow. OA in general requires only that material is available somewhere, i.e. is e-archived, but this covers both informal personal Web sites and more formal, authorised (and hopefully longer-term) repositories.

Subscription charges for HSS journals are generally much lower than for STM (SJPS04), so perceptions about whether subscriptions are affordable or reasonable are rather different in HSS, and a good many HSS journal owners, like learned societies, may feel the benefits of commercial publishing on their behalf far outweigh the down sides.

But it is also the case that library purchasing in HSS is under pressure, so HSS researchers have a growing motivation to examine forms of publication that may increase access. Again, while hard copy publication is regarded, in many HSS quarters, as the only legitimate form of publication, the de facto spread of e-activity generally is beginning to infect HSS researchers and leading them to make papers available, as ad hoc pre- or post-prints, on Web sites. The general continuing increase in the pace of research is also making the long time periods common in traditional hard copy publication less acceptable. As against this, the continued emphasis in HSS on monograph publication makes the OA model less obviously pertinent, even if in STM and some areas of HSS electronic publication and distribution of book-length items like dissertations is common. But the most important influence on HSS views on OA is probably the fact that proportionally less research is grant-supported, so both funding agency interests in OA publishing and the possibilities for institutional flexibility in resource financing are less obvious. It is certainly not clear that, in the absence of grant funding, and without alternative support mechanisms, HSS researchers will take a positive attitude to authors paying for publication.

However it is also not clear that the traditional publication models on which HSS researchers as resource producers currently rely will continue to remain viable, particularly when HSS researcher interests as resource consumers are already subverting the producer view and external changes are further working in favour of consumers who want access to resources.

We believe that pressures towards open access are likely to increase in STM, and that this will have knock-on effects for HSS, in addition to internal pressures in HSS and SS especially. So we think it important that HSS takes an interest in what is happening and, in particular, seeks to ensure that any moves to OA for STM do not damage HSS, and may rather be positively exploited.

**REC 20: that responsible parties in HSS express public support for the principle of wide and ready access to research outputs and other research resources.**
(AHRC/ESRC/Funding councils/British Academy)

**REC 21: that relevant bodies monitor open access developments to ensure that these do not have a negative impact on HSS, and in particular on individual HSS producers through naive applications of 'author pays' policies.**
(AHRC/ESRC/RLN/British Academy)

## 5.5 Technical support

(See 3 Q11)

One operational issue in e-resource access that can be important for researchers is the level and form of ICT technical support they get. While HSS researchers are increasingly e-savvy, they may not be able to get the ICT advice they need for effective working. This technical support applies to both hardware and software, and in relation to software refers both to general software and to resource-access and use software where, for example, guidance may be needed on e-catalogue search or e-data manipulation. The requirement for competent technical advice also applies to researchers as resource providers, for example as database builders.

Technical advice and support on hardware and general software, though uneven, appears to be generally adequate, though generally less comprehensive and immediate than in STM departments, and HSS researchers are more likely to have to rely on central university computing services than on departmental, and hence subject-informed, ICT officers. Thus researchers may in practice proceed more on a go-it-alone basis, or seek advice too late for maximum benefit, than is really desirable. HSS researchers may therefore fail to make best use of available e-resources and tools, and also fail to adopt the most useful resource construction models. In subject areas, especially SS ones, where ICT has long been exploited, there are relatively few difficulties, but AH researchers may fail to realise the need for technically informed advice. There are particular problems in that while HSS researchers may reasonably assume they know how to use the traditional non-e tools of their trade, it does not follow they know equally well how to use the new e-tools that are becoming available, e.g. for searching. At the same time other relevant staff, like librarians, need to be well-informed about and proactive in promoting subject-specific e-resources.

**REC 22: that ICT and software support for HSS researchers be up to date.**
(universities)

**REC 23: that ICT technical support for HSS researchers be ensured at faculty or school level, or be faculty tailored, to promote subject-area as well as general guidance and support.**
(universities)

**REC 24: that university library staff are trained and take a proactive approach to ensuring that researchers are aware of, and know how to use, e-resources, especially secondary ones.**
(university libraries)

**REC 25: that researchers, and departments, develop mutual self-help habits to improve knowledge and understanding of e-resources in their subject areas.**
(universities/researchers)

## 5.6 Repositories

(See 2.2; 4.15)

The issues here are those on the technical side to do with design, curation and sustainability, and those on the organisation side to do with scope, responsibility and funding.

These two aspects are hard to separate, but there is currently a marked lack of clear thinking about repositories and an urgent need for a more thorough analysis of their requirements and implications, especially for long-term service. Even though changing ICT makes it very hard to predict long-term functional and cost conditions, the current Topsy-like situation is likely to be inefficient and expensive. Much material will be deposited without a proper understanding of migration, recovery and future service needs and possibilities; and while realism suggests that we would not in fact, contrary to many ICT buffs, want everything kept, this is not a good reason for general carelessness about future losses. At the same time, valuable material will not be deposited because, for a good many researchers, there is no obvious way of their ensuring that valuable materials are preserved beyond the random life of their own personal Web sites. The general problem is, moreover, one that affects major resource providers, like journal publishers, as much as individual researchers or small groups.

