erpa**studies**

publishing

www.erpanet.org

ERPANET – Electronic Resource Preservation and Access Network – is an activity funded by the European Commission under its IST programme (IST-2001-3.1.2).

Further information on ERPANET and access to its other products is available at http://www.erpanet.org.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int).

ELECTRONIC **R**ESOURCE **P**RESERVATION AND **A**CCESS **N**ETWORK

Table of Contents

Executive Summary

The publishing sector is becoming increasingly aware of the importance and value of digital information and of the problems and challenges that lie in preserving it and providing access to it over time. The amount of information produced, received, and managed in digital form increases daily in the publishing sector, and radically affects the main business functions and processes. Most of the seven organisations that participated in this study understand that the main issues affecting them must be resolved in concert with other publishers and organisations that have a stake in the sector. Controls need to be put in place for the creation, management, and preservation of valuable information, and should be developed with other publishers and business partners. This sort of activity is only becoming visible and much still remains to be addressed.

A fuller knowledge within the sector of the demands and challenges of preserving digital information has to be fostered before action can be taken. Certain larger organisations have begun to consider the loss of valuable digital assets a serious risk and are working to find solutions to the problems of cost, standards, establishing responsibility for action. Problems are also raised by proposed technical solutions such as format change, degradation of hardware and software, and quality and completeness. The remaining, smaller organisations, wary of high costs and untested potential solutions, prefer to wait and see what develops in the sector, especially the products of the developing partnerships with major libraries. The organisations involved in this study present a range of examples of understanding and activity, highlighting the work that has been done, and which still must be tackled.

Chapter 1: The ERPANET Project

The European Commission and Swiss Confederation funded ERPANET Project[1] (Electronic Resource Preservation and Access Network) works to enhance the preservation of cultural and scientific digital objects through raising awareness, providing access to experience, sharing policies and strategies, and improving practices. To achieve these goals ERPANET is building an active community of members and actors, bringing together memory organisations (museums, libraries and archives), ICT and software industry, research institutions, government organisations, entertainment and creative industries, and commercial sectors. ERPANET constructs authoritative information resources on state-of-the-art developments in digital preservation, promotes training, and provides advice and tools.

ERPANET consists of four partners and is directed by a management committee, namely Seamus Ross (HATII, University of Glasgow; principal director), Niklaus Bütikofer (Schweizerisches Bundesarchiv Hans Hofman (Nationaal Archief/National Archives of the Netherlands), and Maria Guercio (ISTBAL, University of Urbino). At each of these nodes a content editor supports their work, and Peter McKinney serves as a co-ordinator to the project. An Advisory Committee with experts from various organisations, institutions, and companies from all over Europe gives advice and support to ERPANET.

---

[1] ERPANET is a European Commission funded project (IST-2001-32706). See www.erpanet.org for more details and available products.

Chapter 2: Scope of the Case Studies

While theoretical discussions on best practice call for urgent action to ensure the survival of digital information, it is organisations and institutions that are leading the drive to establish effective digital preservation strategies.[2] In order to understand the processes these organisations are undertaking, ERPANET is conducting a series of case studies in the area of digital preservation. In total, sixty case studies, each of varying size, will investigate awareness, strategies, and technologies used in an array of organisations. It is anticipated that upwards of 500 organisations, institutions and public bodies will eventually contribute to this research. The resulting corpus should make a substantial contribution to our knowledge of practice in digital preservation, and form the foundation for theory building and the development of methodological tools. The value of these case studies will come not only from the breadth of sectors included, but also through the depth at which they will explore the issues.

ERPANET is deliberately and systematically approaching disparate sectors from industry and business to facilitate discussion in areas that have traditionally been unconnected. With these case studies ERPANET will broaden the scope and understanding of digital preservation through research and discussion. The case studies will be published to improve the approaches and solutions being developed and to reduce the redundancy of effort. The interviews are identifying current practice not only in-depth within specific sectors, but also cross-sectorally: what can the publishing sector learn from the aeronautical sector? Eventually we aim to use this comparative data to produce intra-sectoral overviews.

This cross-sectoral fertilisation is a main focus of ERPANET as laid out in its Digital Preservation Charter.[3] It is of primary importance that disparate groups are given a mechanism through which to come together as best practices for digital preservation are established in each sector.

Aims

The principal aims of the study are to:

- build a picture of methods and match against context to produce best practices;
- accumulate and make accessible information about practices;
- identify issues for further research;
- enable cross-sectoral practice comparisons;
- enable the development of assessment tools;
- create material for training seminars and workshops; and,
- develop contacts.

---

[2] Chapters 2 and 3 are taken from 'Cross-sectoral Development of Digital Preservation Strategies: ERPANET and the Expansion of Knowledge', given at *Preservation of Electronic Records. New Knowledge and Decision-making*, Symposium 2003.
[3] The Charter is ERPANET's statement on the principles of digital preservation. It has been drafted in order to achieve a concerted and co-ordinated effort in the area of digital preservation by all organisations and individuals that have an interest and share these concerns.
http://www.erpanet.org/www/content/documents/Digitalpreservationcharterv4_1.pdf.

Potential sectors have been selected to represent a wide scope of information production and digital preservation activity. Each sector may present a unique perspective on digital preservation. Organisational and sectoral requirements, awareness of digital preservation, resources available, and the nature of the digital object created place unique and specific demands on organisations. Each of the case studies is being balanced to ensure a range of institutional types, sizes, and locations.

The main areas of investigation included:

- perception and awareness of risk associated with information loss;

- understanding how digital preservation affects the organisation;

- identifying what actions have been taken to prevent data loss;

- the process of monitoring actions; and,

- mechanisms for determining future requirements.

Within each section, the questions were designed to bring organisational perceptions and practices into focus. Questions were aimed at understanding impressions held on digital preservation and the impact that it has had on the respective organisation, exploring the awareness in the sector of the issues and the importance that it was accorded, and how it affected organisational thinking. The participants were asked to describe, what in their views, were the main problems associated with digital preservation and what value information actually had in the sector. Through this the reasons for preserving information as well as the risks associated with not preserving it became clear.

