2. STAKEHOLDERS: RIGHTS AND RESPONSIBILITIES

2.1 Who are the stakeholders?

Many different groups have an interest or involvement in digital information. Any strategy for digital preservation will naturally have to take into account the various needs and perspectives of these groups. The stakeholders include:

- Authors
- Publishers
- Libraries
- Archive centres
- Distributors
- Networked information service providers
- IT suppliers
- Legal depositories
- Consortia
- Universities
- Research funders

Another way of identifying the stakeholders is by their main area of interest in relation to digital material. Using this method Table 1 lists ten stakeholders.

Table 1: Stakeholders' interest and impact on the long term preservation of digital material

Stakeholder	**Interest and impact on the long term preservation of digital material**
Initiators	Collection development. Research libraries collect material that is current, published on current technology. Establish the nature and scale of the threat of irretrievable loss for digital material items.
Regulators	Legal deposit; Public Record Office; Copyright. Legislation to preserve ownership for a limited period of time, to ensure a national collection of material is established and to preserve items that are in the public interest.
Creators	Creation of digital records. Lack of control over format of deposited items leads to unmanageable diversity.
Rights owners	Maintain copyright. Preservation of material may lead owners to demand copyright in perpetuity.
Fund-holders	Manage the funds available for preservation activity according to agreed priorities and service levels.
Providers (1)	Initial diversity of formats at publication complicated by new editions in new formats and on new media.
Readers	Access to material. Readers will demand material in current acceptable format for display and inclusion in new digital material.
Archivists	Conserve the archive, whilst preserving the items, and maintain the integrity of the deposited items.
Providers (2)	Provide new editions, which link into the new intellectual context through re-indexing and re-packaging.
Interferers	Make material inaccessible through technological turbulence or blocking publication.

It is interesting to note that roles and interests may change over the long term. For example, copyright will lapse after a certain length of time: will this make it less attractive to the copyright owner to make a contribution to long term preservation?

A relative newcomer to the scene are the interferers. They may be seen as the antithesis of the regulators — although new regulations may be brought in to counter their activities. Sometimes they may simply be a nuisance, obstructing the course of good preservation practice, taking a narrow perspective on minor issues, or delaying the introduction of new measures. At other levels their effects may be far-reaching. Budget cuts, for example, can seriously damage the value of a collection, by restricting intake and causing holdings to be disposed of. Political instability can destroy centuries of preservation — the intellectual heritage of a culture.

This chapter explores the attitudes of the stakeholders to the preservation of digital data, in terms of both their needs and their responsibilities. It has been compiled from two reports: *Responsibility for digital archiving and long term access to digital data* (Haynes *et al.*, 1997), and *An investigation into the digital preservation needs of universities and research funders* (The Digital Archive, 1998). The research for both reports was carried out through interviews, questionnaires and focus group meetings.

The various stakeholders have been divided into two large groups: those involved in the electronic publication cycle; and those involved in the research community. This somewhat arbitrary division is made to reflect each group's major area of interest. There is, of course, a certain amount of overlap between them. For example, individual members of the research community are often authors and sometimes publishers as well.

2.2 The electronic publication cycle

The model shown in Figure 1 represents the electronic publication cycle and identifies the key players at each stage in the process.