The present situation is an ad hoc market mix of public and private with sites run by large publishers; by institutions like the National Archives carrying forward their normal operations in an e-form; by funded provider agencies like the UKDA; and sites run by the many and varied other parties with particular interests, of the kind represented in PORTAL. But there is little in the way of common repository models or experience with the longer-term protocols, and little coherent coverage of UK-originating material. In general there is great variety in the nature of the materials included in such repositories as there are, ranging from raw survey data, say, to processed numerical data, technical papers, to formal publications, and extending well beyond usual notions of 'e-prints'.

Currently, for HSS in the UK, there are the UKDA and its dependent sites, motivated by grant-funded data deposit, which illustrate subject area repositories; and a number of prototype institutional repositories, reflecting the view that individual researchers, regardless of their subject areas, deserve support in sustaining their research materials.

### Technical issues

(See 2.2; 4.7; 4.15)

The technical issues concern choices of document and data format, along with those for appropriate management and preservation software, and for operational protocols. This is an area where go-it-alone is not sensible: there is much to be learnt from collaborative investigation, prototyping, etc., and by taking advantage both of STM activities, where repository site work is more advanced, and of international examples and studies, as well as of, e.g., document handling in general. It is particularly important to be clear about the requirements for user access, to any item, on demand and on a day to day basis: the idea that repositories should be only a backup store of last resort, as the name and use for archives might imply, is completely inadequate. Repositories need to be designed for the long term, and managed as operational services.

The implications of repositories holding a wide range of resource types, with different ownership and publication status and subject to different intellectual property regimes, also need further investigation. While dealing with these matters is in part an organisational one, depending on the type of material deemed acceptable, there are important technical issues of database management, for example covering rights propagation, that require technical understanding and sound software engineering.

**REC 26: that further collaborative work on repository design, software engineering and operational service requirements, plus practical prototyping and long-term condition analyses, should be encouraged and funded as a matter of urgency.**

(RCUK/JISC/RLN)

The real requirements for access tools for effective repository use, and especially metadata requirements, are far from clear, other than in specialised data and subject areas. Thus while repositories can only gain from the current developments in document and data handling languages, notably XML and its successors, this does not imply that repositories will be of no use without full blown ontologies and all the more ambitious facilities envisaged for the Semantic Web. At the same time, it is not obvious that much of the conventional bibliographic apparatus is necessary as a finding tool, though it may be highly desirable for other reasons, and there is a clear need, under many repository models, for the highest standards of object labelling, versioning and so forth.

It is particularly important that repositories should be clearly seen as e-resources in their own right, not as non-e resources that, though digitised, simply carry their conventional library packaging with them. Repositories should be clearly constructed, right from the start, for modern forms of e-access including Web visibility, full text search and the like. Less requirement for ambitious metadata and similar curation should also make repositories less expensive.

The presumption that the correct interface with the external world, notably through search engines, can be painlessly achieved by automatic metadata harvesting, e.g. by using the Open Archives Initiative Framework, needs careful examination. If service providers are supposed to add value by building on the metadata, they have to get appropriate metadata; but equally a supply-side view of what the providers ought to find useful may not be a sensible approach to delivering metadata that is effective from the users' demand side. These technical issues interact with 'political' and organisational ones. Thus the relative merits, for convenient deposit of and effective access to material, of national, institutional or subject-based repositories need urgent study.

It is vital to make a clear distinction between repository proposals that are motivated by bureaucratic or political considerations, e.g. ones for institutional repositories with controlled metadata and national harvesting, and proposals that are motivated by researchers' interests in resource access where, for instance, metadata may be better lightweight: a single model may not meet radically different goals.

**REC 27: that collaborative work on repository design for e-access, within the wider framework of World Wide Web and similar general-purpose e-operations, should be actively promoted.**

(JISC/RLN)

**REC 28: that collaborative work on metadata utility for open Web service access to resources be actively undertaken.**

(UKOLN/JISC)

### Organisational issues

(See 2.2; 2.3; 4.10; 4.14; 4.15)

There are problems for both resource providers and resource consumers with all the repository organisation models.

Thus institutions like universities may have to support substantial repository efforts which are not seen as particularly valuable to research communities. Institutions like major libraries may (find themselves obliged to) develop repository roles for their own holdings, but these are subject to the institutions' funding constraints and may not provide good subject coverage for individual research fields. Subject-based organisations, e.g. major professional societies, may support repositories of community value but their long-term funding may be vulnerable. Funding agency repositories may be valuable to consumers but not mandated or supported for subject coverage across multiple resource types. Publisher repositories may be helpful to consumers, but for some resource types only, and are vulnerable to commercial fortune including termination. Global repositories as illustrated by the Internet Archive suffer from lack of content discrimination. These problems apply across all resource types, though they may currently affect some types more than others e.g. published versus unpublished material. There are particular complexities where many resource types with different formal status and intellectual property status are involved.

These problems imply a need for detailed investigations and cost-benefit analyses of the alternative repository models, in the particular context where UK provision, or encouragement, is important or essential whether to preserve the resources created by UK researchers or to make these and other resources available to UK researchers.