The core of the questionnaire focused on the actions taken at corporate level and sectoral levels in order to uncover policies, strategies, and standards currently employed to tackle digital preservation concerns, including selection, preservation techniques, storage, access, and costs. Questions allowed participants to explore the future commitment from their organisation and sector to digital preservation activities, and where possible to relate their existing or planned activities to those being conducted in other organisations with which they might be familiar.

Ten organisations in each sector, and three people within each organisation are targeted for each study. In reality this proved to be problematic. Even when organisations are identified and interviews timetabled, targets often withdrew just before we began the interview process. Some withdrew after seeing the data collection instrument, due in part to the time/effort involved, and others (we suspect) dropped out because they realised that the expertise was not available within their organisation to answer the questions. The perception of risks that might arise through contributing to these studies worried some organisations, particularly those from sectors where competitive advantage is imperative, or liability and litigation issues especially worrying. Non-disclosure agreements that stipulated that we would neither name an organisation nor disclose any information that would enable readers to identify them were used to reduce risks associated with contributing to this study. In some cases the risk was still deemed too great and organisations withdrew.

Chapter 3: Method of Working

Initial desk-based sectoral analysis provides ERPANET researchers with essential background knowledge. They then conduct the primary research by interview. In developing the interview instrument, the project directors and editors reviewed other projects that had used interviews to accumulate evidence on issues related to digital preservation. Among these the methodologies used in the Pittsburgh Project and InterPARES I for target selection and data collection were given special attention. The Pittsburgh approach was considered too narrow a focus and provided insufficient breadth to enable full sectoral comparisons. On the other hand, the InterPARES I data collection methodology proved much too detailed and lengthy, which we felt might become an obstacle at the point of interpretation of the data. Moreover, it focused closely on recordkeeping systems within organisations.

The ERPANET interview instrument takes account of the strengths and weaknesses from both, developing a more focussed questionnaire designed to be targeted at a range of strategic points in the organisations under examination. The instrument[4] was created to explore three main areas of enquiry within an organisation: awareness of digital preservation and the issues surrounding it; digital preservation strategies (both in planning and in practice); and future requirements within the organisation for this field. Within these three themes, distinct layers of questions elicit a detailed discovery of the state of the entire digital preservation process within participants' institutions. Drawing on the experience that the partners of ERPANET have in this method of research, another important detail has been introduced. Within organisations, three categories of employee were identified for interview: an Information Systems or Technology Manager, Business Manager, and Archivist / Records Manager. In practice, this usually involved two members of staff with knowledge of the organisation's digital preservation activities, and a high level manager who provided an overview of business and organisational issues. This methodology has allowed us to discover the extent of knowledge and practice in organisations, to understand the roles of responsibility and problem ownership, and to appreciate where the drive towards digital preservation is initiated within organisations.

The task of selecting the sectors for the case studies and of identifying the respective companies to be studied is incumbent upon the management board. They compiled a first list of sectors at the very beginning of the project. But sector and company selection is an ongoing process, and the list is regularly updated and complemented. The Directors are assisted in this task by an advisory committee.[5]

---

[4] See Appendix. We include the questionnaire to encourage comment and in the hope that other groups conducting similar research can use the ideas contained within it to foster comparability between different studies.
[5] See www.erpanet.org for the composition of this committee.

Chapter 4: Introduction to the sector

The publishing sector can be categorised into five organisational types: Trade publishers, Periodical publishers, Scientific, Medical, and Technical publishers. Trade publishers produce consumer books, including fiction, non-fiction, children's, and reference books, which are sold mainly through booksellers. Periodical publishers issue magazines and newspapers (publications that appear more or less at regular intervals) and Scientific, Technical, and Medical Publishers mainly produce scientific research information.

Publishing organisations are concerned with the actual publishing of the information, and also the selection, editing, designing and distribution of the material. In addition, publishers promote and maintain relationships with the disseminators of their products such as booksellers and libraries.

The book and learned journal publishing industry in Europe had revenues of €19 billion in the year 2000. There are in excess of 3,320,000 individual titles available from book and journal publishers in the European Union[6], and an estimated 50,000 magazine titles in Europe, reaching an average of 80% of European adults. The combined newspaper and magazine market is worth €81 billion[7]. Publishing in Europe is a very decentralised industry, with the magazine and newspaper being particularly fragmented as the vast majority are either family-controlled or family-owned. The publishing industry has seen many changes in structure and profit in recent years, and these will continue into the near future. In 2000/2001, 286 merger and acquisition transactions were completed, with a total disclosed value of about €10.6 billion[8].

Increased competition from radio and television, a reduction in advertising revenues, freely available information on the Internet, electronic publishing, and the development of new partnerships represent some of the prominent issues highlighted by those within the publishing sector. With the increasing prominence of new technologies, the challenge lies in being able to exploit these and remain competitive.

---

[6] Federation of European Publishers, http://www.fep-fee.be/.

[7] Overton, Anna & Trevor Pritchard, *Quality Portfolios Drive Credit Prospects for Europe's Newspaper and Magazine Publishers,* Standard and Poors, April 2002. http://www.standardandpoors.com/europe/francais/Fr_forum/articles_sectoriels_et_commentaires/Newspaper-and-Magazine-Publishers_08-04-02.html.

[8] *Ibid.*

Chapter 5: Details of the Interviews

A variety of publishers were approached for this case study with the intention of providing representation from a wide range of publishing organisations. Most publishing organisations, when approached, were interested in the ERPANET project and the study, and were keen to take part in some capacity. Although two major publishers subsequently pulled out, we were able to replace them without too much difficulty. In total, we received information from seven publishers, detailed below.

Gruppo Mondadori
http://www.monadori.com/
The main activities of the Mondadori Group are book and magazine publishing, graphics and printing, and information services. The company forms part of the Italian Fininvest Group which is also active in television, cinema, sports, finance and insurance. Its magazine division publishes women's magazines (*Donna Moderna*), TV guides (*Guida TV*), news magazines (*Panorama*), and other publications on topics such as home, travel, and entertainment. It also publishes Hearst (*Cosmopolitan*) titles. Italian Prime Minister Silvio Berlusconi controls Mondadori through his holding company Fininvest. Gruppo Mondadori employs 5,100 staff and has annual revenues of €1.4 billion.