Figure 1. The electronic publication cycle

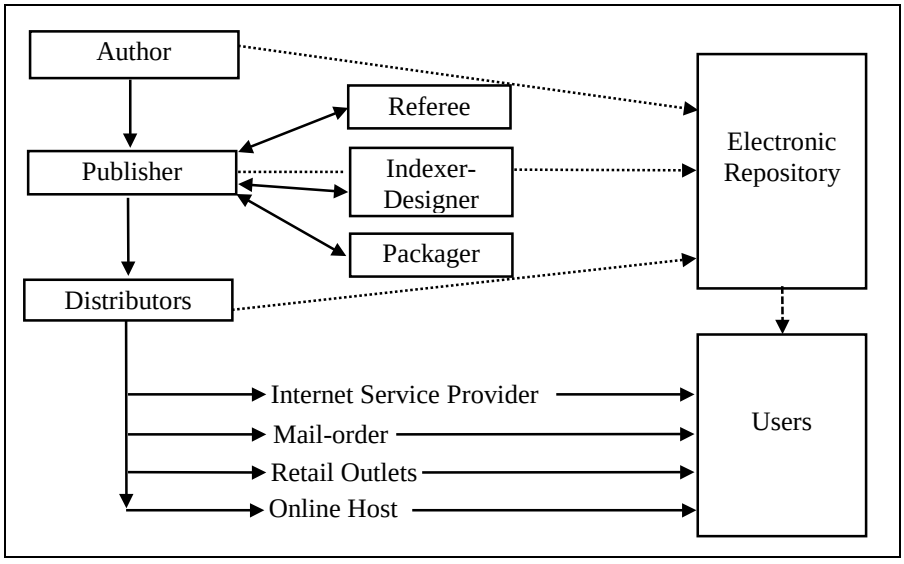

Table 2 shows the priorities and concerns of four of the major groups active in the publication cycle.

Table 2. Stakeholders' priorities and concerns

Rank	Authors, Data Originators	Publishers	Distributors	Repositories
1	Common strategic approach by providers of preservation services (coherence, consistency and interoperability).	Preservation of the functionality of electronic publications.	Criteria for permanent preservation. What is worth archiving?	Standards and common formats.
2	Intellectual property rights.	What constitutes a publication?	Contracts between archivists and information providers. Cooperation between copyright holders to permit voluntary or legal deposit.	Permanence and refreshment of data.
3	Security (protecting against piracy; preservation against catastrophe; preservation of integrity — what is the authoritative text?).	Who should keep digital materials? Rights holders' benefits. Who should pay?	Copyright issues.	Initial capture into an electronic record-keeping system.
4	Financial implications (Who pays? Who benefits?).	Need to use open standards in storing the data (e.g. SGML rather than Microsoft Word).	Funding and costs.	Access.
5	Migration and emulation from one generation to the next.	Rights owners' benefits.	How to organise all aspects of data emulation and migration.	Financial viability and responsibility.

Authors and data originators stress the need for a common strategic approach to archiving digital materials and (not surprisingly) for attention to intellectual property considerations, envisaging that the unilateral approach of libraries to decisions about preservation will have to give way to greater collaboration in future. They see national coordination of digital archives as a more appropriate way forward than a centralised national digital repository, hoping to allow groups with specific interests to manage their own archives whilst ensuring that material is more generally available. They feel that a common approach to archiving digital material entails providing a framework for guidelines (covering such issues as emulation and security as well as best practice in the area) which is not prescriptive, since otherwise it is likely to be ignored.

They see market demand as a major driving force in deciding what is to be archived and suggest that a voluntary approach, funded (at least initially) by the interested parties, might be more appropriate than asking for public funding for an unfocused national resource.

Publishers assume that a national archiving strategy based on the legal deposit libraries will emerge. Accordingly, they see a need to separate the repository function of bodies such as the British Library and their role of providing a document delivery service to make information more widely available. They see the British Library's

Document Supply Centre as a commercial document provider which is, in effect, competing on unfair terms.

They feel that transfer of electronic publications could take place at the point of publication, as long as there are restrictions on release to the general public. If no guarantees can be made to protect publishers' commercial interests, the alternative would be to hand over the material at the point (defined by them) at which the publication is no longer of commercially exploitable value. While an electronic publication is commercially exploitable they feel that it is in the publishers' interests to maintain (and retain) it. (It should be perhaps be noted that this view is not shared by other groups who argue that this approach would leave some material vulnerable to neglect.)

Although publishers accept the idea of preservation for the common good, they argue that the humanistic goal of national collections being maintained and preserved as a service to scholarship is no longer tenable because of the large volume of material and high preservation costs involved.