**REC 29: that study be made of the costs and benefits of institutional (university etc.) repositories to UK HSS researchers, both for research outputs generally and for e-prints.**
(UUK/RCUK)

**REC 30: that study be made of models and mechanisms for subject repositories, as a basis for future decisions to promote UK-based repositories.**
(RCUK/JISC)

**REC 31: that the research funding bodies, especially AHRC and ESRC, continue to investigate best models for repository support for the resources their grants help to create, and possible roles for UKDA and AHDS.**
(AHRC/ESRC)

**REC 32: that JISC be remitted to develop alternative models for very long-term, publicly funded repositories, and to explore design examples in feasibility studies.**
(JISC/BL)

## 5.7 Intellectual property

(See 2.2; 2.3; 4.10; 4.11)

The primary issue here is the impact of IP on access to resources. But since without IP rights in them, creators are unlikely to produce resources and make them available, the issues for creators have also to be considered. The issues for both creators and users are taking sharp new forms under the pressure of ICT, and also have ramifications for resource providers and mediators more generally. Both formal protocols and informal practice are rapidly changing, with effects on legitimate forms of use and on charges for these; the wide-ranging effects for e-resources in particular make it impossible to operate primarily within a national environment.

The situation is extremely complex, as the Royal Society noted, and the Academy is undertaking a review of IP in the HSS context. Our comments here are therefore focused on IP issues in specific relation to access to e-resources.

In the traditional non-e 'physical' model, IP began with the author but was normally assigned to the publisher, whether a body like a learned society or a purely commercial operation. The publisher took the risks of publication by bearing the costs, and lost or profited. Book authors might make money from advances or royalties, or at least expected to gain status; journal article authors normally could only get status. Other forms of resource, e.g. maps or photographs, might be assimilated to one or other of these variants on the basic model. Publishers recovered their costs by sales – in the academic world mainly to libraries. Once published, materials were essentially in the public domain and free to users, with IP rights exercised primarily by locational ties for reading or borrowing, copyright constraints on copying forms and quantities, as well as photocopying charges. Having shaken down over a long period, the relationship between the parties involved had settled to a generally win-win one, with creators and producers, having providers like libraries as their agents, IP protected, and with users not unreasonably constrained in access by others' rights. The same general regime applied internationally.

ICT has disturbed this model, both by shifting effort and cost balances within the publication process even for resources eventually emerging in non-e form and, much more significantly, for e-resources. It has done this both through its direct impact on operations on and with e-resources, and also through its indirect impact on peoples' attitudes: the growth of the Web has generated a large presumption that what is out there is free not merely to get but to do things with, where one can do so many things so easily that anything not perceived as an obvious infringement of the IP holder's rights tends to get done. Thus the way in which existing items can be propagated, and new materials constructed using bits of old ones and propagated, is moving the boundary between the researcher's private workspace and the formal publication step, and putting a new informal but public space between the two. Reusing other's material is no different from what researchers have done with others' non-e materials in their own personal or very local workspace. But ICT is encouraging the growth of informal public space in a way that may, strictly, be breaching formal IPR.

Attempting to enforce IPR that were originally seen as reasonable for the non-e case in the e-environment is making them appear unreasonable so they are, where practically feasible, generally being ignored. Treating a screen display as a copy is perceived as unjustifiable and unacceptable. This reflects the fact that applying existing formal IP rights is shifting the balance between resource owner and resource user, in favour of the owner. It has led to the use of explicit licensing schemes, which for the researcher means: 'you can only do what we will let you do', replacing the older library scheme which meant: 'once you're in the library you can do anything we don't say you mustn't do'. The shift towards owners is, in particular, increasing publisher powers, and owner powers are being further enhanced by the increasingly restrictive character of copyright legislation itself, for example the length of time to which copyright applies.

The investments that publishers are making, for journals in particular, in e-resources, have important benefits for researchers. But it is far from clear, especially when they also hold digitised back runs, that researchers will have the formal ability to access resources as much as they do now, in spite of the access promise of ICT. In the same way, the rights that can be claimed for digital images can hold researchers to financial ransom.

It is not necessary, to ensure researchers' access to e-resources, to advocate a total free-for-all; nor are 'creative commons' necessarily the right solution to present problems. However the current commercial push, for instance in the US, towards greater controls over e-material, even if motivated by interests in quite other materials than those likely to be used by HSS

researchers, is likely to have damaging effects on researchers' access to e-resources. It is therefore important that those with an interest in encouraging research, and especially in funding it, should adopt a more active policy on ensuring access to e-resources than has hitherto been the case. It is crucial, in particular, to combat the definition of any mechanical copying of any kind as an infringement of IP, and the perpetuation of copyright periods through mechanical copying.

**REC 33: that HSS researchers should be ensured the same real rights in access to e-materials as to non-e materials.**

(Licensing societies/Publisher bodies/JISC)

**REC 34: that fair dealing, and fair use, for personal study and non-commercial research should apply equally across all resource types, including audio-visual content.**

(Licensing societies/Publisher bodies)

**REC 35: that resource creators should not give away exclusive rights to publishers, but should be encouraged to retain basic rights including those of e-posting and other use of their own materials.**

(RCUK/UUK/British Academy/researchers)

**REC 36: that not-for-profit partnership approaches should be encouraged for the digitisation of older materials, and to the creation of secondary e-resources.**

(RLN/British Academy /Funding bodies)

**REC 37: that funding and other national bodies should collaborate to establish sound intellectual property models for e-resources.**

(Funding councils/RCUK/JISC/British Academy)

## 5.8 Organisation and funding

(See 2.2-2.5; 3 Q2, 7, 12; 4.10; 4.13-4.16)

Overall, UK bodies, whether at the national or local level, are doing much to establish e-resources and enable access to them. A good deal of this activity is a natural, but also ad hoc, response to the march of ICT. For example, digital databases are created on research projects, copyright repositories are receiving digital deposits, libraries and other agencies are negotiating deals for e-journals, universities and their departments are involved in sustaining resources created by individual researchers on their Web sites. However, while most of this activity has been reactive, there has also been proaction, whether by research councils requiring e-resource preservation, libraries seeking secondary e-resources, or government agencies managing resource sites or offering guidance on resource formation and maintenance.