Hachette Filipacchi Médias
http://www.hachette-filipacchi.com/
Hachette Filipacchi Médias has a presence in 34 countries. It edits and publishes 42 magazines in France relating to current affairs, television, women, leisure, travel, and young people. Hachette Filipacchi Médias is the leading foreign publisher in the United States, Spain, Italy, China, and Japan, and has also developed a strong Internet presence. The company employs 8,939 people in total, 3,942 internationally, and has annual revenues of €2.2 billion.

HMSO (Her Majesty's Stationary Office)
http://www.hmso.gov.uk/
Founded in 1786, HMSO manages and regulates the utilisation and licensing of the reuse of all information produced by the UK government that is protected by Crown copyright. HMSO manages the licensing of Parliamentary Copyright, and oversees the printing and publication of all UK legislation and related official materials in traditional print formats and via the Internet. It also advises government departments on all aspects of official publishing. HMSO oversees the production of the Civil Service Year Book (CSYB) which is the definitive official reference for central government in the United Kingdom and is published twice a year in print and updated on the internet.

Kluwer Academic Publishing

http://www.wkap.nl/
Kluwer provides services for academic and corporate researchers in humanities & social sciences, environmental & plant sciences, physical sciences, behavioural sciences, engineering & computer sciences and biosciences. The company's headquarters are in Dordrecht, The Netherlands. The Kluwer Online gateway at www.kluweronline.com, reaches millions of researchers, scientists and professionals in more than 50 countries with electronic Journals, eBooks, Custom Books, and eReference Works. At the time of the interview, the parent company was Wolters. Since then, Kluwer Academic Publishers has been divested and is no longer part of Wolters Kluwer NV. Candover and Cinven, two financial investors, now own the majority of the equity in this company. Kluwer (http://www.wolterskluwer.com), a

multinational (25 countries) information services company with annual sales of more than €3.8 billion, employing approximately 20,000 people in Europe, North America, and Asia Pacific.

Office for Official Publications of the European Communities (Publications Office)

http://eur-op.eu.int/general/en/index_en.htm
The Office for Official Publications of the European Communities, based in Luxembourg, is the publishing house of the institutions and other bodies of the European Union (EU). It is responsible for producing and distributing EU publications on a variety of media. Its organisation and operation are laid down by Decision 2000/459/EC, ECSC, Euratom. Under the provisions of the EU Treaties, the publication of certain titles, such as the *Official Journal of the European Communities* or the *General Report on the Activities of the European Union* is a legal obligation. Other titles are regarded as essential for the development of the Union and its policies, while some publications are intended as information channels either for the general public or for professionals. Currently, the Office publishes daily in the 11 working EU languages, as well as in Gaelic and in other languages as required. The Publications Office employs 522 staff and has an operating budget of €56.42 million.

Oxford University Press

http://www.oup.co.uk/
OUP as it exists today took shape in the 1650s. OUP began to expand internationally in 1896, opening its first office in the United States that year. Today OUP USA is Oxford University Press's second major publishing centre after Oxford, producing annually nearly 500 titles. Music, journals, and electronic publishing have all been introduced within the last 75 years. OUP is now one of the largest publishers in the UK, and the largest university press in the world, publishing more than 4,500 new books a year, with a presence in over fifty countries, and employing some 3,700 people world-wide. It has a diverse publishing programme that includes scholarly works in all academic disciplines, bibles, music, school and college textbooks, children's books, materials for teaching English as a foreign language, business books, journals, and dictionaries and reference books.

Random House

http://www.randomhouse.com/home.pperl
Random House, Inc. is the world's largest English-language general trade book publisher and is a division of Bertelsmann AG, one of the largest media companies in the world. RH produces around 8,000 new publications a year and has more than 100 publishing companies across 13 countries.

Random House, Inc. assumed its current form with its acquisition by Bertelsmann in 1998, which brought together the imprints of the former Random House, Inc. with those of the former Bantam Doubleday Dell. Together, these groups and their imprints publish fiction and non-fiction, both original and reprints. They appear in a full range of formats – including hardcover, trade paperback, mass market paperback, audio, electronic, and digital. The readership spans the full age range from adults to young adults and children. Bertelsmann employs 74,000 people in more than 300 companies and operating units. Annual sales in fiscal year 99/00 were $17.6 billion.

Chapter 6: Circumstances

On the whole, few complications were encountered in locating and gathering information from the participants in the publishing sector. Some of the organisations had large multi-national structures, while others were smaller and more compact, but all were most helpful in finding suitable personnel within the organisation who were able and willing to discuss and share their thoughts on issues surrounding digital preservation.

A common feature of these organisations was a fragmented internal structure where departments have little practical interaction with each other. This made it quite difficult to locate individuals who could speak of organisation-wide practices. This, however, meant that a range of participants in the same organisation had quite a different perspective on the organisation's priorities and activities which enabled an analysis of its working realities. Perspectives were very much shaped not only by the position held in the company but also the position that the department had within the organisational structure.

For the majority of the seven organisations, more than one person was questioned in order to provide the study with a variety of perspectives. Where there was more than one person, we aimed to ensure they were located in a different part of the organisation with differing positions, competencies, and responsibilities. Contributing to this study were editors, electronic publishing editors, project managers, information officers, e-business consultants, information technology managers, archivists, library managers, marketing managers, and business systems managers. Their views therefore represent a variety of organisational and preservation perspectives.

Chapter 7: Analysis

This section presents an analysis of the data collected during the case study. It is organised to mirror the sequence of topics in the questionnaire.

- Perception and Awareness of Digital Preservation

- Preservation Activity

- Compliance Monitoring

- Digital Preservation Costs

- Future Outlook

Perception and Awareness of Digital Preservation

The consensus amongst participants was that a general awareness and understanding of digital preservation exists in the publishing sector and shows signs of development and growth. However, this understanding lacks depth and maturity, and the overall impression given was that this is a topic of concern that does not register very high on the agenda of publishers and publishing organisations. The exception are those organisations involved in electronic publishing, where this has become a more prominent issue since e-publishing began about five or six years ago.