Learned society publishers may take a different view from commercial publishers since they have always accepted a greater responsibility for meeting the needs of the intellectual community they represent. An archiving role may also be in their own interests.

The need to review the assumptions under which the deposit libraries operate is also highlighted by the **distributors**. The question of who should make the decision about whether or not to archive an item will require collaboration between librarians, archivists, and publishers. They feel that it is unrealistic to depend on the publishers alone to ensure preservation of digital publications, even in the short term.

This group see preservation and access as separate issues, whilst recognising that preservation only has a purpose if access is allowed to the material. However, they too accept the idea of preservation for the public good.

The group envisage various possibilities, from a systematic to an accidental approach. Taking a broad historical perspective, they argue that the sheer volume of digital material favours accidental preservation, with those items for which there is a continuous demand most likely to survive.

Although some types of digital publication, such as CD-ROMs, are seen as sufficiently coherent to be collected and archived, others, such as Internet sites, are not. Hypertext links present particular problems because of the difficulty of preserving the ability to link with other sites.

Contracts between archivists and information providers, as well as cooperation amongst copyright holders to permit voluntary or legal deposit, are seen as part of the necessary response to the breaking down of traditional divisions of functions between publishers and repositories.

Distributors question whether the collecting strategy of the past centuries, based on keeping every edition, is still relevant today. Whereas in the past the printing process was sufficiently laborious to ensure that publishing was relatively rare, electronic material can change so quickly that this approach may no longer be appropriate.

Predictably, the **repositories** take a somewhat different view of access and deposit issues. They suggest that in the future preservation will require both an access copy and an archival copy of the data. The problem of legal deposit as it stands is that books on deposit are as accessible as any other book held. However, for electronic documents the same principle probably should not apply. They see the access and preservation roles becoming much more distinct for electronic documents.

They feel that the source of the data should determine how large data sets should be treated. Commercial publications should have strictly limited access with the emphasis on preservation. Public data (such as that coming within the remit of the Public Record Office) should be more widely available.

Repositories could save ‘pointers’ to the material while publishers maintain the actual archives (and at the same time exercise some control over access).

They acknowledge a ‘heritage role’ for digital archives, requiring access by the academic community. This is likely to lead to a number of new operating arrangements to secure appropriate access. For, example the British Film Institute has based its policy on the access role rather than purely on preservation, partly enabled by technology. The Data Archive has issued CD-ROMs of data for exclusive scholarly use and, where appropriate, proscribes use by commercially sponsored academics. This system depends on password access and undertakings by researchers not to use the data beyond agreed terms. A single-user licence could limit access to networks, but difficulties would remain over downloading of data and their subsequent exploitation.

When considering specific questions about the preservation of digital material the group as a whole offers a very diverse range of views.

Of the organisations contacted, three-quarters either have a policy on digital archiving, are working towards a policy, or feel that a policy is necessary. However, there is a also a perception that the responsibility for developing an archiving policy lies elsewhere, perhaps with an international or professional association, or through the requirements of legislation. Although some organisations have comprehensive policies in operation, others are strongly biased towards physical storage, with little consideration of access issues or migration to new media.

Views on who should be responsible for holding and archiving digital materials vary widely:

- A National Digital Repository, or variations on this theme, along with the appointment of a National Digital Preservation Officer. However, some doubt that a central repository would be able to fulfil the needs of all the different depositors.
- National libraries/legal deposit libraries/the British Library.

- The National Sound Archive is seen as a good example of a voluntary system of deposit.
- A system of distributed repositories, or a body to coordinate different repository agencies.
- Government, government departments or the public domain.
- Rights holders should be responsible for archiving their own material.
- Specialist agencies and those with historical responsibility for archiving.

Many representatives of this group see legal deposit as the most appropriate mechanism for ensuring that digital material is preserved, although there are fears about access and further thought is needed as to how the system would work in practice. Most agree that a voluntary code is unlikely to be satisfactory, although it could be useful initially until legislation is introduced.