Our Review shows a mass of complicated relationships among the various institutional providers, especially at the national level. The formal and financial connections embed numerous resource bodies with overlapping remits, resource interests, and activities. This maze-like confusion partly reflects the fact that much e-resource creation and access provision has grown bottom-up from particular initiatives, and also the fact that different agencies have different subject orientations, provider constituencies and user communities, as well as different funding responsibilities and powers. However this multiplicity of entrenched provider constituencies is not in the best interests either of best resource provision or of resource users.

Universities, as the researchers' collective main home, though centrally funded are primarily focused on their own goals and conditions. Their priority is the institution's interests and needs, not the greater national good. Thus with respect to research resources, the relationship

between both their library operations and their grant-funded research and the wider needs of the research community at large for resource provision and preservation are relatively weak. So while the resources available to an institution's researchers vary widely, there is great resistance to top-slicing approaches to develop national resources and a definite preference for institution-limited optimisation. The way in which, for example, post-1992 universities have developed local collaborations illustrates how effective such strategies can be.

Among the central bodies, major national providers like the British Library, National Archives and MLA are outside the HE sector and also have their own various skills and priorities. The Library is a major collector of non-e materials but is also keen to establish a good and effective digital deposit apparatus; but it can only influence others through the weight of its holdings and by leadership example. It has encouraged the recently-funded RLN as a body, but it is not yet clear what the RLN's operational effects, given its relatively modest finances, will be, and especially how what is seen as an elite-oriented organisation will support the greater research community's good. TNA's influence, though vital for the class of materials it deals with, is limited by the types of resource it deals with in relation to the wider range of researcher interest.

The MLA's concern with wider access, for everyone, to all knowledge resources, and with lifelong learning, means that its influence on resources for the 'professional' researcher are more indirect than direct. The research councils collectively, and the AHRC and ESRC specifically, are very much concerned with the reuse of resources they have funded as an integral element in new research, and thus support major data archives. But these, though important, are only part of the resources researchers need, and are themselves essentially heterogeneous aggregations. JISC plays an important technical role for e-resource development and conservation, but has multiple disjoint activities and major commitments to support for teaching and learning rather than research. Again, while JISC and major libraries or library consortia have an important role in facilitating access to commercially-produced resources, notably journals, much of this is ad hoc in the sense of dealing with particular resource sources and, as noted, does not ensure that the resources themselves are available to all UK researchers.

Unfortunately, while there is no need in principle for resource users to know about, or negotiate, the formal and funding organogram maze, the complicated aggregation of resource providers with its mixture of present and absent connections is in practice an impediment to users' access to resources. There are too many resource sites, and resources, each built in some particular e-style with particular e-content materials, without any regard for the amenity and convenience of the e-street as a whole. Steps are being taken, e.g. the Common Information Environment initiative (CIEWWW), to bring more coordination and collaboration among providers, and some parts of the town are rather more coherent and better signposted than others. However it is too often the case that there is a surplus of resource shops for one type of good in one area and a marked lack of any shops for another type elsewhere.

In relation to funding for e-resources, the overall model is that any shift from non-e to e-versions of resources, or development of new e-resources is done primarily within existing funding frameworks and fund supply. This largely implies a continuation of the current distributed mix of responsibilities and interests, with some place and scope for accommodating particular, specifically-funded initiatives for infrastructure like the Grid or particular digitisation projects. Much of the centralised funding relevant to e-resources of interest to researchers is channelled through JISC, but there are two issues here: whether HSS gets enough attention compared with STM, and also whether research interests get enough attention compared with teaching and learning. Moreover the distribution of JISC funding over a wide range of activities and sub-bodies means that the available money is unlikely to be being used to best advantage. In particular, the project-based culture, sometimes morphing projects into services, but often failing to follow up, is not necessarily the best way to ensure that users' needs are properly served.

There are particular problems in the tension between the private and public sectors in e-provision and access. The private sector is willing to make substantial investments, for example because a major journal publisher can expect to reach a world-wide clientele. But while the public funding sector may gain in the short term, the costs to public sector research in the longer run may be higher than they would have been with more public investment up front. The STM community is becoming more aware of these issues, and seeking suitable responses other than purely reactive ones. The HSS community needs to wake up to them, and address the problem of what responses may be feasible in the context of the kinds of research resources and needs they have.

As a whole, the major problem in organisation and funding is that e-resource provision is fragmented, and too often supply rather than demand led. There are presumptions about the way that, and ease with which, the individual researcher can be expected to navigate among multiple resource providers, and search for and identify resources, that are not obviously grounded in researchers' actual needs and skills. The UK's multiple, distributed resource provision encourages poor cost-benefit ratios for providers and users alike. However we do not believe that heavy-duty, top-level government-instituted attempts at unification, even if feasible, are the only appropriate forward strategy, though some government-led underpinning infrastructure provision is essential. What is required is more operational collaboration among resource holders grounded in direct interaction with researchers themselves. Active researchers should be included on relevant decision-making bodies, and also consulted directly.