A dearth of information available to the publishing community on digital preservation inhibits the development of awareness and discussions are carried out in a predominantly *ad hoc* manner. Indications were there that external sources such as 'solution providers', including IBM and Microsoft, as well as the library and archival communities do help to raise awareness, but there is no authoritative source for publishing organisations to find information relating to their digital preservation problems.

Common initiatives developed to tackle issues of potential strategies and common standards are reportedly too narrow in focus. Significant developments in digital preservation have yet to materialise that will impact the operations and preservation activities of these publishing organisations. In only one case was it reported that external advice, a survey of external information sources and the activities of similar organisations was incorporated into the organisation's own digital preservation policies and strategies.

The Main Problems

A clear **understanding of the issues** central to digital preservation remains the primary problem to the development of solutions in this sector. A consensus of what digital preservation is, and how it affects organisations and the digital assets they hold remains to be realised. Another problem lies in the establishment of **responsibility**. This is a particular problem for the publishing sector as the main parties struggle to reach consensus about who must ultimately take responsibility for the maintenance and preservation of digital products. Frequently mentioned was the lack of widespread awareness of **standards**, despite the desire, available for the preservation of the kind of information created, used, and received in this sector. In addition, **costs** of digital preservation activities were emphasised as problematic due to a lack of sophisticated understanding and a general impression that they could

only be prohibitive for many organisations. Consequently, the difficulties inherent in presenting a strong business case for encouraging greater investment and commitment to digital preservation were only aggravated. Furthermore, **format changes**, **degradation of hardware and software**, and the **quality and completeness** of the information all posed concerns.

Asset Value and Risk Exposure

Awareness of the value of the data and information held by each organisation was investigated. Perception of value impacts both the risk exposure and information management activities, in particular enabling the prioritisation of action. Some of the types of information discussed in the study included books, journals, newspapers and magazines which are created and kept in digital form, including content, cover designs, artwork, bibliographic data, and images. Recent changes in practices mean that instead of a print copy being sent to the printer, they now receive an electronic copy. As a popular novelist can expect to be reprinted for the next twenty to thirty years, these electronic copies and related data need to be maintained in the long-term. In addition to this, business information such as production and sales costs are preserved, as well as data on sales and royalties. Print copies are almost always kept of all publications by the organisations questioned, in addition, some organisations also print out important business documents such as contracts and electronic mail communication. Some concerns were raised about organisations' Internet and Intranet sites that were not being actively preserved.

Responses on value ranged from the belief that digital information in itself has no real value in the organisation to a wider understanding of information as an exploitable asset with a high reuse value. These values, however, were often held on an individual basis in the absence of a clear organisational position.

Seventy five percent of respondents indicated that no risk or business needs analysis had been conducted by their organisation with regard to the preservation of digital information. This was not something that some organisations wished to spend money on. Of those that had conducted one, the analyses were narrow in focus, relating to a specific project or type of business process, and the information associated with it. Information was generally scarce in this area and participants were not aware of the specifics of any outcomes and what actively resulted.

Regulatory Environment

The regulatory environment consists of laws and regulations governing sector-specific and the general business environment as well as mandatory and voluntary standards and codes of best practice. The legal risks to the organisation proved to be the most compelling reason given for preservation, with business and financial coming a close second. Contracts and permissions must be kept permanently and these are increasingly produced in digital format. Some organisations also felt there was a historical impetus to preserve information, but others were adamant that if there was no financial gain to be made from it, and no legal requirement, digital information would not be kept. There were also some that felt that even though risks to digital material existed, they were minimal, as data and information could, in their opinion, 'almost always be recovered'. The regulatory environment governing this sector is not as strict as in other sectors and preservation activities in publishing organisations are not governed in a prescriptive manner. Publishers have more allowance to make decisions on information management and preservation without external censure.

Copyright legislation impacts publishers worldwide. Without copyright, publishers would not be able to trade profitably. Copyright is a protection that covers published and unpublished literary, scientific and artistic works, whatever the form of expression, provided such works are set in a tangible or material form for a fixed period of time. Increasingly, issues of copyright, translation rights, and fair use are becoming more complex with the publication of material in a greater range of formats.

Preservation Activity

Organisations typically direct their preservation activities in response to the value of the assets they hold and the risks that these and the organisation in general are faced with. A drive towards action is indicated through the development and recommendation of policies and strategies and the practical activities of selection, preservation, storage, and provision of access to the organisation's valuable assets. Despite the current apathy towards digital preservation in many areas, there are some positive initial activities being undertaken, although these were mainly on an individual or small-scale basis.

Policies and Strategies

Policies and Strategies provide evidence of the accountability and information about an organisation's preservation activities. Few formal policies exist, however, within the majority of publishing organisations questioned. This reflected the lack of a systematic approach in most of the organisations to retention and preservation. The development of policies and strategies remains relatively inert, and publishers tend to prefer to wait and see what is developed and what will emerge from within and external to the sector, and, in addition, what the market will demand. Some reported that while there was 'nothing on paper' (i.e. no formal policies or strategies) there were some general ideas about what should be kept, although not necessarily how. Some organisations did have specific policies for email, text, images, and transfer, but many respondents indicated that a good strategy was to keep everything. Mission statements were also felt to be good indications of organisational policy.

Where there was policy, or some minimum organisational direction in this area, it tended to result from the work of small task-forces or working groups set up within the organisation to tackle one aspect of the problem, or from individuals within the organisation who had a personal interest. Rarely did digital preservation policy or direction come from higher management levels, although a few were developed by the IT department which commonly held a strong central position in most of the organisations. This, to some extent, explains why most of the policies that were reported to be in place do not apply to the organisation as a whole. However, the fragmented nature of many publishing organisations will also have some bearing on this.

Some organisations however, do have stricter policies and strategies in place, particularly in regard to a certain publication. The *Office for Official Publications of the European Communities (Publications Office)* for example, has very strict organisation-wide controls over the handling and preservation of the *Official Journal* and the legislation that it publishes.