Most also agree that once a digital object has been selected for preservation it should be kept forever, although some see that there are problems with the practicalities of this aspiration.

Opinions differ as to the format in which material should be kept. Some argue for the maintenance of digital materials in a standard format. Others express the view that archived material should be kept in its original format so as not to distort appearance or content. They suggest that any transformation or migration will lead to loss of data, but acknowledge that there are also problems in reading and accessing old technology.

The most common response to the question of funding is that it should be a national responsibility, funded by the government through taxation, although some feel that others have some role to play in funding — publishers, the creators of information (for example, the academic community), libraries, users.

Perhaps the greatest divergence of opinion is on access, with views ranging from the extremes of denying all access (i.e. material should be held purely for preservation) to making material freely available at no cost. Most agree that access should be allowed, but with restrictions, and that access and copyright should be subject to negotiation.

2.3 The research community

Electronic resources are being funded and created within the research community at an unprecedented rate, both in terms of data and of ‘electronic paperwork’. The increase is due to the ubiquity of hardware and software, and the electronic infrastructure now available to researchers. The funding agencies are supporting projects which produce electronic materials. Large quantities of electronic materials are being created within UK universities. All of these are expected to increase.

There is now a growing concern that as more digital information is produced — much of it held only in digital form — it is essential to develop strategies for the selection and preservation of such material.

There is evidence to suggest that, although some representatives of the research community have a clear understanding and recognition of the importance of digital preservation, in the sector as a whole there is a lack of awareness of the need for preservation policies for electronic research materials. There are strong arguments for continuing to make digital information available, including: the need for scientific advances to build on what has gone before; principles of openness and replicability; making greater use of limited and expensive resources; and reducing the burden on respondents to surveys.

Many **funding agencies** report an increase in the number of projects creating electronic materials, and expect this trend to continue. As well as researchers within the university sector, some funding agencies also support individual researchers, in-house research, local authorities and school teachers. They also fund research in collaboration with other research agencies and with commercial companies, and these collaborations often go beyond national borders. Any policies for digital preservation will therefore have to accommodate the diversity of grant recipients as well as the variety of research carried out.

Examples of electronic data produced by research include: scientific experimental data; epidemiological data; social surveys; clinical trial data; metadata providing a description of metadata files; electronic teaching materials; musical and choreography notation; audiovisual files; and multimedia databases.

'Electronic paperwork' has increased the amount of electronic material available for preservation. Electronic mail, bulletin boards and mailing lists have become an important means of communication between academics. Funding agencies are also introducing systems to enable the online completion of grant applications. Hardware and software are being developed to meet an ever widening range of tasks: for example, in laboratory research many of the instruments used now produce an automatic electronic output.

Funding agencies agree that the electronic materials produced by researchers constitute a valuable resource and should be preserved. The reasons for preservation include making the best of resources through secondary analysis, for evaluation and replication of past research, and the investigation of scientific misconduct. There are some concerns, however:

- the independence and impartiality of the funding body could be compromised by secondary analysis
- quality control over secondary analysis could not be guaranteed
- respondent confidentiality and respect for the sensitivity of data could be placed in jeopardy
- the security of preservation technologies (for electronic data) is suspect
- sensitive or commercially valuable data might fall into the hands of unsuitable individuals or bodies.

Only a minority of funding agencies have established policies and/or guidelines regarding the preservation of electronic materials. The firmest of the policies are those of the Economic and Social Research Council (ESRC) and the Natural Environment

Research Council (NERC). ESRC researchers are obliged to discuss data access at an early stage and to offer any electronic materials to The Data Archive at the University of Essex. NERC's policy is that grant recipients must satisfy the funder that they have the willingness, expertise and commitment to preserve data themselves or that they have made arrangements with an appropriate data centre to take responsibility for their data at a specified time point. While many research funders recognise the value of secondary analysis, ESRC is alone in actively encouraging secondary analysis over primary data collection.