It is not the Academy's role to propose major changes in organisational structure or funding regimes. However we make the following recommendations on these matters, for e-resource provision and access.

**REC 38: that key players collaborate more closely and strategically than at present in the provision and sustenance of e-resources, in planning, organisation and funding.**
(Funding bodies/JISC/RLN/BL/CURL)

**REC 39: that the relevant funding bodies, notably AHRC and ESRC, develop policies for longer-term e-resource requirements and reuse that enhance the quality and value of research community resources.**
(Funding bodies)

**REC 40: that joint bodies with a provision role, notably JISC and RLN, develop well-grounded researcher-oriented strategies for e-resource provision and management, that provide sound long-term models and not just viable responses to the current situation in public/private provision.**
(JISC/RLN)

**REC 41: that HSS researchers be directly involved in the development of e-resource provision and management strategy by JISC and other meta-organisations like RLN.**
(JISC/RLN)

**REC 42: that the funding bodies collectively promote the investigation and characterisation of operationally sound and financially viable models for long-term e-resource preservation, curation, and access services.**
(Funding bodies)

**REC 43: that funding should pay particular attention to secondary e-resources and support for resource discovery, emphasising wide-coverage and light-weight materials and methods**

**that enable researchers to gain some access to what there is, but are also well-designed to mesh with current ICT operational structures and tools like the World Wide Web and Web search engines.**

(Funding bodies)

**REC 44: that the British Academy monitor developments in the provision of research resources, and especially e-resources, and encourage cooperation within the UK and between UK and other parties to promote the timely provision of e-resources for HSS research.**

(British Academy)

**REC 45: that the British Academy take the lead in encouraging HSS learned societies to develop policies on e-resources.**

(British Academy)

## 5.9 Conclusion: Overall combined recommendations

The single global message of this Review to resource providers and access mediators is:

*Learn from the way researchers are actually using e-resources already and get joined up.*

Taking the various Specific recommendations under the preceding separate issue headings together, this leads to the following Overall recommendations. The more important among these are starred.

* **OVERALL REC A: that relevant institutions and bodies adopt a coordinated and coherent strategic approach to e-resource provision and access, based on research community needs.**
* **OVERALL REC B: that e-resource conversion by resource holders pay particular attention to secondary before primary e-provision.**

**OVERALL REC C: that all those providing e-resources address means and mechanisms for access from general information discovery systems such as Web engines.**

* **OVERALL REC D: that the national institutions, funding bodies and library representatives collectively address the development of licensing and fair use protocols for e-resources that balance the claims of providers and users.**
* **OVERALL REC E: that the national institutions and funding bodies conduct an in-depth analysis of the requirements and options for long-term e-resource curation, preservation and use.**

**OVERALL REC F that HE and other research institutions ensure that HSS researchers have sufficient access to appropriately-trained technical support staff.**

**OVERALL REC G: that HSS researchers actively seek guidance on access to, and provision of, e-resources.**

**OVERALL REC H: that HSS researchers actively promote user community interests to both e-resource funders and providers.**

## 5.10 Organisations and groups to which recommendations are addressed

Note that individual organisations or groups may also be subsumed under higher level headings, e.g. 'Funding councils' under 'Funding bodies', or 'Museums' under 'Collection-owning institutions'.

| | *Recommendations* |
|---|---|
| AHDS – Arts and Humanities Data Service | 6 |
| AHRC – Arts and Humanities Research Council | 5, 20, 21, 31 |
| Archives holders | 14 |
| British Academy | 20, 21, 35, 36, 37, 44, 45 |
| BL – British Library | 1, 2, 3, 12, 16, 18, 32, 38 |
| Collection-owning institutions<br>i.e. libraries, archives, museums, learned societies | 7 |
| CURL – Consortium of University Research Libraries | 1, 2, 11, 12, 13, 38 |
| ESRC – Economic and Social Research Council | 5, 20, 21, 31 |
| Funding bodies<br>i.e. Funding councils; Research councils; Government departments; Lottery agencies | 7, 10, 19, 36, 38, 39, 42, 43 |
| Funding councils<br>i.e. HEFCE; SHEFCE; HEFCW; DELNI | 20, 37 |
| JISC – Joint Information Systems Committee | 5, 6, 8, 13, 16, 17, 19, 26, 27, 28, 30, 32, 33, 37, 38, 40, 41 |
| Joint Committee on Legal Deposit | 18 |
| Learned societies | 2 |
| Licensing societies<br>i.e. Authors' Licensing and Collecting Society; Publishers Licensing Society | 17, 33, 34 |
| Major libraries<br>i.e. research libraries; national libraries | 9, 10, 11 |
| Museums | 15 |
| MLA – Museums, Libraries and Archives Council | 15, 16 |
| National Council on Archives | 4 |
| Publisher bodies<br>i.e. Publishers Association; Committee on Academic and Professional Publishing; Association of Learned and Professional Society Publishers | 33, 34 |
| Publishers | 2 |
| RCUK – Research Councils UK | 26, 29, 30, 35, 37 |
| RLN – Research Libraries Network | 1, 2, 3, 8, 11, 13, 16, 21, 26, 27, 36, 38, 40, 41 |
| Researchers | 25, 35 |
| TNA – The National Archives | 4, 14, 16 |
| UKDA – UK Data Archive | 6 |
| UKOLN | 28 |
| Universities | 9, 22, 23, 25 |
| UUK – Universities UK | 29, 35 |
| University libraries | 24 |

# Appendix 1

## Working Group membership

The Review's work has been carried out by a Working Group of 10 members, appointed by the British Academy. Under the chairmanship of Professor Karen Spärck Jones the Working Group met three times between May and December 2004 to oversee the direction of the Review and to consider its findings.

**Chair**

**Professor Karen Spärck Jones FBA**
Emeritus Professor of Computers and Information
University of Cambridge

**The other members of the Working Group**

**Professor Robert Bennett FBA**
Professor of Geography
University of Cambridge
Chair, British Academy Research Committee

**Dr Clive Field**
Director of Scholarship and Collections
The British Library

**Mr Richard Fisher**
Executive Director, Humanities and Social Sciences Publishing
Cambridge University Press

**Professor Roger Kain FBA**
Deputy Vice-Chancellor and Montefiore Professor of Geography
University of Exeter

**Professor David Rhind CBE, FRS, FBA**
Vice-Chancellor
City University

**Professor David Robey**
Professor of Italian
University of Reading
Director of the AHRC ICT Arts and Humanities Research Programme

**Professor Seamus Ross**
Professor of Humanities Informatics and Digital Curation
University of Glasgow

**Professor Kevin Schürer**
Director of ESDS and the UK Data Archive
University of Essex

**Mr David Worlock**
Chairman
Electronic Publishing Services Ltd, London

**Secretary**
Mr Geoff Smith
Formerly British Library

**British Academy**

**Mr P W H Brown**
Secretary

**Ms Abigail Cooke**
Website Content Manager

# Appendix 2

## Terms of reference

### British Academy Policy Review: Research and Information e-Resources

Information technology is changing the nature of resources for research. Old resources are taking new forms, e.g. electronic journals, and new resources are appearing, e.g. web sites. These electronic, or e-resources can be grouped under the headings of electronic publishing; computer databases and archives; and digital libraries.

The aim of the policy study is to identify, for the Humanities and Social Sciences (HSS),

1) what HSS needs for e-resources are;
2) whether these needs are specific to HSS, as opposed to researchers generally, in the nature of the resources or forms of access to them;
3) how, if these needs are special, they can be factored into current national policies, strategies, and practices.

Most of the current e-resource action is driven by the scientific and technical community's pressures, interest and needs, notably publisher consolidation and reaction to it (e.g. ScienceDirect and Public Library of Science), and large-scale collaborative data gathering (e.g. genome database projects), with the advantage in many cases of substantial funding and government or other institutional support (e.g. Biobank). These developments are having knock-on effects on other important matters, such as intellectual property ownership.

At the same time, library-based initiatives are in the UK, notably the Research Support Libraries Group's proposal for a Research Libraries Network, and the SUNCAT project for a national serials catalogue, that are important for HSS researchers. Thus one aspect of the study is to relate HSS e-resource needs to current and likely changes in the UK's library-based resource provision.

The specific terms of reference for the study are therefore:

1. To assess the development of existing initiatives in e-resource provision, to determine how HSS will be affected.
2. To assess the impact of the way the publications market for e-resources operates on HSS.
3. To assess the impact of other institutional or structural factors (e.g. research council protocols) on e-resources for HSS.
4. To assess the implications of the very wide dispersion of HSS research resources, to determine whether this has disadvantages for the provision of HSS e-resources.
5. To assess whether some subjects (e.g. modern languages), or institutions (e.g. learned societies) are or would be disadvantaged in the provision of e-resources.
6. To establish whether the British Academy could have any particular role in relation to gaps or deficiencies in e-resource supply for HSS.

# Appendix 3

## List of organisations and persons consulted

| 1. Organisations | Representatives |
|---|---|
| *Collection-owning bodies* | |
| British Library | Lynne Brindley; Clive Field; Richard Boulderstone |
| The National Archives | Sarah Tyacke; Alison Webster; David Ryan; Liz Hallam Smith; Vanessa Carr |
| Cambridge University Library | Peter Fox; Patricia Killiard |
| National Library of Wales | Andrew Green |
| National Museums of Scotland | Wendy Turner; Brian Martin |
| *Learned Societies* | |
| Philological Society | Paul Rowlett |
| Royal Anthropological Institute | David Zeitlyn |
| Economic History Society | Robert Allen |
| *Research Councils* | |
| Arts and Humanities Research Council | Michael Jubb; David Robey |
| Economic and Social Research Council | Jeremy Neathey; Astrid Wissenburg |
| *Data Archives* | |
| UK Data Archive/ESDS | Kevin Schürer |
| Arts and Humanities Data Service | Sheila Anderson |
| *National agencies* | |
| Joint Information Systems Committee | Malcolm Read; Loraine Estelle |
| Museums, Libraries and Archives Council | Chris Batt; David Dawson |
| *Individual experts* | |
| Coalition for Networked Information | Clifford Lynch |

**2. Researchers**

See Section 3 of the report for details of the researcher survey conducted.

**3. University libraries**

The heads of library services of the following universities were written to. Responses were received from those marked with *

*Older universities*

Aberdeen; *Birmingham; *City; East Anglia; *Exeter; Lancaster; Leeds; *Leicester; Liverpool; *Southampton; Wales Bangor

*Post-1992 universities*

Anglia Polytechnic; Brighton; *Central Lancashire; Glamorgan; Nottingham Trent; Plymouth; Robert Gordon; Sheffield Hallam; Sunderland; *Westminster; *Wolverhampton

# Appendix 4

## Survey letter and questionnaire

### Covering letter

**British Academy: Policy Study on Research e-Resources for the Humanities and Social Sciences**

Dear Researcher

I am writing to you as Chair of the Policy Study's Working Group. The British Academy is investigating the provision of, and access to, research e-resources, and we expect our Study's findings and recommendations to be significant inputs on this increasingly important matter to such bodies as HEFCE, AHRC and ESRC.

Further information on the Study is available at:
http://www.britac.ac.uk/news/release.asp?NewsID=142

We are particularly concerned that researchers in the Humanities and Social Sciences should be able to take full advantage of electronic developments, and should not be placed at a disadvantage because of strategies and practices specifically oriented towards research in science and technology.

We are following this up with British Academy Fellows and award-holders, among others. We are therefore approaching you as an HSS researcher, for your help in providing us with first-hand information about your experience with research e-resources, both by comparison with non e-resources with which you are familiar, and in their own right.

We would be grateful if you would complete our survey questionnaire, electronically or on paper as you prefer, and return it as soon as possible (at the latest by 10th September), as indicated below.

Thank you in advance for your help.

Karen Sparck Jones

Emeritus Professor of Computers and Information
University of Cambridge

## Survey questionnaire

Please provide specific details as requested. We also welcome additional comments.

**Use of non–electronic resources.**

1.1 What are the three most important types of non-electronic primary research resource you use (eg journals, data archives, manuscripts, maps …)?

1.2 Can you access these easily? (Yes/No).

1.3 If not, why not? Please comment.

**Availability of resources in electronic form.**

2.1 Are any of these three important resource types also available, to a significant extent, in electronic form? Please answer Yes or No for each of the resource types.

2.2 If so, can you access them easily? Please answer Yes or No for each of the resource types.

2.3 If not, why not? Please comment.

3 Where the same resource is available in both non-electronic and electronic form which do you prefer to use, and why?

4 Are you using e-resources that are wholly new, or sufficiently different from their non e-versions to offer quite novel research possibilities (eg digital maps)? If yes, please list the types of resource.

5 What other primary resources (both electronic and non-electronic) do you sometimes use? Please list.

**Tools, software, research mechanisms.**

6 Do you make significant use of data processing programs or tools (eg text corpus analysers, statistical modelling tools)? If yes, please list examples.

7.1 What traditional resource discovery tools, i.e. secondary resources, do you rely on most (eg catalogues, bibliographies, abstracting and indexing services, colleagues)?

7.2 Are there e-tools for discovery that are particularly important to you? (eg portals, search engines, alerting services). If yes, please list.

7.3 Do you have any difficulties in accessing or using these e-tools? If yes, please describe.

8 Has the way you do your research already changed, or is it changing, through new e-materials (eg digitised image files), programs (eg machine translation systems), e-discovery tools (eg Web portals), or basic e-capabilities (eg email, PC 'office' software)? If yes, please describe.

**Resource creation**

9 Are you yourself an e-resource contributor or creator through the provision of data (eg transcribed records), or of literature (eg publication of your papers in e-journals or on your department's Web site)? If yes, please describe.

10 Do you have your own Web pages? (Yes/No).

**Technical capabilities and support**

11.1 How do you normally work electronically (eg own laptop, Computing Service user machine, machines in university library)? Please indicate type of equipment, ownership, location.

11.2 Are you able to get technical IT help (eg on file transfers, document formats, hardware problems)? If yes, please indicate from whom (eg institution's support service, departmental colleagues).

11.3 Are you able to get expert guidance on availability and use of e-resources? (If yes, please describe from whom (eg institution's library service, departmental colleagues).

**Gaps in provision**

12 What resources or tools would you most like to have electronically that you don't have already? Please list.

13 Other comments. Please give any other comments you think would be helpful to the study.

**Personal information – in confidence**

14.1 Your subject area (eg Medieval history, Greek literature, Social psychology, Econometrics).

14.2 Your position (eg Professor, postdoctoral fellow, lecturer, research assistant).

14.3 Your institution. Please give institution name.

14.4 If you are a British Academy Fellow or award holder, please indicate whether you are an FBA, post doctoral research fellow (PDRF) or small research grant holder (SRGH)

Thank you for your help.

# Appendix 5

## List of background documents and sources of information

We have provided short citation forms for references of two types, for ordinary documents and for Web sites, which we hope are self-explanatory. Note that all Web sites used for citation forms were visited in early 2005.


**AIAWWW** AHRC ICT in Arts and Humanities Research Programme
http://www.ahrbict.rdg.ac.uk/

**ARXWWW** arXiv.org e-Print archive
http://arxiv.org/

**BAPWWW** The British Academy. PORTAL.
http://www.britac.ac.uk/portal/

**BCCP02** Weiss, Peter. Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts. Summary Report. US Department of Commerce. National Weather Service. 2002
http://www.weather.gov/sp/Borders_report.pdf

**CCHWWW** Commission on Cyberinfrastructure for the Humanities & Social Sciences
http://www.acls.org/cyberinfrastructure/cyber.htm

**CEDWWW** CEDARS: CURL Exemplars for Digital Archives project
http://curl.bham.ac.uk/projects/cedars.htm

**CIEWWW** Common Information Environment initiative
http://www.common-info.org.uk/

**CRRWWW** Computing Research Repository (CoRR)
http://xxx.lanl.gov/archive/cs/intro.html

**DCCWWW** Digital Curation Centre
http://www.dcc.ac.uk/

**DCMWWW** Dublin Core Metadata Initiative
http://dublincore.org/

**DIGWWW** DigiCULT: Technology Challenges for Digital Culture
http://www.digicult.info/pages/index.php

**DMAM04** Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education. A joint report by The Electronic Publishing Innovation Centre (EPIC), in partnership with Key Perspectives Limited, 2004.
http://www.jisc.ac.uk/uploaded_documents/E prints_delivery_model.pdf

**ENRWWW** EnrichUK: the gateway to a lottery-funded collection of 150 sites supported by the New Opportunities Fund.
http://www.enrichuk.net/

**ECAWWW** Electronic Cultural Atlas Initiative
http://www.ecai.org/

**ESCWWW** e-Science: the UK e-Science Programme
http://www.rcuk.ac.uk/escience/

**EEUWWW** eEurope 2005 Action Plan
http://europa.eu.int/information_society/eeurope/2005/index_en.htm

**FAIWWW** JISC Focus on Access to Institutional Resources (FAIR) Programme
http://www.jisc.ac.uk/index.cfm?name=programme_fair

**FGGI01** Borgman, C.L. From Gutenberg to the global information infrastructure. MIT Press, 2001

**GALWWW** Gallica: bibiothèque numérique de la Bibliothèque nationale de France
http://gallica.bnf.fr/

**GCOWWW** Google Checks Out Library Books
http://www.google.com/intl/en/press/pressrel/print_library.html

**HCGR04** House of Commons Select Committee on Science and Technology
14th report: Scientific Publications: Free for all? the Government Response 2004
http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/1200/120002.htm

**HCSC04** House of Commons Select Committee on Science and Technology
10th report: Scientific publications: Free for all? 2004.
http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm

**HHHWWW** Humbul Humanities Hub
http://www.humbul.ac.uk/

**IDPWWW** IDP: International Dunhuang Project
http://idp.bl.uk/

**INTWWW** The Internet Archive.
http://www.archive.org/

**IPDM03** Lavoie, B.F. The incentives to preserve digital materials: roles, scenarios and economic decision-making. Office of Research, OCLC Online Computer Library Center, 2003
http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

**IRCA04** ICSU Report of the CSPR Assessment Panel on Scientific Data and Information. International Council for Science. 2004.
http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf

**ISIWWW** ISI Web of Science/Web of Knowledge
http://www.isinet.com/

**ITHS93** Information technology in Humanities scholarship; British achievements, prospects, and barriers. British Library R&D report 6097, 1993.

**JSCWWW** JISC Committee for the Information Environment (JCIE). Scholarly Communications Group
http://www.jisc.ac.uk/index.cfm?name=jcie_scg

**KSOE03** Keeping science open: the effects of intellectual property policy on the conduct of science. The Royal Society, 2003.
http://www.royalsoc.ac.uk/document.asp?id=1374

**LOCWWW** LOCKSS (Lots of Copies Keep Stuff Safe) Program
http://lockss.stanford.edu/

**LUNWW** Lund Principles
http://www.cordis.lu/ist/directorate_e/digicult/lund_principles.htm

**MOAWWW** Making of America
http://www.hti.umich.edu/m/moagrp/index.html

**PERWWW** Perseus Digital Library
http://www.perseus.tufts.edu/

**RSEC03** Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. 2003
http://www.cise.nsf.gov/sci/reports/atkins.pdf

**RSLG02** Research Support Libraries Group: Report, 2002.
http://www.rslg.ac.uk

**RULI02** Researchers' Use of Libraries and other Information Sources: current patterns and future trends. Final Report. (Study commissioned for Research Support Libraries Group 2002).
http://www.rslg.ac.uk/research/libuse/

**SCRWWW** Scran
http://www.scran.ac.uk

**SJPS04** Scholarly Journal Prices: Selected Trends and Comparisons.
Sonya White & Claire Creaser. LISU. October 2004
http://www.lboro.ac.uk/departments/dis/lisu/pages/publications/oup.html

**SOMR04** Survey of Outstanding Material for Retrospective Conversion and Retrospective Cataloguing in CURL Libraries, CURL, 2004.
http://curl.bham.ac.uk/projects/retrocon.htm

**SOSWWW** SOSIG: Social Science Information Gateway
http://www.sosig.ac.uk/

**SSCWWW** Joint CURL/SCONUL Scholarly Communications Group
http://www.sconul.ac.uk/activities/sch_comm/

**TELWWW** The European Library
http://www.europeanlibrary.org/

**UNCWWW** United Nations Common Database
http://unstats.un.org/unsd/cdb/cdb_help/cdb_quick_start.asp

**VLMWWW** Virtual Library museum pages
http://vlmp.museophile.com

**WORWWW** OCLC WorldCat
http://www.oclc.org/worldcat/

**WTPS04** Wellcome Trust position statement in support of open access publishing
http://www.wellcome.ac.uk/doc_WTD002766.html