Participants had more information to provide about organisational strategies and the range of practices that had been implemented in the absence of formal

organisational policy. This less structured method of preservation seemed to dominate the responses and most strategies and practices that were in place had been developed by a department, or to solve a particular problem at a particular point in time. This meant that there was no significant effort to tackle the problem as a whole, or in any consistent manner. The existence of a library or an archives was considered by some to be a good organisational strategy, but often these library and archive staff felt that it was their own initiative and personal intervention that constituted the organisation's preservation activities.

Staff in the participating organisations were made aware of digital preservation policies, strategies and practices mainly through employee handbooks and the organisation's intranet site. In a few cases, there was no information made available for staff. What did exist was developed in-house. Usually this was either a department working independently, or with the IT department. In one instance, the archive would work with departments to try and build in preservation practices. There was some response to the question about policy update and renewal with reports of common organisation renewal every 2-3 years or every 5 years, or that the organisational practices were in a state of continual evolution. Generally, however, there was no apparent need for regular renewal or update, and responses indicated that as things were working well there was no need to make changes, and that when the need arises there would be further examination of the situation.

The lack of desire or perceived need for formal organisational policy was consistent with the overall lack of consensus about the value of digital information and the need to maintain and preserve it.

Selection

This section aimed to develop an understanding of how an appreciation of asset value was translated into the selection and management of that asset. Questions addressed how, in practice, organisations determined what needed to be kept, and for how long.

One third of respondents stated that there were selection, classification, and retention processes in place in their organisations. One company did have a records management software system that handled the classification and disposition of electronic documents and records. For some organisations, such processes only applied to certain types of information, such as publication data. For other types of information such as business and operations data, there was no intervention, and this would commonly be left to sit on the system until it was eventually removed. In most cases each department in the organisation set its own rules about what should be kept and disposed of. For those where there was no set organisational policy, there usually was general guidance and the organisation chose not to intervene any further. Only in one case were selection, classification, and retention policies linked and implemented across the entire organisation.

Selection and retention of digital information was the responsibility either of the department as a whole, managing editors, a specific unit within the organisation, or the IT department. There was little mention of any sort of organisational records management, or a dedicated unit responsible for the management of information assets across the entire organisation, although some mentioned that this was an initiative under consideration. Because of the lack of any centralised or formalised selection controls, all but a few respondents stated that there was no way to ensure that preserved information was complete, accurate, and identifiable. In some

companies it was the responsibility of the users of the information that had to ensure its accuracy.

Preservation

All of the organisations, despite the lack of formal policy, took steps to preserve certain types of information identified as being of some value. They either handled their preservation activities in-house or part in-house, part out-sourced. Stored information was often kept off-site, and some activities were contracted out, but managed by the organisation itself. The responsibility for the actual preservation processes lay in the most part with the IT department, and in some cases with the archives department. In others it was the responsibility of the offices or departments themselves (the producers of the information), or some identified staff member within them. Knowledge about any training or advice that was available for the people who were responsible for preservation activities was sparse, with answers ranging from 'none' to 'nothing specific'. Those that were interviewed who had preservation responsibilities mentioned having to source digital preservation information themselves, and also the lack of official competency that came with their position. For others, these activities were in addition to their 'main' responsibilities, and they also again lacked any formal training.

When asked about the organisation's awareness of external standards, best practices, and guidelines available on preservation, there was a wide range of responses. While a third of respondents were not aware of any, two thirds indicated an array of publishing, book industry, archival, library, and information technology standards, best practices, and guidelines. Despite the positive response, there was an indication that it was only the individuals being questioned that had knowledge of these, and not necessarily the organisation as a whole. If these individuals were not in a strong position in the organisation, there was little hope of these standards and guidelines having much impact. By far the standards, best practices, and guidelines which had the most impact and formed the bulk of responses were those from the ICT world. These were considered 'common sense' best practices and included regular backups, secure access, and offsite storage. Some storage standards were also mentioned including the BS5454 Standard (Recommendations for the storage and exhibition of archival documents) although these were mainly focussed on paper-based documents.

External factors also impacted the preservation activities of the organisations questioned, in particular, standards imposed by the book industry and libraries placing demands on publishers for the provision of content in a uniform, accessible format, such as PDF (Portable Document Format). This has also increasingly been the case with the production of online journals where there are also developing metadata requirements to comply with. Some publishers mentioned the use of standards developed by their 'technology partners', for example, the adoption of Microsoft's E-Book and digital rights management standards. Some of the respondents indicated that they were aware of external standards such as these, but that little use is made of them due to a preference for in-house developed procedures.

All respondents indicated that they did have a dedicated storage area for digital information to be stored. These ranged from standard IT network servers to libraries, archives, and RAID systems. Redundant copies of digital information were kept by all organisations, with the exception of one who indicated that all data was kept on a live server. In addition, hard copies were kept of most publications by almost all organisations, with some also retaining an image of each printed page in PDF format.

Most respondents related that their organisations were using standard data formats to manage and preserve their information, and that these were predominantly the formats in which the information had been created. Only two respondents indicated that there had been any sort of format conversion for the purposes of preservation. Some of the most popular formats employed were PDF, Microsoft Word, Microsoft Excel, Quark design files, TIFF, XML, and SGML. (One respondent noted that their organisation wanted to ensure that they follow the industry format, but that they were still waiting for a publishing industry format to emerge.) In the meantime, the relationship between the publishers and the wider sector was relatively balanced as neither is in a position to exert too much pressure on the other. If a bookseller were to impose a standard format of information transfer that was unacceptable to the publisher, they could simply withhold the title as owners of copyright. Neither is in a position to make too many demands, and this may in fact perpetuate the current situation.

There was a distinction between those who produce online journals and other texts, and those who do not with regards to metadata creations and maintenance. Publishers of digital journals have stricter metadata standards to follow, imposed by those who disseminate them, such as specific information inserted into an SGML (Standard General Markup Language) header that must accompany the content. Most of the rest of the publishers employ a version of the Dublin Core metadata schema for resource discovery, and a few others have none at all that they are aware of. Only a third of respondents stated that metadata creation is automated.