Selection is one of the biggest problems. Limitations of resources mean that no archive can accept all material that it is offered. Materials preserved must include contextual information by which the materials can be understood and used correctly. The prediction of what will be useful in the future needs to be improved in order to avoid the accumulation of material of low value — and this will require an understanding of what constitutes value. Some datasets require constant updating in order to maintain their value and it is unclear who has responsibility for ensuring the integrity of the updated material.

There are also concerns about ownership. Ownership of electronic materials is rarely established clearly. Funded data collected by universities is the property of those universities and the research councils have no official authority over what happens to it, although they may exercise some unofficial leverage.

Ownership of data brings responsibilities in terms of access and preservation — although not all of the research funders who claim ownership accept the preservation of materials as their responsibility.

It is widely suggested that the creation of a culture of data sharing is needed, best achieved through a campaign promoting awareness of data preservation and its value. A statement of national policy and guidelines would enable agencies to be more confident in their requirements of researchers and would help to allay fears about the safety of materials once preserved, and the ease with which they can be located and accessed. Some agencies favour a centralised repository rather than a dispersed network, while others feel that a central body might threaten their independence. Most agencies agree that there should be national funding for the preservation of electronic resources, but not all feel able to make a contribution to such a fund.

There is extensive creation of digital materials by **university research centres and projects**. These include: numerical data derived from experiments and surveys; papers and reports in the form of text files; qualitative research data, such as interview transcripts and field notes; databases from historical sources; multimedia, audiovisual materials and World Wide Web pages; electronic tutorials and teaching materials; electronic mail messages; project administration files, agendas and minutes of meetings.

The sharing of electronic materials appears to be common within the academic sector. In most cases distribution is in the form of CD-ROM or over the Internet, particularly the World Wide Web, often through the production of specific gateways. Access to

materials is most often achieved through informal agreements with the creators and copyright holders.

Many researchers agree that the re-analysis of data is a central principle of scientific scholarship and that secondary analysis can also bring cost benefits. On the whole they do not see a serious problem in predicting which digital materials are likely to be useful in the future. Nor do many consider that materials can only be fully understood by the original researcher, although they do recognise the difficulties involved in understanding other people's work.

Many researchers do not seem to know whether their funder has a policy regarding the preservation of the electronic materials they are creating, although most are strongly in favour of a policy being formulated. Some projects have their own internal preservation policies and some offer their materials to an archive, such as the Data Archive or an Arts and Humanities Data Service (AHDS) centre. The problems experienced when attempting to establish a preservation policy are:

- lack of national guidance
- lack of interest in out of date materials
- uncertainty about technical standards
- insufficient resources to carry out preservation.

Reasons for not archiving materials include: a lack of awareness about archives; the absence of archives suitable for the materials being created; the need for adequate metadata; and concerns about the legitimacy of lodging material in an external archive when the intellectual property rights are not clearly defined. In most cases the decision to deposit materials is dictated by the policy of the funding agency.

Many researchers consider the preservation of digital materials to be the joint responsibility of the projects generating the materials and the funding agencies. Ownership is a central concern. It conveys control, including the right and the responsibility to preserve the material; it is often closely linked to copyright; and it confers responsibility for ethical issues, accuracy and usability. In terms of funding, the general feeling is that fewer funds for primary research in order to fund preservation would be unacceptable, despite the cost benefits of secondary analysis in the long run.

In common with the funding agencies, many researchers favour a national body to advise on preservation policy and monitor all relevant developments in standards of best practice. A distributed network is also favoured to allow greater security and reliability, distribute work and storage capacity, avoid a monopoly situation, and allow for regional or discipline specific differences.

Although **universities and higher education institutions** (HEIs) are creating large and increasing quantities of electronic materials, none of the universities surveyed has established any procedures, policies or guidelines covering the preservation of electronic materials at their institutions. The reasons for this include insufficient resources to allocate to the task, lack of knowledge about electronic preservation, or

conflicting priorities. In some cases the problem of data preservation has simply not been recognised.

Most vice-chancellors agree that the preservation of electronic materials should be a joint responsibility of researchers and the research funding agencies, although some see it as the remit of 'other organisations', such as the Higher Education Funding Councils' (HEFC) Joint Information Services Committee (JISC) and the British Library. Others stress the importance of flexibility, with the solution dependent on the nature of the electronic material. Again, ownership is seen as problematic; ownership of electronic materials is often unclear and digitisation exacerbates the problem of what constitutes intellectual property.

Questions of funding elicit a mixed response from vice-chancellors. Some suggest that preservation should be funded 'nationally' or 'centrally' from the public purse, or from HEFC funds, while others believe that the funding agencies should foot the bill. Other proposed sources include the private sector, and funds generated from commercial partnerships or from end users. Few believe that universities should pay for preservation directly, although some feel that their institution would be willing to contribute resources to it.

They agree that a national body responsible for advising on preservation policy would be useful, and favour expanding the role of an existing institution over the establishment of a brand new body, the most suitable candidates being the British Library, the National Preservation Office or the Consortium of University Research Libraries.

Vice-chancellors also suggest that a campaign to raise awareness of the problem of preserving electronic materials and a programme of education regarding preservation techniques would be useful. These should target a wide range of groups including the creators of digital materials, universities, publishers, database hosts, government departments, commercial organisations with a high level of involvement in research, central support services and research funders. This group also feels the need for a wider programme of research to tackle the range of political, legal and technical issues that remain unresolved.

Taking the research community as a whole, three main areas of concern are apparent:

- a lack of awareness of the need for preservation policies for electronic research materials;
- a need for advice, standards and national policy to plan for preservation;
- the need for a centre, or a distributed network of centres to provide preservation facilities for those without the facilities and resources to provide their own long-term preservation.

It is also clear that the form which centres of preservation should take and the responsibility for undertaking this work cannot be addressed until some serious issues have been debated and clarified. These issues centre on the link between ownership of the materials and the rights and responsibilities to preserve. The funding of research in the UK is very complex, making it impossible to draw a simple line between

ownership and responsibility for preservation. Preservation is by its very nature a complex, expensive and, above all, a long-term commitment, while many research projects are short-term and shifting in funding, location and staffing. The potential value of the material they produce is inestimable, however, and in many cases its loss will be a loss to the nation's cultural and scientific heritage.

2.4 Conclusions

There is always a danger that surveys will over-represent those with an interest in the topic and that conclusions will be based on a self-selecting, unrepresentative minority, despite strenuous efforts to avoid this outcome. The views of non-respondents are likely to be just as interesting as the contributions of those who did participate — if only we could get at them. Nevertheless, it is possible to draw some broad conclusions about the perspectives on digital preservation of both communities.

1. There is concern across all sectors that resources are being lost and agreement on the need for a campaign to promote awareness of data preservation.

2. There is an acknowledgement of the role of the creators of digital information and their responsibility for its long-term preservation. This is linked to concerns about the ownership of digital material and the protection of intellectual property rights.

3. There is common concern about the costs of preservation, especially since the scale of costs involved is an unknown factor.

4. There is no consensus, however, on how digital preservation might be financed, although many indicate that some form of national funding is necessary.

5. There is also agreement on the need for collaborative developments, and for shared and agreed policies.

6. All express the need for guidance — a national policy and guidelines covering preservation of electronic materials. Many feel that a central national body should lead on preservation policy and monitor all relevant developments in standards and best practice.

7. Overall there is a lack of established policies and guidelines and evidence to suggest that the majority of organisations have not thought through the implications of digital preservation. Clearly, there are exceptions to this, and it is to those organisations which have taken a lead that we must look for examples of best practice.