Digital preservation practices outlined in the responses were based on either *ad hoc* decision making and the evolution of previous practices in an organisation, or what available finances allowed. In some cases particular solutions were implemented as a result of a vendor relationship, with full adoption of a particular vendor's solutions and systems. For those printing out to paper as a preservation solution, it was because a reliable, economical, and sufficient system had not yet been found to meet the organisation's needs. In their estimation, an electronic solution would not be as easily accessible, and would be too expensive and time consuming to implement.

Access

All organisations reported that there were access controls in place to protect the organisation's digital assets and related contextual information, such as access privileges, passwords, logins, security logs, firewalls, and IP (Internet Protocol) filtering. For older information that was not immediately accessible online, there were additional controls in place such as a formal request procedure through an IT helpdesk. In general, IT departments were responsible for security arrangements. Copyright and privacy concerns were two of the most important reasons given for the restriction of access to the organisation's information. It was considered very important that only staff were only allowed access to products and upcoming publications. For some, controls were stricter, with only the department or unit responsible for the information able to access it without special permission.

To improve access, several organisations hope to increase access for both staff and third party clients and make their systems more robust and integrated. Some indicated that their organisation was considering ways to improve version control, and others talked about an increased role for scanning. Better relationships with partners such as major libraries were also seen as a way to make digital assets more widely available.

Compliance Monitoring

Monitoring of action is essential to ensure procedures and practices are implemented according to any organisational policies as well as external requirements and that they are meeting anticipated outcomes. Only a quarter of respondents stated that there was any monitoring and audit of the preservation process in the organisation. For the rest, it was enough that the organisation and its departments were following standard IT procedures, and that when information was used it was checked by staff. There seems to not be enough resources available for this sort of monitoring activity, and so organisations can only issue guidelines and hope these are followed.

Checks to preserved materials were also not undertaken in any formal or prescribed manner. In one case, however, systems were monitored daily, with availability statistics created and analysed monthly, and with random controls and publications orders being generated by automated scripts. For the rest, any checks that were being undertaken (and most respondents suspected there were) were probably being undertaken by the IT department, and in some cases the IT department and the organisation as a whole were audited either by an external company or their parent company.

The lack of formalised preservation processes in the first place is more than likely the cause of the lack of audit and monitoring of preservation activities. Monitoring an audit process can only be expected when there is a system in place to actually check.

Digital Preservation Costs

A major component of any organisational decision-making is cost. Cost was identified as one of the main problems associated with digital preservation at the outset, and remains one of the most uncertain. There are no specific cost examples available for organisations to make digital preservation decisions. Without a sound business case to present to organisational management, there is concern that investment in and development of preservation activities within organisations will be made increasingly difficult.

A reluctance to discuss and a lack of awareness about organisational finances and spending on digital preservation activities may have been in part because of a lack of formal cost analysis. With the exception of one organisation, there was no knowledge of where the money for digital preservation activities came from, or what section of the budget had provision for them. For most, the costs were assumed to be part of the IT budget, whereas others mentioned that preservation was funded on a project basis, where the organisation would devote money to a specific scheme or department. There was also no clear indication about the overall percentage committed by the organisation in relation to total expenditure. This sector was not aware of any available sources of funding within the publishing sector allocated for digital preservation issues. There was mention of some third currently willing to invest money in finding solutions, for example some National Libraries in Europe, a number of Higher Education Institutions, and some project money from foundations.

Only a quarter of respondents were satisfied with the amount of money that was currently being spent by their organisations on digital preservation activities. There was a degree of confidence displayed about committed resources, as the increase in investment in IT was seen as an increase in investment in digital preservation. For

the respondents who were not satisfied, concerns were raised about the narrow focus of spending and the resulting activity, which has yet to result in a comprehensive solution to the problem.

While there remains little formality and no clear notion of the costs of digital preservation activities, it will be difficult for those concerned with the preservation of digital information to present a sound business case to support their claims. Without statistics, it proves very difficult to make assessments on risk and plan ahead to incorporate digital preservation plans into an organisation's future.

Future Outlook

About a half of the organisations believed that no changes to digital preservation policy or strategy were required immediately or in the short-term unless a better methods or solutions were developed and made available. The other half, however, believed that imminent change was needed. All but one respondent stated that the money allocated for preservation would have to rise in the future to meet digital preservation demands. It was clear that investment would only increase, however, when there was a better understanding of the need, and once digital preservation was recognised as a business necessity and became more of a priority. For those concerned with technology, there was a clear need for the amount of money to increase as it was considered only natural that as technology will have a more important role to play, increased spending will be needed to accommodate this.

Respondents indicated that the biggest priority for investment was the development of common standards, such as a format solution being developed or agreed upon for storage. Increasing the numbers of personnel dedicated to digital preservation activities was also seen as a valuable way of allocating funds. There was also a call for greater co-operation and discussion, especially from the smaller organisations that feel that they are being left out of some of the developments in this area. Furthermore, the problem of responsibility should be resolved, with one respondent indicating that it should be the joint responsibility of libraries, governments, and publishers to find a solution to digital preservation problems. Another remarked that the digital preservation problem was like the Y2K problem, as it will only become relevant and important when the situation becomes critical. Only then will significant investment be directed at the problem. This echoes the view of many who stressed the need for a more clearly defined understanding of the issues and of the needs of publishers. Investment was also considered important to tackle a reduction in paper, the development of migration and audit procedures, emulation solutions, more research and development, quality control, and more integrated content management systems.

Questions about the organisation's perception of its preservation efforts provided some surprising answers. Over half of the respondents stated that their organisation's believed that their preservation activities were satisfactory, as there had been no major problems to date. An example made by one of the respondents was typical, they believed, of the nature of their organisation in regards to digital preservation. When first approached by ERPANET about participation, the request passed through a significant number of people before an individual was found that understood the problem well enough to become involved. It proved difficult to get a sense of the success or otherwise of the organisations' preservation activities as too many were still in the early stage of development.

The most important areas requiring attention included:

- finding technological solutions to digital preservation problems;

- the development and revision of preservation policies and strategies;

- wider recognition of the problems and challenges; and

- increasing the visibility of individuals in organisations with digital asset management and preservation responsibilities.

While most respondents would like to see increased cross-sectoral and intra-sectoral activity, many felt that this would not happen in the short term. It was considered unlikely that publishers would work very closely with each other, and it was felt that management would probably not be very keen due to the cost and time involved. Several respondents believed that it was important to work with other organisations like libraries and technology partners. One respondent stressed that by doing so it might help to clear up the ambiguity over responsibility, and instead of having three partners worrying about the same problem, it would be better if there was only one. There was a definite demand for more information resources on this topic, but no consensus about where these should come from. Initiatives were important and the sector working together as a whole was considered central. How this was actually to be achieved remained uncertain.

Chapter 8: Conclusions

Despite a good awareness of digital preservation issues and the value of digital assets, there is little cohesive thinking in terms of strategies, policies and action to meet the challenges posed by digital preservation. Where there is a clear risk, and a strong motivation for preservation, such as legal obligation, many organisations choose, in the absence of more suitable preservation methods, to make paper copies. As organisations evolve and continue to increase the amount of business and transaction conducted in electronic form, the relationship between information and the long-term practicalities of its preservation will need to be extended into the digital environment.

Certain assumptions about digital preservation are held by many in the sector and these look likely to remain in place and govern decision making until they are proved unsuitable. Most organisations do not consider themselves or the loss of digital information a serious threat.

Digital preservation in this sector is becoming more relevant, especially in the area of digital publishing. Changes in technology have allowed publishers to make their products available in many more formats, and have triggered concerns about digital preservation in the industry. The magnitude of the problem, and its immediacy are not prime concerns at present, but may become so if the examples of some are communicated with others. There is little in the way of intra and cross-sectoral activity in this sector, but recent high profile relationships and agreements between publishers and other institutions may encourage the sharing of ideas and strategies. The challenge remains to incorporate this into an organisation's methodology and system design, although reported initiatives sound promising.

A 'wait and see' mentality dominates this sector. There is a belief, however, that a common and practical solution can be found which will be adopted by the industries within, and related, to the sector. It is expected that industry formats, for example, will emerge as a result of co-operation between certain larger publishers and their activities with major libraries. Many of those concerned with digital preservation are hopeful that these relationships will provide smaller organisations with best practice guidance and a better understanding of responsibility.

The publishing sector remains relatively poised for action, with only a few major companies recently committing themselves to a more comprehensive involvement in tackling problems that have been identified. Greater investment is needed in all areas, but this will not arrive until good solutions are proposed, practical and standardised steps can be taken, and the costs and benefits fully explored.

Appendix: The Interview Instrument

ERPANET Case Study

Administrative Section

Interview Details
Organisation Details
Disclosure/Privacy Information
Tracking of Activities
Perception and Awareness of Digital Preservation

We would like to begin by asking you a few questions about your general impressions of digital preservation, and the impact that it has on the _____ sector. We will use the term 'digital information' throughout to refer to all forms of digital data, records and information.

1. Is there a general awareness in the _____ sector that the long-term preservation (more than five years) of digital information is an important issue?

2. To what extent does the sector recognise the importance of preserving digital information in the long-term?

3. What are the main problems associated with digital preservation in the _____ sector?

4. From what sources have you heard about the issues surrounding digital preservation?

5. What values does digital information have in the _____ sector beyond the original purposes for which it was created?

Understanding How Digital Preservation Affects Your Organisation

We would like to focus on how some of these digital preservation issues affect your own organisation

6. What type of information is digitally preserved in the short and the long term in your organisation?

7. What are the reasons that digital information is preserved in your organisation:
   - ❑ Legal requirements
   - ❑ Financial requirements
   - ❑ Business requirements (e.g. document important decisions and activities)
   - ❑ Historical value
   - ❑ Other (Please specify)

8. What risks is your organisation under if digital information is not preserved in the long-term?
   - ❑ Legal risks
   - ❑ Financial risks
   - ❑ Business risks
   - ❑ Historical value
   - ❑ Other (Please specify)

9. Has the organisation conducted a risk analysis and/or business needs analysis with regard to the preservation of information? If yes, can you indicate the main results?

Actions Taken: Policies, Strategies, Standards and Practices Developed

The questions in this section aim to explore some of the actions that the organisation has undertaken to deal with the preservation of electronic records. It will examine the above as well as selection, preservation, storage, and access activities.

Policies, Strategies, and Standards

10. Is there any collaborative effort across the _____ sector to tackle common digital preservation issues?
    - ❑ Conferences
    - ❑ Newsletters
    - ❑ Journals
    - ❑ Common Institutions
    - ❑ Collaborative Projects
    - ❑ Other (Please specify)

11. Has your organisation attempted to find information external to the sector regarding preservation?
    If yes, please indicate the sources
    - ❑ Government agencies
    - ❑ Higher education institutions
    - ❑ Archives
    - ❑ Libraries
    - ❑ Museums
    - ❑ IT Specialists
    - ❑ Other (Please specify)

    Please specify the kind of information provided and how useful it proved to be.
    _____
    _____
    _____
    _____

12. Do you cooperate with other institutions in the research and development of policies, strategies, and standards? In what way?

13. How useful is this common effort in applying it to your organisation's own needs?

14. Do you have any specific organisational policies that relate to the preservation of information?

15. Who (and what) was/is involved in the creation of these policies?
    - ❑ Management
    - ❑ Employees
    - ❑ Special task force in the organisation
    - ❑ Results of internal analyses (e.g. risk analysis)
    - ❑ External sources, models, advice
    - ❑ Other (Please specify)

16. Do these policies apply across the entire organisation?

17. How are these policies implemented?

18. Has your organisation developed preservation strategies, standards, and practices and implemented them?
    - ❑ Yes
    - ❑ No
    If YES, Please specify.

_____
_____
_____
_____

19. How were they introduced and implemented (e.g. by department, with training)?

20. How, and under whose responsibility have these been established?
   - ❏ External Advice/Sources/Models
   - ❏ Survey of information resources
   - ❏ In-house solutions developed
   - ❏ Other (Please specify)

21. How often are your preservation policies and strategies updated and renewed?

Selection of Digital Information for Preservation

22. Do you have a selection policy, or classification and retention policy that determines what information in your organisation is to be preserved?
   - ❏ Yes
   - ❏ No
   If YES, Please specify.

_____
_____
_____
_____

23. Is your classification and retention schedule linked and implemented across the organisation?

24. Who is responsible for the maintenance and implementation of these schedules?

25. How do you ensure that selected information is complete, accurate and identifiable?

Preservation of Digital Information

26. Does your organisation take care of its preservation activities itself, or are these outsourced?
   - ❏ Outsourced
   - ❏ In-house
   If outsourced, what reasons were behind this decision, and who carries out the preservation activities?

_____
_____
_____
_____

27. Are there specific individuals in your organisation responsible for the preservation of digital information?

28. What positions do these people hold in the organisation, and what are their responsibilities and competencies?

29. What type of training or advice is available for them?

30. Is your organisation aware of any external standards, best practices, and guidelines available on preservation?
   - ❏ Yes

    ❑   No
If YES, Please specify.

_____
_____
_____
_____

31. Are these specific to your sector?

32. Where did you learn about them? Please specify your sources.

33. Which of these standards, practices and guidelines do you use?

34. What technologies do you use for preservation? For the following list of current techniques, please specify which ones you use and for what kind of information.

Technique	**Specify Type/Technology Used**	**Information Preserved**
Print to Paper		
Scanning		
Save on Disk		
Save on Other Media		
Emulation		
Migration		
Microfilm/Microfiche		
Other		

35. On what grounds were these techniques chosen? Please specify your answers.
    ❑   External Advice
    ❑   Trials and Evaluations
    ❑   Recommendations
    ❑   Intra-sectoral standards available
    ❑   Other
Please provide as much information as possible about why these decisions were taken.

_____
_____
_____
_____

36. What data formats do you use for preservation?
    ❑   Standard data formats
    ❑   Others
Please specify for both answers

_____
_____
_____
_____

37. Do you convert the information to be preserved into other data formats for technical (or other) reasons?

38. What metadata do you use to describe both your digital information and the processes of storage and preservation? Does it follow any standards available (Dublin Core or others)? Can you provide a copy of the metadata set?

39. Is the collection and production of metadata automated?

40. Who is responsible for the transfer of information into long-term storage?

41. How often (if undertaken) does digital information migrated or refreshed?

Storage of Digital Information

42. Do you have a particular storage area for digital information to be preserved?
    - ❑ Yes
    - ❑ No

    If Yes, how is this organised and equipped?

    _____
    _____
    _____

43. Do you keep redundant copies of the digital information to be preserved for safety (or other reasons)?

Access to Digital Information

44. How is information protected from inadvertent or unauthorised access and manipulation?

45. Does your preservation solution allow direct access to the digital information stored (i.e. are they stored in an executable format)? If no, how is the access provided?

46. What access issues does your organisation face?
    a. Copyright
    b. Privacy Issues
    c. Access Security and Privileges
    d. Others (Please specify)

47. How does your organisation intend to provide access to digital information into the future?

Digital Preservation Costs

48. Did your organisation attempt to undertake a cost benefit analysis concerning its investments in preservation?

49. Has this analysis been assessed in light of your actual preservation activities? Did it prove to be accurate?

50. To which section of the budget are the economic resources for your preservation programme allocated?

51. What percentage of the organisation's budget is spent on preservation? Can you compare that to some other area of the organisation's activity?

52. Is the organisation attempting to address amortisation issues in the preservation budget?

53. Are there available sources of funding within the _____ sector allocated for digital preservation issues?
    - ❑ Yes
    - ❑ No

    If Yes, please specify

_____
_____
_____
_____

54. Are you satisfied with these cross-sector services?

55. If no, what would you like to see available? [i.e. what would you think could best be solved in common in your sector?] Would you be willing to engage financially in such information?

56. Are there other external sources available for digital preservation activities, (e.g. government grants, cross-sector funds)?
    ❑   Yes
    ❑   No
    If Yes, please specify
    _____
    _____
    _____
    _____

Monitoring of Actions

After having identified what has been undertaken in your organisation with regard to preservation activities, we would like to find out about how these efforts have been monitored.

57. Is the preservation process audited on a regular basis?

58. Is compliance to policies, standards, and strategies audited on a regular basis?

59. Is compliance to other requirements (legal, business etc.) audited on a regular basis?

60. How often are checks made to the preserved material, (e.g. for signs of deterioration)?

61. Please specify the criteria used for these audits.

62. Who performs these audits? (e.g. Internal/External)

Future Requirements

We would like to ask about the areas in which there is a need for additional attention in your organisation and the sector as a whole.

63. How long do you predict that your current preservation policies, strategies, and solutions will meet your organisation's preservation needs?

64. Is the amount of money allocated for preservation going to change in the future? Will it need to be changed?

65. If more funds were available, what could/would they be used for?

66. What conclusions has your organisation come to about its preservation efforts? Are these satisfactory?

67. What preservation efforts are remaining to be addressed within your organisation?
    ❑   Further data to be preserved

     ❑   Revision and adjustment of preservation policies and strategies
     ❑   Additional resources dedicated to preservation
     ❑   Technological solutions
     ❑   Other (Please specify)

68. Would you like to see more cross-sectoral or intra-sectoral activity with regard to preservation?

69. Are there any other areas in which you would like to have more information made available on digital information? Where do you expect this information to come from?

Thank you very much for your valuable contribution.

CONTACT DETAILS

ERPANET Coordinator

George Service House
11 University Gardens,
University of Glasgow
Glasgow, G12 8QQ,
Scotland

Tel: +44 141 330 4568
Fax: +44 141 330 3788

Coordinator@erpanet.org

ERPANET STAFF

d i r e c t o r s
Seamus Ross, Principal Director
Niklaus Bütikofer, Co-Director
Mariella Guercio, Co-Director
Hans Hofman, Co-Director

c o o r d i n a t o r
Peter McKinney

e d i t o r s
Andreas Aschenbrenner
Georg Büchler
Joy Davidson
Samir Musa

www.erpanet.org

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK