Mind the gap

Assessing digital preservation needs in the UK

Prepared for

DigitalPreservation**Coalition**

by

Martin Waller
Robert Sharpe

Tesella
Scientific software solutions

Digital Preservation Coalition, 2006

© 2006. Digital Preservation Coalition

First published 2006

About the Digital Preservation Coalition (DPC)
The DPC was established in 2001 and formally launched at the House of Commons in February 2002. It is a cross-sectoral membership organisation formed to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.

About Tessella
Tessella Support Services plc specialises in the application of innovative software solutions to scientific, technical and engineering problems. Tessella has over 20 years of proven expertise in the area of reliable and authentic long-term preservation of electronic records, both for government and scientific organisations.

Availability:
The report will also be freely available from the DPC website, www.dpconline.org

Foreword

This report is the culmination of almost three years of data gathering and preparation. The Digital Preservation Coalition identified a digital preservation needs assessment for the UK as a major priority for its work programme. In doing so, we wanted to achieve a number of specific objectives, but at a broad level we wanted to achieve two major goals:
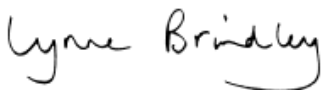
- Firstly, we wanted to move beyond "preaching to the converted" and reach a much wider audience than those who are already aware of the pressing issues involved in digital preservation.

- Secondly, we wanted to build a detailed picture of the current situation in the UK, which would enable the development of well informed strategic planning. In other words, we wanted to move beyond scare stories, beyond vague threats and provide concise but compelling arguments for engaging in digital preservation activities, based on evidence and objective analysis.

Since its launch in February 2002, the Digital Preservation Coalition has actively engaged in an advocacy and awareness campaign which has certainly contributed to a greater level of awareness of the issues surrounding digital preservation. What we felt was needed next was to investigate more precisely the extent to which this awareness had led to action.

- What are the major obstacles standing in the way of developing digital preservation programmes?

- What are the major incentives?

- Are there any sector-specific differences?

All of these issues need to be considered so that we can accelerate planning and preparation, informed by objective data analysis rather than conjecture and assumptions.

There is much to be done and we know there will never be enough funding to do everything. We believe that this report will help to ensure that scarce resources are used to best effect. It should act as a catalyst to ensure that the momentum which has already been achieved - to safeguard valuable digital resources in the UK - does not founder for lack of a clear roadmap.

Lynne Brindley
Chair, Digital Preservation Coalition

Contents

Report structure

This report is designed to be accessible to a variety of audiences. While it is also intended to be readable cover to cover, this section will help identify which sections are of most interest to an individual reader:

- Section 3 covers the methodology of gathering the information contained in this report.

- Section 4 discusses the drivers that cause organisations to store digital material for a long time and hence uncovers the nature and type of digital material that needs digital preservation.

- Section 5 discusses the current state of practice within the UK and reveals a gap between the needs of organisations and what is currently achieved. It demonstrates that many organisations face very similar problems as well as uncovering some sector-specific issues. However, it is not all bad news as it also highlights some organisations that have had some success at addressing and solving the problems that are involved which shows that progress is achievable.

- Section 6 discusses the gap in more detail and identifies 17 core UK Digital Preservation needs. Most of these apply to all organisations although some are sector-specific.

- Section 7 then summarises these needs and discusses the current state of progress in addressing these. In some cases the progress is very encouraging suggesting that it is possible to learn from the best practice of others, while in other cases recommendations for further actions are identified.

Executive Summary

The continuing pace of development in digital technologies opens up many exciting new opportunities in both our leisure time and professional lives. Business records, photographs, communications and research data are now all created and stored digitally. However, in many cases little thought has been given to how these computer files will be accessed in the future, even within the next decade or so. Even if the files themselves survive over time, the hardware and the software to make sense of them may not. As a result, 'digital preservation' is required to ensure ongoing, meaningful access to digital information for as long as it is required and for whatever legitimate purpose.

This report is the culmination of the UK Digital Preservation Needs Assessment (UKNA) study carried out for the Digital Preservation Coalition (DPC) to reveal the extent of the risk of loss or degradation to digital material held in the UK's public and private sectors. This has been an on-going exercise over three years with inputs from a wide variety of sectors including education, libraries, archives, museums, local and central government bodies, scientific research institutions, and organisations from the pharmaceutical, financial, manufacturing and engineering, media, energy, chemical, and publishing industries.

One of the main findings of the report is, not surprisingly, that the volume and total value of digital information needed in the 'information age' is increasing:

- Digital information is increasingly useful. Its preservation and reuse allows high quality search and retrieval regardless of geographical location, helps to improve the quality of decision-making and enables future analysis with fresh perspectives or new techniques.

- There are a number of other drivers to retaining digital information including complying with legal and regulatory obligations, protecting the investment of time and effort put into creating the information in the first place and addressing the risks of litigation and loss of prestige through not preserving records adequately.

- Failure to address the issue of digital preservation would generally result in the loss of corporate memory or of key cultural material, in many cases leading to financial loss.

However, organisations often do not have good solutions to the long-term preservation of data:

- The trend of increasing user demand for digital information is placing substantial pressure on the existing information infrastructure and working practices.

- Most organisations do not know the volume of digital information they need to preserve.

- Most organisations also have to preserve some digital material originating from outside of their organisation over which they have limited control.

- Many organisations also do not know how long they need to preserve information. The majority of those organisations that can specify a lifetime, need to preserve information for 50 years or more. Over such timescales, digital information is highly vulnerable to software and hardware obsolescence. Although many organisations have facilities in place that could preserve paper documents for this length of time, few have comparable strategies in place for digital material.

- The loss of digital data is commonplace, and in some circumstances seems to be accepted as an inevitable hazard.

In addition, there is considerable confusion about how to address the problem:

- Very few organisations have some kind of digital preservation strategy in place.

- Most organisations are hindered by a lack of clear responsibilities for digital preservation, a problem complicated by the need to involve a range of staff with different skills and the need to involve users at all stages of the information lifecycle.

- Digital preservation is very much a new discipline which is still being developed, and for which there are few people with the necessary skills. As a result it is a complex activity to undertake and is often perceived as risky.

- A successful digital preservation solution needs more than just management buy-in, it needs awareness and commitment at all levels throughout the organisation and often collaboration with other organisations.

- It can be hard to put together a strong cost-benefit justification because the main benefits are often intangible or are "public goods" distributed across time and a range of organisations. The common project-based funding model can hinder digital preservation activities as this often fails to place a value on assets that outlive a project and can be reused elsewhere.

As a result of analysing these findings, a series of key UK digital preservation needs have been identified. Evidence indicates that steady progress is being made toward meeting some of these needs, but in addition, it was found that some further actions are required.

Key recommendations for organisations:

- Organisations should create a long-term proactive information/knowledge management plan. This should start with an information audit and then, once the problem has been appropriately sized, the task of creating a solution should be properly resourced. This should include the assignment of clear responsibilities.

- Organisations should perform regular information audits to measure (and continue to measure) their digital preservation needs, and to ensure that these are being met.

- Organisations should consider the long-term value of digital material when putting together plans and budgets.

- All organisations need to encourage an international 'market' for digital preservation tools by linking up with other projects around the world and engaging with software vendors. This would deliver economies of scale and reduce risk for individual institutions.

- Organisations should consider the long-term preservation characteristics of the formats they use. They should work together and with software vendors to encourage the development of open file format standards.

Key recommendations for UK Government, funding bodies and regulatory bodies:

- Funding bodies should support research into the long-term value of digital information and models of how that value may change with time.

- The UK Government and funding bodies should promote, through seed funding, the creation of more digital archives across relevant sectors and organisations. Promote collaborative regional or national repositories to meet the needs of small and medium size organisations. These can then be promoted as exemplars for other organisations to follow. The overall aim should be to set up a network of trusted repositories.

- Promote wider awareness and use of relevant resources available from the DPC, the DCC and others.

- In formulating policies, the UK Government should take into account their impact on the digital information lifecycle. In particular, the sooner digital preservation activities can be carried out, the more effective and economical they can be.

- The UK Government should ensure that programmes and policies are co-ordinated nationally and on a wider scale (especially within the EU)..

- The UK Government needs to work with the digital preservation community and the legal profession to establish clearer guidelines for ensuring that digital information is legally admissible.

- Regulatory bodies need to work within their sectors and the digital preservation community to map out a framework, which will allow organisations to store information in an exploitable form while retaining the ability to satisfy regulatory concerns.

1 Threats and opportunities

The digital revolution continues to open up many exciting new opportunities in both our professional and leisure lives. We rely more and more on computers to digitally store our documents, our photographs and our data. In our headlong rush to 'go digital', little thought is given to how we will access these computer files in the future. There is a real danger that much important digital information will be lost through lack of awareness and poor planning for the future.

Our increasing reliance on digital information brings many new opportunities, and we must be prepared to rise to new challenges.

Case study: Viking Lander data

When the US space agency NASA sent two Viking Landers to Mars in 1975 to find out whether life might exist on the red planet, it was assumed that the datasets painstakingly compiled by scientists at the time would be available for future generations of scientists on magnetic tape.

Yet, just a few decades later, despite the space agency's best efforts to keep the tapes in a climate-controlled environment, time has left them cracking and brittle. Furthermore, when scientists attempted to re-use some of the data in the late 1990s, they found that they could not decode the formats used. In the end they had to track down old printouts and retype everything.[1]

Courtesy NASA/JPL-Caltech

The preservation of paper documents is well understood and is supported by an entire professional infrastructure of librarians, archivists, conservators, etc. The preservation of electronic documents, in contrast, is a relatively new discipline with very real challenges that are often poorly understood. The rapid pace of technological change, the continual deterioration of electronic media and the difference in perspective between archivists and IT staff mean that preservation of digital data is rarely carried out in a methodical and sustainable manner.

What makes long-term storage of digital information difficult?

- **Deterioration of the storage medium:** tapes, disks and even CDs and DVDs have a finite lifespan (manufacturer claims on CD lifespans can be as low as 20 years),
- **Obsolescence of the storage medium** (e.g. 5¼" disks, though commonplace 20 years ago are now obsolete, and most new machines no longer have a 3.5" floppy disk drive),
- **Obsolescence of the software** used to view or to analyse them (e.g. WordStar cannot easily be run on modern machines),
- **Obsolescence of the hardware** required to run the software (e.g. software to view Domesday project only ran on BBC Micro computers),
- **Failure to document the format adequately** (A common example is preserving tables of numbers without preserving an explanation of their meaning) ,
- **Long-term management:** the need for continuous active management over long time periods and the high cumulative risk of organisational or administrative failure.

---

[1] See www.cbsnews.com/stories/2003/01/21/tech/main537308.shtml (last accessed on 03/01/2006)

Digital preservation is more than just records management. It is not sufficient to simply store the bitstream. The meaning and format of the content along with any associated functionality must also be preserved. Much research has been carried out over recent years to identify the best strategies for long-term retention of digital materials, and good sources of advice and guidance exist.

"Digital capture, storage and copying devices (plus media) are likely to rapidly become more capable and cheaper. From a positive perspective this will improve the opportunities for the origination of high quality digital images/facsimiles, but alternatively can only accelerate the cycle of obsolescence"

Source: MLA survey

Experience has shown that it is considerably more efficient to address the preservation issues for digital information as early as possible rather than to make up for deficiencies at a later stage. For example, assessing the long-term importance of digital information and recording details about its format can often be done relatively easily at the point of creation, whereas it can be a costly exercise requiring significant input from domain experts at a later stage. Hardware, software and media obsolescence issues are more effectively dealt with before obsolescence occurs, as illustrated by the considerable effort required to rescue the 1986 Domesday Project (see panel below).

Digital preservation does not only affect the world of academia and research; it is also important that today's businesses respond appropriately. In a recent judgement in the US, Morgan Stanley had more than $1 billion awarded against them as a result of their failure to preserve and hand over certain documents required by the courts. The Securities and Exchange Commission in the US are also considering fining the same bank over $10 million – specifically for failing to preserve email documents.[2]

Case Study: The 1986 Domesday project

Copyright The National Archives, Catalogue reference: E 31/2/2 f.238a

The original Domesday book, which was produced on vellum in 1086 for William the Conqueror, is still readable today at The National Archives, over 900 years later.

To mark the 900th anniversary of the original Domesday book, the BBC launched the 1986 Domesday project. This consisted of thousands of photographs and maps combined with statistical, written and visual information to produce a picture of Britain in 1986.

The project was recorded onto 30cm laserdiscs and could be viewed with software running on BBC Microcomputers. Less than 20 years later, not only are the 30cm laserdiscs obsolete, but also the necessary hardware. Separate projects were launched by The National Archives and by Leeds University to rescue the data and the software that made up the 1986 Domesday project. This was only achieved thanks to a surviving laserdisc player and more than a year's effort by specialist teams.[3]

The threat of litigation is only one of many reasons for preserving digital information. Well preserved and catalogued digital data is a valuable asset opening up many new opportunities such as faster and more widely available access, more powerful search capabilities, more flexible data mining and generally greater re-use opportunities. For many organisations, digital information represents a considerable investment of time and effort. Preserving the information for future access allows the greatest benefit to be derived from this investment.

---

[2] For more on Morgan Stanley case, see www.cfo.com/article.cfm/4340197/c_4340354?f=archives&origin=archive (last accessed on 03/01/2006)

[3] For more details on the operation to rescue the Domesday Project data, see www.ariadne.ac.uk/issue36/tna/ (last accessed on 03/01/2006).

2   Context

2.1   What is Digital Preservation?

Throughout this report, we have used the term 'digital preservation' to refer to all the actions required to ensure ongoing, meaningful access to digital information for as long as it is required and for whatever legitimate purpose.

2.2   When should Digital Preservation be performed?

Digital preservation differs from 'traditional preservation' in that digital information needs to be actively and continuously managed – it cannot be simply stored and left on a shelf. In addition, the timeframe during which action needs to be taken is short. Problems can arise in as little as five years and certainly more than twenty years would be highly likely to cause difficulties if information has not been properly managed during that time. When we talk about 'long-term' we are therefore referring to anything needing to be kept for more than twenty years, a relatively brief period for print materials.

Because of obsolescence and the impermanence of digital media, digital preservation should start as early in the lifecycle of a digital resource as possible, preferably when actually planning to create digital information. Decisions made at that time, regarding formats and essential documentation, will have major implications for the ability to retain those materials for even the short term, let alone medium to long-term. Case studies in this report provide ample evidence of the ease and speed with which effective access can be lost if digital information is not appropriately and actively managed.

The sense of urgency in progressing digital preservation activity, is heightened by knowledge of the ongoing trends in the output of and user demand for digital information and the consequent increasing volumes requiring digital preservation. These trends are so dramatic that they are putting and will continue to put substantial pressure on the existing information infrastructure and working practices. We need to find solutions for digital preservation as soon as possible to have sufficient expertise and practice to apply them to the information deluge, which is approaching over the next decade.

2.3   Who should care about digital preservation?

Digital preservation has implications for a much wider group of participants than is the case with printed materials. It is not a case of leaving it to someone else as that will inevitably mean that material will be effectively lost. The short timeframe during which action needs to be taken means that, while it may be *technically* possible to recover digital information (as in the example of the Domesday project), it is not *practically* feasible to go to this much trouble for every digital resource.

Individuals and organisations creating digital information, whether through digitising non-digital materials or creating 'born digital' resources, will need to take steps to ensure their material is safely stored and has appropriate documentation so that its purpose and context are clear enough to enable others to manage the information over time. Organisational corporate records originate almost exclusively in digital form and there may be legal or efficiency drivers to retain them in digital form and to ensure that they remain accessible and authentic. Researchers and scholars rely increasingly on digital sources of information and are themselves major creators of digital resources.

Digital preservation therefore impacts on a very broad and diverse range of stakeholders, including many who would not have regarded themselves as having a role in traditional preservation.

3 Background and methodology

Previous studies undertaken by the Digital Preservation Coalition (DPC) and its members include a DPC Members' survey in 2003[4], an investigation of real-life scenarios of circumstances in which digital materials become vulnerable to loss, and in 2004-2005 a sample survey of local and regional organisations funded by the Museums, Archives and Libraries Council (MLA)[5]. These investigations comprised the first two stages of the DPC's UK Needs Assessment (UKNA). This report is the culmination of the two earlier surveys, and a more detailed, wider survey undertaken in 2005.

3.1 Objectives of the study

In August 2005, the DPC commissioned Tessella to carry out the latest part of the UK Digital Preservation Needs Assessment (UKNA) study with the following objectives:

- Raise the profile of digital preservation in the UK.
- Articulate what is required in terms of infrastructure (i.e. storage capacities, training etc.).
- Provide persuasive arguments in favour of digital preservation.
- Influence data creators, who have a key role in the longer-term feasibility of preserving digital resources.
- Engage senior policy and decision makers likely to be involved in approving funding for digital preservation activities.
- Identify and create a prioritised list of key priorities for further action.
- Draw attention to the important work that has already taken place.
- Identify, list and prioritise specific challenges within identified sectors (libraries, archives, research, etc.).
- Identify the digital preservation issues of common concern across all sectors.
- Articulate the risks associated with *not* addressing issues relating to digital preservation in the UK.
- Align the recommendations with appropriate UK government agenda and policies and also relevant EU activity.

3.2 Methodology

The latest part of the UKNA study draws information from a number of sources:

- Existing sources of data on digital preservation activity in the UK were analysed and synthesised.
- Additional up to date and quantitative information was gathered via an online survey run by Tessella.
- Further qualitative information was gathered via a series of interviews of respondents with different responsibilities and interests around digital preservation and from a wide range of organisations.

The online survey was in the form of a questionnaire carried out between September and November 2005. Invitations to fill in the questionnaire were sent to over 900 individuals from a wide range of organisations in different sectors. The selected individuals all had an assumed interest in digital preservation as part of their professional responsibilities, and included a range of roles including records managers, archivists, librarians, but also IT managers and data producers. 104 responses were received, giving a good response rate of over 10%. These included respondents from education, libraries, archives, museums, local and central government bodies, scientific research institutions, and from organisations in the pharmaceutical, financial, manufacturing and engineering, media, energy and chemical, and publishing sectors.

---

[4] DPC Members' survey can be viewed at www.dpconline.org/graphics/reports/index.html (last accessed on 03/01/2006)

[5] MLA regional study can be viewed at www.mla.gov.uk/resources/assets//M/mla_dpc_survey_pdf_6636.pdf (last accessed on 26/01/2006)

4 Why is digital preservation important?

Today, much information is delivered only in digital form, and increasingly via the web. Digital communications are the norm. In the government sector, for instance, the Modernising Government Agenda has set a target for all dealings with UK

"Digital preservation allows long term access to the lifeblood of the organisation"

Source: 2005 DPC survey

Government to be deliverable electronically by 2008. In the private sector, it is has been reported that 70% of UK companies now use e-mails for contract negotiations, HR letters and financial transactions[6]. Without suitable preservation and access planning, such information would become inaccessible after relatively short timescales.

Respondents to the 2005 DPC survey felt a strong sense of urgency:

- 87% recognised that a failure to address the issue of digital preservation would lead to loss of corporate memory or of key cultural material,

- Over 60% felt that their organisation could lose out financially – either through loss of income or through increased operational costs.

In contrast, fewer than 20% of the organisations surveyed have some kind of digital preservation strategy in place.

4.1 Legal requirements

Our legal framework requires organisations to preserve adequate records of business processes, communications and many other types of data.

Some commonly cited legal drivers:

- The **Freedom of Information Act 2000** is a major legal driver in the public sector. It applies to all information held by public authorities, including electronic records. It came into force at the start of 2005 (replacing the '30 year rule'). It implies both preservation of electronic records and some degree of record curation so that records can be made available when requested.

- The **Data Protection Act 1998** applies to all information about individuals, including electronic records. It requires that the information is accurate and up-to-date, secure and kept no longer than is necessary for the original purpose for which it was collected. The greatest impact of the Data Protection Act is to ensure that retention periods for electronic records are agreed and implemented consistently.

- Recently, the **Sarbanes-Oxley Act of 2002** has been introduced in the US as a result of a series of high profile scandals. Its purpose is to ensure corporate accountability, and it requires a structured approach to retaining records (including e-mails) and an ability to access them in a way that preserves meaning. All companies with a significant US shareholder base must comply with the provisions of this legislation.

- Public authorities and bodies carrying out a public function are subject to the **Environmental Information Regulations 2004**, whereby they must respond to any requests for access to environmental information that they hold. This has implications for preservation and for curation of digital records. The Environmental Information Regulations bring into UK law the UNECE Århus Convention on Access to Environmental Information.

---

[6] Survey by Industry Association AIIM Europe, www.aiim.org/article-pr.asp?ID=29428 (last accessed 27/01/2006)

The importance of the legal framework for digital preservation was recognised by 84% of the respondents to the 2005 DPC survey, who indicated that there were legal drivers for digital preservation in their organisation. Indeed, 73% recognised that their organisation would fail to comply with legal requirements if it failed to preserve its digital information adequately.

4.2   Accountability

The UK government has long accepted the importance of preserving digital records to ensure that it is accountable. This is now enshrined in law through the Freedom of Information Act 2000 and is one of the main drivers for The National Archives' digital preservation programme[7].

Accountability is also a key motivation for business, partly as a result of recent regulations (e.g. the Sarbanes-Oxley Act), but also in order to protect themselves from litigation. Recent high profile cases, such as the Morgan Stanley examples cited in section 1 of this document, illustrate the importance of being able to search and recover archived emails in a prompt and legally admissible manner. The importance of digital material was highlighted by the previously quoted AIIM survey, which revealed that over 70% of UK companies use e-mails for contract negotiations, HR letters and financial transactions.

Despite the direct penalties that can be incurred through litigation, for many organisations it is the fear of damage to reputation as a result of loss of data that can be the more significant driver.

4.3   Protecting the long-term view

In the business world, long-term preservation of access to digital data is important for ensuring business continuity and for supporting decision making with a long-term view. In the cultural world, many assets are now in digital form and it will therefore be necessary to find ways to preserve these for future generations. In research, preserving data may be crucial for identifying long-term trends (e.g. meteorological observations).

"No real consideration seems to have been given to digital preservation beyond the storage of the data"

Source: 2005 DPC survey

The 2005 DPC survey found that 81% of the respondents who were able to specify a lifetime for their digital information had to keep at least some of it for more than 50 years, a timescale which makes it highly vulnerable to software and hardware obsolescence. Although many organisations have facilities in place that could preserve paper documents for this length of time, few have comparable strategies in place for digital materials.

4.4   Protecting investment

Many organisations possess valuable intellectual assets, which are supported fully or partially by digital records. 64% of respondents to the 2005 DPC survey preserve digital data to protect Intellectual Property (IP) and 22% do so to support patent applications. In many cases it is important to keep both a record of the asset itself, and also a full audit trail of how it was derived.

Even if there are no formal IP requirements for preserving digital data, much data represents a considerable investment of time, effort and money, and it would therefore be foolish not to preserve it adequately.

4.5   Enabling future re-use opportunities

We are shifting from an industrial economy to a knowledge economy, where repositories of digital information and the tools to mine, analyse and re-purpose them represent a society's intellectual capital. Effective and affordable digital preservation strategies and systems will transform archives into valuable assets.

---

[7] For more details on The National Archives' digital preservation programme, see www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm (last accessed on 03/01/2006).

In academic research, data re-use is of considerable importance.

- It allows subsequent research to be carried out without the time and cost overheads of re-creating similar data.

- It allows data from a number of different sources to be combined to give a wider data set.

- It allows teaching to be carried out based on real data sets.

- It allows greater interdisciplinary and cross-disciplinary interaction.

- It also allows future re-analysis of data with alternative or improved techniques.

4.6   Lost information is lost forever

For some digital data, it is the uniqueness of the event that it records which makes it particularly valuable for future retention. Many types of historic scientific data, such as meteorological measurements or raw measurements taken from a volcano eruption or an archaeological dig, are not repeatable and can be of great value. Similarly, records of news events (e.g. mobile phone pictures of the 2004 tsunami) cannot be repeated. Although the data may have been exploited at the time of generation, there are many ways in which it may be usefully re-used in the future.

4.7   User expectations in the information age

A suitably designed digital repository can expose a vast amount of data through simple user-friendly interfaces such as the text-based searching that users have come to expect in the "google age". Such ease of access considerably enhances the value of preserved data by making it easier for potential users to be aware of it, to access it and to reuse it.

Over 80% of respondents to the 2005 DPC survey recognised that their organisation would benefit from the improved access to information that could be brought about by having a suitably catalogued and searchable digital repository.

"There is an expectation that searching should be immediate (probably based on experience of Google). A few years back, visiting academics might have started to plan their trip a year in advance and be willing to invest considerable effort in finding material when they arrive. Now there is an expectation to be able to search remotely and a tendency to complain if the search takes longer than a minute."

Source: 2005 DPC survey

4.8   Business efficiency

A full transition to digital business processes can bring substantial gains in efficiency and quality. It can allow information to be accessed quickly and easily from any location, it can support risk management and business continuity planning and it can optimise the legal admissibility of records. Digital business processes are often seen as a driver for establishing an Electronic Document and Record Management System (EDRMS). This does not, however, resolve the full digital preservation challenge. Sustaining digital business processes over the long-term is only possible if they are supported by suitable digital preservation activities in order to keep accurate records of activities and make them available for re-use.

This has clearly not yet occurred in the mainstream as, for example, 50% of organisations which responded to the 2005 DPC survey still print out documents in order to preserve the hard copy, thereby making it harder to copy, edit or search the documents, to distribute them widely or to build new business intelligence through automatic processing and analysis.

Digital processes can be especially important in R&D where speed of development can be a major factor in determining the success or failure of a project. In the pharmaceutical industry, for example, the time it takes to get a drug to market is a critical factor affecting profitability. Any opportunities to speed up this process, either by improving the efficiency of business processes or by making use of results from previous experiments or trials, could bring considerable benefits.

Should we preserve digital or paper copies of important documents?

Storage of printed copies of important documents is generally accepted as a reasonably fail-safe method of preservation. The dangers are well understood and can generally be mitigated.

In contrast, preserving key documents in electronic form over the longer term can be perceived as more problematic and risky because our experience of it is relatively new and practice still evolving. As a result, it is not uncommon for organisations to bypass the problems by printing out digital documents and archiving the paper copies.

In some circumstances this approach can be a justifiable and pragmatic one. It does, however, have substantial drawbacks in terms of cost and lost opportunities. Finding, accessing, copying, editing, transferring and re-using the digital file are all considerably easier than they are for a paper document. Shelf space restrictions and the time and effort involved in printing and filing a document can also be strong arguments against this approach.

5   What is the current state of practice?

The increasing coverage of digital preservation issues in the press nationally and internationally demonstrates a growing awareness of the problems and risks. But does this mean that organisations and businesses are now resolving the problems?

5.1   A commitment to digital preservation?

The 2005 DPC survey found that there was a management commitment to digital preservation in 52% of cases. This, however, only translated in 33% of cases to a well-defined responsibility structure for digital preservation. Only in 20% of cases was there adequate funding and in 18% of cases was there a digital preservation strategy in place (see Table 1).

	Yes	No	Don't know
Is there a high level commitment to digital preservation in your organisation?	52%	33%	16%
Is there adequate funding allocated to digital preservation in your organisation?	20%	54%	25%
Are there clear responsibilities for digital preservation in your organisation?	33%	55%	12%
Does your organisation have a digital preservation strategy, or any other policy or strategy document which specifically refers to digital preservation?	18%	77%[8]	5%

Table 1: Commitment to digital preservation *(source: 2005 DPC survey)*

The earlier survey of DPC members showed a similar pattern in that, despite a clear commitment within the organisations to digital preservation, this was not, in the majority of cases, adequately embedded within their formal statements and procedures. This was later confirmed by the MLA regional survey in which just 10 out of the 26 organisations surveyed had a corporate planning document referring to digital preservation and only 6 out of 26 had a digital preservation policy.

Another survey carried out by industry association AIIM Europe, which represents the information management community, revealed that less than 45% of companies have policies on the length of time e-mails are retained and how or by whom they are stored.

Most organisations are at the early stages of thinking about digital preservation. There is a growing awareness and commitment to it and this is gradually trickling through to creating practical strategies. The MLA regional survey found that several of the surveyed organisations that did not have a digital preservation strategy in place were in the process of putting one together. Indeed, the 2005 DPC survey found this to be true for a large proportion of organisations (43%) – though further interviews revealed that in practice this tended to be in the very early stages.

"There is a high level feeling that [something] should be done, but with no practical action. The lack of action seems to be due mostly to not knowing how to proceed."

Source: 2005 DPC survey

---

[8] Includes 43% who replied that a strategy was being created.

5.2   Organisational structure and responsibilities

Another striking result from the 2005 DPC survey is the impact of the lack of clarity in responsibilities for digital preservation. Table 1 shows that 55% of respondents felt their organisation does not have clear responsibilities for digital preservation. The survey also revealed that 53%, almost the same proportion, felt that this was a barrier to their digital preservation programme.

In some organisations there is simply a gap in responsibilities and no-one has the formal responsibility for preserving and curating digital material – should the data creators look after their own data? Should the archivists who traditionally look after paper records extend their remit to digital materials? Should the IT staff who already look after the digital systems extend their remit to cover digital data?

In other organisations, the problem is more complex and is characterised by insufficiently defined boundaries between the different roles involved in digital preservation activities. This can lead to differing expectations and to tensions. IT staff may feel that a suitable backup strategy is sufficient for long-term preservation, archivists may fail to understand some of the technological issues, and data creators may simply fail to recognise the importance of making their data available to others

As a result, it is necessary to have a coordinated cross-disciplinary team which takes into account the differing roles of data creators, record managers, archivists, IT staff and any other stakeholders. This in turn relies on having sufficient management buy-in to make it a viable option.

Finally digital preservation is a long-term activity and over time responsibility for care of digital materials may pass between organisations. Uncertainties during the handover of responsibilities and different interests over time and between organisations may impinge on a clear definition of roles and responsibilities.

"No overall corporate information manager. No real understanding of records, archives and digital preservation with IT staff."

"Perceived split between IT specialists and records management (how vs why)"

"Ordinary Joe IT managers have to see that they can do something for their companies and for themselves by getting interested in digital preservation"

"There is currently no designated responsibility for digital preservation, though the organization is aware that the problem exists"

Source: 2005 DPC survey

5.3   Volume and lifetime of information

Evidence was found in all three UKNA surveys that a large proportion of organisations have not assessed the volumes of material they need to preserve. In particular, this was true of over 33% of respondents to the DPC members' survey, and of 55% of the respondents to the 2005 DPC survey.

Where respondents were able to provide estimates, the latest survey revealed large volumes of data to be preserved. These were concentrated around the 1-10TB range (see Figure 1).

The 2005 survey showed that 49% of respondents did not know of any defined lifetimes for their main type of digital data. Of those that did know the lifetime, 81% need to keep their main type of information for longer than 50 years. In particular, archives and libraries often aim to keep digital material indefinitely. There are also cases in the commercial sector where lengthy retention schedules are required. For example, pharmaceuticals
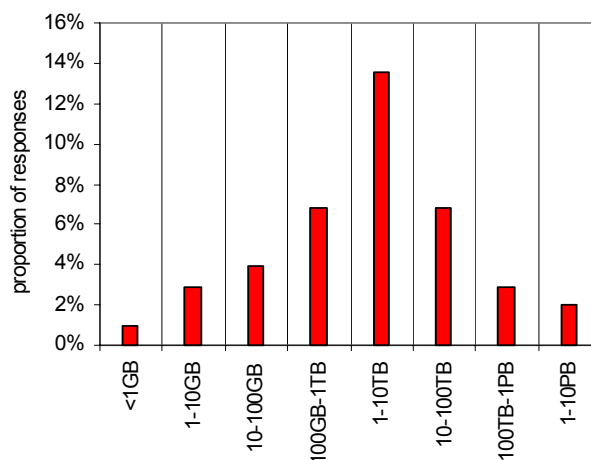
Figure 1: Volume of data to be preserved
(Source: 2005 DPC survey)

and manufacturing and engineering companies need to keep records for periods of time exceeding the lifetime of the drug or product they relate to. In the financial sector, organisations must keep financial records for the lifetime of the customer account plus a further seven years.

These timescales are long enough for substantial evolutions in software and hardware technologies to occur. These organisations must therefore have adequate strategies and adequate solutions in place for preserving this material. Given the volumes involved, relying on resolving the problem manually at the time of access to the data (as in the earlier Domesday project and Viking Lander examples) would be an extremely costly and impractical, if not impossible, task.

"The digital material that has survived from the past is mainly ad-hoc and some is stuck on backups of old systems where, for example, DAT tapes may exist but the effort required to find the equipment needed to read them and then to work out what records are is too large to be practical: which means that, in practice, the records are lost."

Source: 2005 DPC survey

5.4   A strong dependence on digital information

The large volumes of digital data for preservation reported in the section 5.3 would have been unthinkable a few years ago. Our growing dependence on digital information is also demonstrated

"Acceptance of the importance of digital resources is generally taken for granted."

Source: 2005 DPC survey

- by the frequency with which digital communications are used for contractual negotiations,

- by the increased use of online publications (many of which no longer exist in paper form),

- by the growing reliance on digital information for patent applications and for preserving intellectual property,

- by our growing reliance on the internet as a source of information,

- by the growing popularity of digital communications in all walks of life.

Failure to take steps to address the issue of digital preservation could lead to irrevocable loss of much of this material.

5.5   Data loss

Given that many organisations do not know the extent of the problem they face, it is not surprising that the loss of digital data is commonplace, and in some circumstances seems to be accepted as an inevitable hazard. Only 29% of respondents in the 2005 DPC survey could say they have not lost access to some digital information, as a result of it being impossible or too expensive to recover. Even when referring to their most important type of data, this proportion only rose to 43% (see Table 2).

"At my archives, we are now receiving information in digital form almost every week. In endless sizes and formats, up to 15 to 20 years old. Most of it is impossible to read for us today and some of it has permanent value, but it is not available on paper, so it is really lost. It is a great and GRAVE problem how we can be sure to preserve our information today for the future. This is not a problem for the future but for today."

Source: *www.findarticles.com/p/articles/mi_qa3937/is_200007/ai_n8917888*

	Yes	No	Don't know	Don't want to answer
Does your organisation have any inaccessible data?	36%	29%	31%	4%
Has any of your main type of data been lost?	28%	43%	29%	
Any data in danger of becoming inaccessible?	48%	21%	27%	4%
Are any file formats that have been used now obsolete?	38%	35%	27%	

Table 2: Loss of digital data *(source: 2005 DPC survey)*

31% of respondents did not know whether their organisation had lost any digital data (see Table 2). This suggests a lack of confidence in their digital processes, and that there is not enough proactive management of digital data.

Equally concerning results were found by a survey carried out by industry association AIIM Europe. It suggests that 40% of UK companies are worried that their electronic records may not be accurate, accessible and trustworthy.

"We are currently cataloguing a collection of personal papers amongst which is a box of about forty 3" Amstrad disks. I know that there are companies which will migrate this data to newer formats, but we currently have no way of knowing what is on these disks, whether it is material replicated elsewhere in the collection, and whether migration will be worth the cost."

Source: mailgroup posting

In some cases the data that has been lost is of little consequence. However, the 2005 DPC survey revealed some more significant examples of loss including research data, business records, and even on-going drug submissions that are relying on paper or microfiche copies of the original digital materials which have themselves been lost. The main reasons given for loss of data are:

- Obsolete file formats / software / media
- Insufficient metadata
- Corrupted files on portable media
- Uncontrolled number of file formats
- Insufficiently documented proprietary file formats
- Inaccessible data at point of donation to archive
- Software updates or emulations not fully compatible with data
- Data physically lost

A risk-based (or value-based) approach:

Initial regulations set out by the US Food and Drug Administration for preserving digital records in the pharmaceutical industry, were found to be too onerous to comply with in practice. Implementation would, amongst other things, "significantly increase the cost of compliance" and "discourage innovation ... without providing a significant public health benefit".

As a result new guidelines were issued introducing a risk-based approach to digital preservation. This approach involves balancing investment against the value of the records. In the case of the FDA guidelines, the value is evaluated in terms of the risk to public safety, but in other cases it could be, for example, based on the likelihood of re-use and/or the potential benefit.

Few resources are targeted at the low value records. These may simply be left as the original bitstream, in the knowledge that with some effort the original data will be recoverable if necessary. For mid-value records, simple migration tools would be used to ensure that the data remains accessible – though possibly with some loss of information or loss of functionality. At the high value end of the scale, any migration or emulation exercise will need to be accompanied by a verification that no loss of data has occurred.

5.6 Origin of information to be preserved

The complexity of digital preservation increases considerably when the preserving organisation has to accept files from outside the organisation, which can be in a range of unexpected formats. Figure 2 looks at how prevalent this problem is.

The 2005 DPC survey showed that 84% of organisations have to preserve digital files created outside the organisation. The majority of these organisations have little or no influence over the formats of these files.

This was particularly highlighted in the responses received from the government sectors and from university libraries where more than 80% of respondents not only acquire data from outside the organisation, but are also aware of having no more than a limited influence over the formats that this data comes in. This means they have to cope with unexpected file formats, and as a result, with unpredictable technical difficulties that can only be resolved with human intervention. This in turn leads to increased costs.

"As a local authority records office, we receive material from a very wide range of organisations and private depositors, sometimes in salvage conditions. The resulting wide range of formats and potential lack of reliable metadata create a potentially enormous task, and may result in it being pointless trying to preserve records which could potentially be historically very important."

Source: 2005 DPC survey
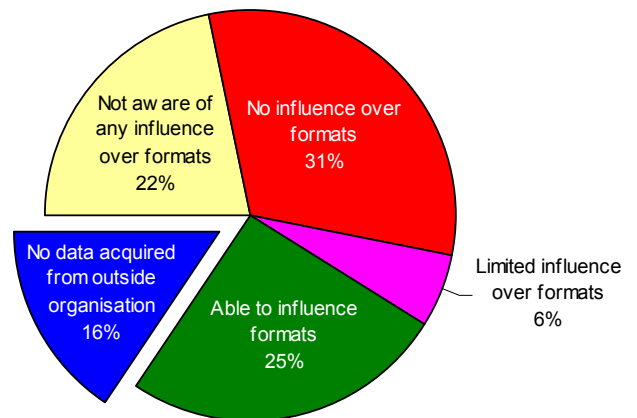
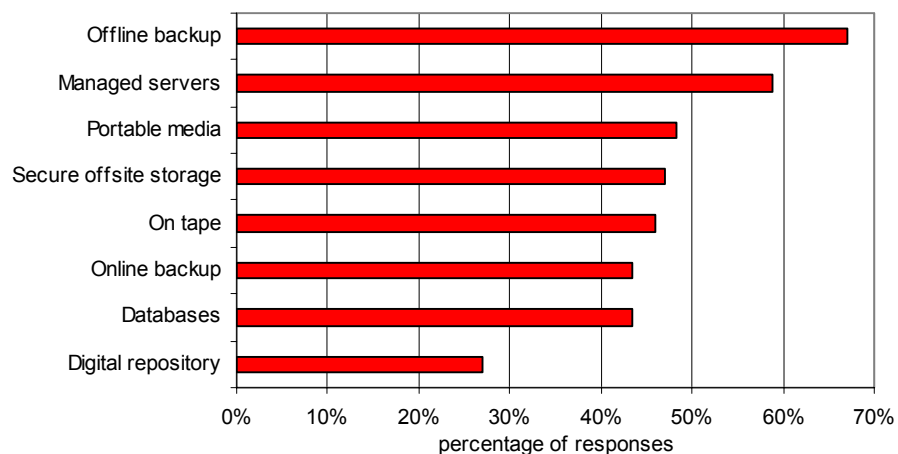**Figure 2:   Influence over formats of externally sourced data** *(Source: 2005 DPC survey)*

5.7 The limitations of current solutions

Long-term storage of data

The 2005 DPC survey asked about approaches taken by organisations to address the issue of long-term storage of digital data. The main responses are given in Figure 3.

Figure 3: Current long-term digital storage solutions *(Source: 2005 DPC survey)*

For the vast majority of organisations, a combination of solutions is used. For example, offline backups may be used for data that must be preserved but is only rarely accessed. This can be taken a step further by using secure offsite storage to ensure that the backup media themselves are protected in a separate location. Online backups, managed servers, databases, or digital repositories allow continued access to the data.

However, 48% of respondents said their organisation uses portable media (e.g. CDs, DVDs, floppies) as a means of long-term storage. In the MLA regional survey 22 out of 26 organisations used CDs for storage of data. This is rarely the exclusive method used, but it is still a cause for concern given the unreliable and unpredictable nature of portable media for more than a limited number of years.


Why not use CDs/DVDs for digital preservation?

The UKNA study found that a high proportion of organisations use portable media for long-term storage of digital records. Although this might seem like a pragmatic approach, it is in fact fraught with dangers. Not least is the danger of obsolescence of the medium itself as technologies and requirements evolve (e.g. 5¼" floppy disks).

CDs are widely quoted as having lifespans of up to 20 years. The reality is that, in most cases, they will be stored under less than ideal conditions. Light, dust, moisture, normal handling can all take their toll. As a result, CDs and DVDs can be inherently unreliable as a medium for preservation over considerably shorter timescales than this.

Portable media are by their very nature an offline solution. This has considerable disadvantages when compared with an online digital preservation solution. In particular, data access and searchability are very restricted, and any active management of the data (e.g. media and format migration) is a very slow, resource-intensive exercise.

Current strategies for digital preservation

The 2005 DPC survey also asked about organisations' current strategies for digital preservation. The main approaches given are listed in Figure 4.
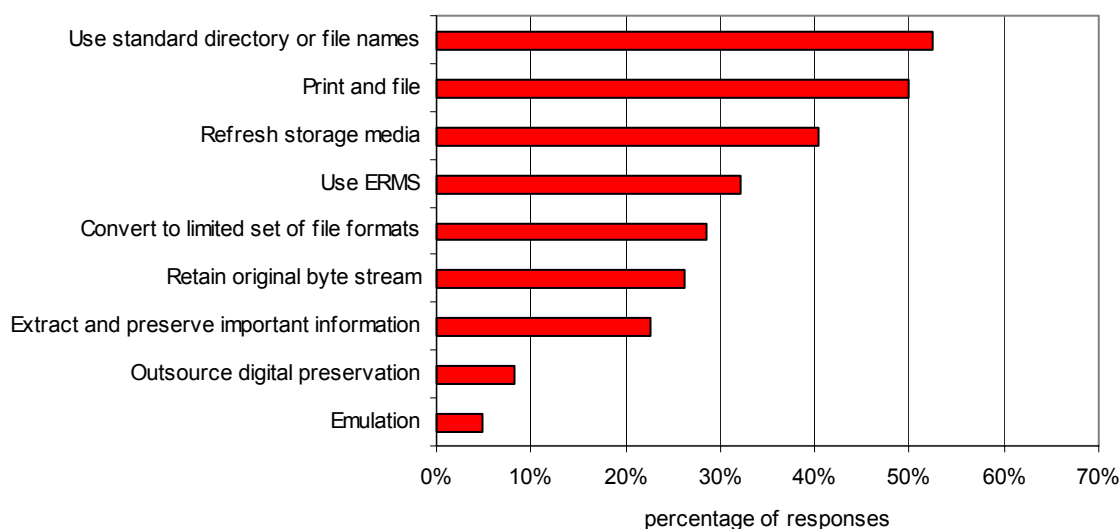
Figure 4: Current approaches to digital preservation *(Source: 2005 DPC survey)*

The two most widespread digital preservation strategies – to use standard directory structures or file names (52%) and to print out digital data and file the paper copies (50%) – may be pragmatic approaches to preserving digital data, but both have severe limitations.

- Using standard directory structures and file names does not begin to address issues of obsolescence of software or of formats. Furthermore, this approach only allows a single organisational structure for the data, and does not promote flexible searching capabilities.

- Printing out and filing the paper copy (50% of responses) was, in over 10% of cases, the exclusive method used for long-term preservation of digital information. This gets around problems of obsolescence, but it does have some significant drawbacks from the point of view of access (e.g. geographical restrictions to access, searchability and re-use limitations), and from the point of view of efficiency (e.g. shelf space limitations, and time and effort spent physically printing and filing).

The popularity of these strategies highlights that digital preservation as a discipline is still in its infancy and many organisations feel they do not have the necessary expertise to manage their own digital preservation activities, but instead adopt a pragmatic, non-ideal solution.

Using metadata for long-term access

Metadata is information describing a digital item. It should include details of the file format, the context and the content of the item. The importance of metadata in a digital archive cannot be overestimated. It adds a whole new dimension, supporting the discovery of archived information, providing an extra richness of contextual or descriptive information at the point of access, and providing technical information to support the active management of digital materials over time.

The creation of metadata can be relatively straightforward as long as it is done at the optimum time – usually as early as possible in the resource's lifecycle – hence the need for digital preservation planning as early as possible in the data creation process. However, only 31% of respondents to the 2005 DPC survey felt that sufficient metadata was supplied at the time of acquisition by the archive. In all other cases, either insufficient metadata gets stored or an exercise to enhance the metadata is required – with the resulting overheads and quite possibly the need for involving domain-specific experts.
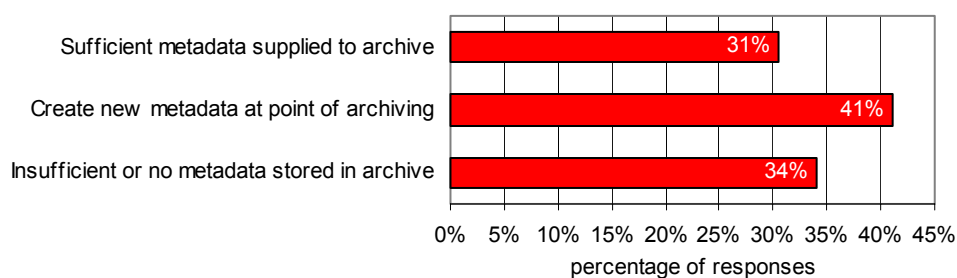

**Figure 4:	Creation of metadata** *(source: 2005 DPC survey)*

Case study: Schools Census

The Schools Census is a survey of schools in England and Wales, believed to have been started in about 1946, and first recorded on computer media around 1975. The data is held by the National Digital Archive of Datasets (NDAD) on behalf of The National Archives.

Although the data itself had been carefully preserved, some of the earlier metadata was lost during a migration exercise in 1991-92. As a result the meanings of the columns of data were no longer clear.

Some 'digital archaeology' was carried out to recover the early metadata. With the help of two County record offices and a few schools, some copies of completed survey forms were located, along with instructions for completing the forms. The completed forms were matched up to the corresponding digital records, and the meanings of the columns became clear.

However, without this extra 'digital archaeology' effort, the carefully preserved and migrated data would have been meaningless as a result of the failure to preserve the corresponding metadata.

Source: www.ariadne.ac.uk/issue39/ndad/ (last accessed on 11/01/2006)

5.8   Key UK digital preservation initiatives

Despite the somewhat gloomy picture painted so far, there are a number of positive developments that are worth noting. Major organisations that have played a leading role in preserving non-digital resources, such as the British Library and The National Archives (TNA), have also extended this mandate to digital materials. Such organisations also play an important role in providing digital preservation advice to others.

The National Archives provides public access to its Digital Archive, which stores electronic records of government. In 2004, the Digital Archive won the inaugural Digital Preservation Award sponsored by the DPC.   TNA has also developed the PRONOM service, which holds information on how to access different file formats. The latest version includes the Digital Object Identifier (DROID), which is a freely downloadable tool that can detect the format of files. TNA also provide support directly to government bodies and more generally through freely available guidance notes.

The British Library (BL) is actively involved in a range of research projects and working groups both locally and internationally in collaboration with other institutions and companies. It is also developing a Digital Object Management System (DOM) capable of dealing with heterogeneous digital materials on a large scale. These may either be created or purchased by the British Library, or else acquired through voluntary or legal deposit. The British Library plays a key role in supporting UK scientific research, which is increasingly creating and needing to access digital resources.

Other organisations have extensive experience in managing digital data. **The UK Data Archive** (UKDA) holds the largest collection of digital data in the social sciences and humanities in the UK. **The Arts and Humanities Data Service** (AHDS) manages a digital archive to support research, teaching and learning in the arts and humanities. **The Natural Environment Research Council** (NERC) has created a number of digital archives to store research data. Many of these bodies are also involved in digital preservation research and provide support to other organisations in their field. Some organisations, such as AHDS, UKDA and the University of London Computer Centre (ULCC), also provide a service to others by providing secure storage and management of digital materials on behalf of others.[9]

There are organisations which focus on providing support for digital preservation and curation. These include the **Digital Preservation Coalition** (DPC) which was set up to help coordinate and drive joint action in the UK and the **Digital Curation Centre** (DCC) which provides a national focus for research and development, and which promotes expertise and good practice.   Such collaborative ventures offer an opportunity to reduce risk through shared experience and to reduce costs by achieving economies of scale and by developing shared services. They therefore provide the potential to help small and medium size organisations overcome the substantial barrier to getting started.

The Joint Information Systems Committee (JISC) has been a key driver in initiatives aimed at embedding digital preservation and curation within HE and FE institutions. JISC has provided significant funds for digital preservation research and is also one of the co-funders of the Digital Curation Centre and a founding member of the Digital Preservation Coalition..

Also in the UK, the **Office of Science and Technology** (OST) has established a high-level e-infrastructure steering group in response to recommendations arising from the Government's ten year Science and Innovation Investment Framework (July 2004). This has in turn recognised digital preservation and curation as one of six key issues to be taken forward by its working groups.

At a broader European level, the **EU Commission** has provided funding through their IST programmes. They have funded projects such as Presto and its successor, PrestoSPACE led by the BBC. The Presto projects aim to develop a coordinated approach to the preservation of digital audio-visual media. FP6 included digital preservation in its themes under the 5th call and FP7 is expected to focus on supporting cooperation between universities, industry, research centres and public authorities in order to gain leadership in key scientific and technological areas.

UNESCO has endorsed a Charter on the preservation of digital heritage, complemented by *Guidelines for the Preservation of Digital Heritage* (2003). These documents have provided a

---

[9] Details of these and other repositories can be found in the DPC's *Directory of Digital Repositories and Services in the UK* www.dpconline.org/graphics/guides/index.html#directory (last accessed 05/01/2006).

focus for further action and are indicative of how seriously this issue is taken. The recently released *Dynamic Action Plan* (2005) aims to coordinate the digitisation of cultural and scientific content in EU member states and includes a section on digital preservation, reinforcing the need to include consideration of digital preservation when digitising content but also noting that digital preservation 'has still not been embedded in service or policy development'[10].

A European Task Force has been established, tasked with addressing issues associated with maintaining permanent access to the digital scientific record, which has the objective of drawing up a strategic programme[11].

This is by no means an exhaustive list. It is just some of the many initiatives and activities which are indicative of a growing realisation that this is an issue which needs to be addressed with increasing urgency.

5.9   Adoption of standards

In addition to the creation of real-life successful solutions, significant progress has also been made towards the adoption of agreed standards. The 2005 DPC survey indicated that many organisations use global standards such as the OAIS reference model[12], which provides a common framework of terms and concepts for digital archives, and the PREMIS[13] working group guide to core metadata for supporting the long-term preservation of digital materials. These standards are important because they provide some guidance to organisations with little experience of digital preservation and they ensure a certain degree of interoperability between organisations. The PREMIS dictionary and data model was only released within the last year, and yet it was widely quoted in the 2005 DPC survey, suggesting a high degree of interest.

Some global digital preservation standards

The OAIS (**Open Archival Information System**) reference model describes the characteristics of digital archives. Written by the Consultative Committee on Space Data Systems, it has become widely accepted especially in terms of defining a common terminology and is now as ISO standard (ISO 14721:2003). In particular, it describes six functional entities that must be present within an archive: Ingest, Storage, Access, Preservation Planning, Data Management and Administration. This emphasises how digital preservation is about a lot more than just storing things. There are many archives in place around the world that claim to be 'OAIS-compliant'. OAIS is also a key component of the *Audit Checklist for the Certification of Trusted Digital Repositories* referred to in 6.4.

PREMIS (**PREservation Metadata Implementation Strategies**) is an international working group, set up by Online Computer Library Center (OCLC) and the Research Libraries Group (RLG) to come up with a data model that can be used in digital preservation. The final draft of this report was published in summer 2005 and immediately received widespread attention in the worldwide digital preservation community. This project also won the 2005 Digital Preservation Award sponsored by the DPC.

---

[10] *Dynamic Action Plan for the co-ordination of digitisation of cultural and scientific content.* November 2005. p.8

[11] For further information on the Taskforce see tfpa.kb.nl/ (last accessed on 24/01/2006)

[12] For more details on OAIS, see ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf (last accessed on 03/01/2006).

[13] PREservation Metadata Implementation Strategies (PREMIS) aims to define a core set of metadata which is broadly applicable across the digital preservation community. More details can be found at www.oclc.org/research/projects/pmwg/ (last accessed on 03/01/2006).

5.10 Sector-specific differences

Having outlined the general picture above, this section discusses those issues and problems that apply to only some of the sectors interested in digital preservation.

Private and public sector differences

The public and private sectors had a lot in common in terms of the responses to the 2005 DPC survey. However, some notable differences were revealed. For instance, funding restrictions were found to be much more acute in the public sector. 80% of public sector responses quoted lack of funding as a barrier to their digital preservation activities, where as this was as low as 31% in the private sector, where the attitude seems to be that funding will be made available once there is a sufficiently strong justification for digital preservation. It is perhaps significant that the industries in which digital preservation has developed most include those driven by the legal imperative, notably the pharmaceutical industry.

This prominence of the legal driver in the private sector is also reflected in the challenges that are raised. In the private sector, the legal admissibility of preserved digital records came out as the most common barrier to digital preservation (56% of survey responses). In the public sector, however, this was only the 12th most important barrier out of 17 alternatives that were offered.

Digital preservation in regulated industries

In the pharmaceutical industry, new drug products may take 10 years or longer to develop. Products are subject to rigorous inspection by national regulatory bodies to establish their safety for human or animal use, and their effectiveness for intended purposes. Records and data that support a marketed drug's development and testing must be available for regulatory inspection during the compound's lifetime, a period generally assumed to be 40 years or longer.

Regulations include the international GxP Good Practices regulations, but also compliance with rules imposed by the US Food and Drug Administration (in particular FDA 21 CFR part 11) is required for obtaining permission to market a drug in the US – hence of great importance to UK pharmaceuticals.

Such external regulation compliance requirements have created an environment in which pharmaceuticals have developed an awareness of digital preservation issues, and several large companies have invested heavily in creating digital preservation solutions. The preservation of data has opened up other significant opportunities, such as electronic patent submissions and protection of Intellectual Property, and also data mining and data re-use opportunities, which will potentially speed up compound discovery.

Heavy competition within the pharmaceutical sector has led to solutions being developed independently and technologies being fiercely protected.

In the financial sector, the Financial Services Authority regulations in the UK, and Basel II internationally have implications for preserving records. In 2007, many organisations in the financial sector will also be subjected to the MiFID regulations imposed by the European Union. MiFID has significant requirements for preserving digital communications in order to support transparency of trading operations.

Digital preservation in research

Most modern research generates digital data. Preservation of this data enables open access to it. This in turn allows research results to be validated, and for new or alternative hypotheses and methods of analysis to be tested. Open access to research data also facilitates the education of new researchers, and enables the exploration of topics not envisaged by the initial investigators. It also permits the creation of new data sets when data from multiple sources are combined.

"Getting systems in place to support data preservation for sharing enhances the long-term value of the research time invested in generating the data"

Source: 2005 DPC survey

The Research Councils are increasingly encouraging digital preservation by making submission of research data to a digital archive a condition of funding for research projects.

Preserving audiovisual materials

Audiovisual media and technologies have traditionally been impermanent, and deteriorate over relatively short time scales. Digitisation provides the solution to this problem and has become widely accepted as the preferred method for audiovisual preservation. The audiovisual archiving community, therefore, has a strong incentive to address the challenges of digital preservation. Many audiovisual archives have been created and cooperate via bodies such as the Film Archive Forum[14].

Libraries and archives

A particularly significant aspect of digital preservation in libraries and in archives is that their remit is to preserve digital material produced outside of the organisation. As a result, there is little or no control over the format of the data that comes in, leading to additional levels of complexity and additional overheads that do not exist in digital preservation in other organisations. In addition, the material being transferred may not have been well preserved in the period between creation and transfer leading to potentially expensive remedial action being required, for example to ensure completeness and/or to create a description that will enable future users to find and make use of the material.

Research libraries and academic publishers face particular archiving and preservation challenges arising from new business models based on licensing of access to electronic journals and collections. When libraries no longer have a physical copy they can preserve and access, issues arise over continuing access and long-term preservation of the materials to which they have subscribed. Trusted third-party digital preservation services are increasingly seen as essential to underpin the development of electronic publishing.

A further level of complexity is created by the need to comply with copyright and privacy laws. One problem is in the identification of material that needs to have access restricted. Another issue is the paper paradigm of borrowing and returning breaks down in the digital world. A physical book can be borrowed from a library only once and needs to be returned before it can be borrowed again while a downloaded document is not 'returned' and it is difficult to ensure that the borrower does not copy it and distribute it more widely.

---

[14] A list of members of the Film Archive Forum can be found at www.bufvc.ac.uk/faf/members.htm (last accessed on 03/01/2006).

6 UK Digital Preservation Needs

The UKNA surveys have revealed that the level of implementation of digital preservation solutions is significantly lower than would be expected given the awareness and commitment that were measured. Digital preservation is a new and complex activity, and this section explores some of the reasons for the lag between commitment and action.

6.1 Growing awareness

The 2005 DPC survey was targeted at organisations (and individuals within organisations) whose role would suggest that they should be aware of the issues of digital preservation. This resulted in reporting of impressive levels of awareness of the problems, which is a testament to the advocacy by bodies such as the DPC and the DCC and increasing coverage in the national press. While such activities have raised the profile of digital preservation, some more awareness raising work needs to be done.

"The [...] archive is a corner of a much larger enterprise that considers an archive as a place where email goes when it dies -- not as an online service with wide-ranging functionality across that enterprise"

Source: 2005 DPC survey

A successful digital preservation solution needs more than just management buy-in; it needs awareness and commitment at all levels throughout the organisation and often collaboration with other organisations. Extra effort is required of all creators of digital data to submit their data to the archive. In some cases this can be minimal, but often some effort will be required to make sure the data is in a suitable format and that it is associated with suitable metadata. In many cases data creators may be reluctant to do this as they may see little benefit to themselves. Indeed they will primarily be benefiting others by giving them greater access to their data. In some cases, this can be seen as a serious disincentive. In the academic environment, many funding bodies have got around this problem by managing their own digital repositories and by specifying submission of data as a condition of funding.

> **Need 1:** Awareness of the impact of digital preservation needs to be more commonplace and spread beyond the current core of informed individuals and organisations. In particular, awareness needs to be greater amongst data creators.

6.2 Translating awareness into action

Raising awareness of the digital preservation problem is just the beginning. Organisations also need to come up with an action plan to resolve the issues. One of the key barriers to actions described above was the lack of clear responsibilities within an organisation.

"There is an awareness of the problem within the organisation, but no practical solution."

"Senior managers say [digital preservation] has a very high priority but there are lots of high priority projects so maybe it needs more attention. Problems include not being sure what to do but perhaps more importantly, is not being clear who is responsible"

"Buy-in from the rest of the organisation would also help. The archives are a small part of the organisation compared to content creators so it is difficult to influence the way things are done.'"

Source: 2005 DPC survey

> **Need 2:** Organisations need to go beyond awareness of the problem and create a proactive plan for dealing with the issues. This should include appointing clear responsibilities for each part of the plan.

The first step in such a plan should be to understand how much the problem actually affects the organisation concerned. For example, if an organisation has been using digital material for a long time, then they are likely to have an immediate problem with legacy information while newer organisations may not have such issues but may be storing up a problem for the future.

Need 3:	**Organisations need to start action by taking stock of the digital materials they hold and are currently creating along with the time periods over which they should be preserved.**

6.3   Funding justification

Digital preservation is not a straightforward activity to fund. Over 70% of people interviewed for the 2005 DPC survey felt that lack of funding was a problem. Guidance needs range from help putting together a case for influencing senior management (for example by providing relevant cost-benefit examples) to the more technical requirements for help in planning digital preservation activities.

Funding models

A strong sense was revealed through the three UKNA surveys that digital preservation is a new activity on top of those already undertaken, and that for the foreseeable future it will not replace current preservation activities. As a result, in order to be successful, digital preservation needs to be considered as a separate activity in its own right, with its own distinct allocation of resources.

The common project-based funding model can hinder digital preservation activities in that project managers are focussed on the successful conclusion of their own project and do not have a long-term view on the future preservation and accessibility of the data and documents produced during the project. Indeed, it is often a different part of the organisation – or even a different organisation altogether – which will derive the benefits from the long-term preservation of information.

"In today's restricted budget situations, there is little emphasis or understanding with regard to digital preservation issues i.e. as long as "stuff" is backed up then all is ok! As a result, resources are difficult to come by and maintain."

"[A key digital preservation challenge for my organisation is] deciding on what should be kept and what can be dumped. To do this effectively would add overheads to each project and there is no management commitment to do this as it would affect the profit margin"

"Digital preservation is still in the R&D stage, and is funded using project money. If a permanent digital preservation function is to develop, it will need more stability than this"

Source: 2005 DPC survey

The MLA regional survey revealed a strong sense that no long-term funding is being provided for the management of digital material. Given the nature of digital preservation, it can only be successful if there is a commitment to long-term funding.

Two JISC-funded projects should provide useful guidance to organisations wishing to assess, and justify, the costs of their digital preservation programme.

- The LIFE (Lifecycle Information for E-Literature) http://www.ucl.ac.uk/ls/lifeproject/ is a collaborative project between the BL and UCL and is investigating methods of attributing costs of digital preservation based on identified stages of its lifecycle.

- The eSPIDA (An effective Strategic model for the Preservation and disposal of Institutional Digital Assets**)** project www.gla.ac.uk/espida/ is based at the University of Glasgow. This two-year project seeks to develop and implement a sustainable business focussed model for digital preservation, as part of a knowledge management agenda in higher education institutions.

Need 4:	**Project based funding needs to take into account the long-term value of the information produced by the project and the cost of its retention.**

Cost-benefit analysis

It may be possible to produce some good estimates of costs, but the lack of practical examples makes it extremely difficult to put a value on the benefits of a digital archive, particularly as many of these are intangible and are spread over a relatively long time-scale.

"Costs (and indeed all kinds of resources) are very difficult to quantify and forecast reliably. There remains a lack of standards and benchmarks, and this makes it hard to compare ourselves (and costs) with other organisations"

Source: 2005 DPC survey

The highest profile published example of a cost-benefit analysis for a digital archive was prepared for the Washington State digital archive[15]. This lists detailed costs, but is not able to put a value on the benefits.

This lack of a clear cost-benefit justification for digital preservation programmes, along with the lack of expertise and the fact that this is a relatively new discipline create a perception that digital preservation programmes are risky. This clearly hinders management buy-in.

Need 5:	**It needs to be easier to build a business case for the long-term preservation of digital material, e.g., a cost-benefit analysis. In particular, there needs to be the ability to measure digital asset value taking into account any long-term commercial value (e.g., by enabling future decision-making) and the risks of losing information (e.g., risks to public safety or health).**

6.4   Repositories for all

No commercial off-the-shelf, digital archiving solution exists at present, and any solution has to be created to suit the individual needs of an organisation. The problem can, therefore, only be resolved with considerable internal effort towards developing the necessary strategies and the necessary IT infrastructure. Not all organisations will wish to take on the challenges posed by digital preservation and will instead want to use a third party repository. However, this approach is currently very risky since there is no established way in which an organisation can measure the suitability of a potential solution provider. Hence, there is a need to be able to trust outsourced solutions.

"There is a general lack of awareness of the need for digital preservation among smaller organizations. The issues are not known about nor understood."

Source: MLA survey

Many respondents to the 2005 DPC survey referred to digital preservation standards, such the OAIS[16] framework and the PREMIS[17] guidance for metadata. In general, these were found to be useful, but it was felt that there was still a need for more standards, and in particular that these should be supported by freely available practical examples.

Need 6:	**Internationally-agreed standards are needed to help guide digital preservation projects and thus reduce the risks involved.**

Some respondents also felt that it would be useful to underline standards with a certification process. A step towards repository certification has recently been taking by the publication of a draft Audit Checklist for the Certification of Trusted Digital Repositories[18] by the Research Libraries Group and the National Archives and Records Administration in the US. Of course, such a checklist could easily be applied to in-house solutions as well.

Need 7:	**Organisations need to be able to place digital material into a repository (whether in-house or provided by another party) with measurable confidence.**

---

[15] Details of the Washington State digital archive can be found at www.digitalarchives.wa.gov/Content.aspx?txt=background (last accessed on 03/01/2006).

[16] The Reference Model for an Open Archival Information System (OAIS) is a framework for digital archives. See ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf for more details (last accessed on 03/01/2006).

[17] PREservation Metadata Implementation Strategies (PREMIS) aims to define a core set of metadata which is broadly applicable across the digital preservation community. More details can be found at www.oclc.org/research/projects/pmwg/ (last accessed on 03/01/2006).

[18] See www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf (last accessed on 03/01/2006)

Where they exist, the national, regional or domain-specific repositories are found to be of great value, and there are requests for more to be created. In particular, the MLA regional survey suggested the provision of a national or regional central storage provision for museums, libraries and archives, possibly along the lines of the AHDS.

"Perhaps a centre of expertise could be established - a service which archives digital private papers acquired by several institutions and provides guidance to creators of digital private papers and to curators in other libraries"

Some public organisations have a clearly identified final repository for their digital material. For example, Government departments are obliged to deposit their records with The National Archives. However, many other public organisations (e.g., local authorities) are not currently provided for at all and so they need to do the best job they can to retain their

"A wider agreement on the responsibilities of repositories which hold these sources would assist at a strategic level"

Source: 2005 DPC survey

information in-house. Even in cases where there is a clear 'final resting place' for information, there are complications: for example, a government department may need to retain digital material itself for substantial periods of time before passing such records to TNA, which could easily render the records obsolete before such a transfer occurs. Hence, there is a need for all public organisations to be able to identify a repository for their digital information at various points in their lifecycle.

Need 8:	**All public organisations need to have an identified repository where their material can be sent to be preserved at all times in the lifetime of that information. This needs to include those organisations that are unable to justify the capital expenditure involved in performing digital preservation (e.g., because they have only a small volume of digital material to store).**

The particular challenges of web archiving

In just a short time the Internet has become an essential mass means of sharing and distributing information, a medium for both social and scientific communication, e-commerce, and much more. Often material is published in no other form than on the Internet.

The dynamic and topical nature of the Internet means that information and content frequently change, are updated or disappear without notice and often without leaving any trace. Increasingly this includes best practice guidelines, professional regulations and legal compliance information that are replaced as they become outdated or are superseded.

Web archiving initiatives have been launched to help capture the informational, cultural and evidential value of web based materials.

- The dynamic nature of the web imposes time constraints on gathering web based material, and – once lost – material can rarely be recovered.

- Preserving a record of best practice guidelines, professional regulations and legal compliance information provides a body of information for future research.

- Web archiving supports the preservation of the intellectual output of organisations, or even countries, as the basis for building future knowledge. The re-use of information can reduce future research costs.

Increasingly organisations have a duty not just to make corporate information available but also to make it available for the long-term. Co-operation between organisations and web archiving initiatives, such as UK Web Archiving Consortium (UKWAC – www.webarchive.org.uk/) in the UK, can reduce the burden of this compliance whilst ensuring the longest possible life and availability for important information.

6.5   A new discipline

A whole professional infrastructure has been built up around the preservation of paper documents. In contrast, digital preservation is still very much in its infancy. This, in turn, means that organisations must have access to staff with the necessary digital preservation expertise. There is currently a shortage of such staff, so most organisations rely on developing existing staff. There is therefore a strong need for digital preservation training, such as the programme recently set up by ULCC, the DPC, Cornell University and the British Library, and funded by JISC[19]. Organisations, especially in the public sector, also rely heavily on guidance from bodies such as TNA, the DPC, and the DCC.

The need for greater provision of guidance and training support has featured in all three UKNA surveys. In the 2005 DPC survey this was requested by as many as 89% of respondents.

Some areas have better access to guidance than others. In higher education and research, support is provided by the DCC, by the UKDA in the social sciences and humanities, by AHDS in the arts and humanities, by the Research Councils in the sciences, and by JISC more generally. Support to local government is provided by TNA and support to libraries is provided by BL. The DPC also provides more generally available support.

Training needs have also been expressed at a wide range of levels including technical, organisational and strategic. It also needs to cater for different levels within the organisation, both operational and senior management. The new Digital Preservation Training Programme (DPTP – www.ulcc.ac.uk/dptp/) has been developed in partnership with Cornell University, whose successful workshops in the US provided a model for the UK-Based programme. The DPTP provides an intensive, residential programme structured around the OAIS model and received excellent feedback from the October 2005 pilot, suggesting this has the potential to meet a growing need. One practical suggestion made by respondents to both the DPC members' survey, and to the 2005 DPC survey, was that digital preservation training should form part of the professional training for conservators, archivists and librarians.

Need 9:	**The new discipline of digital preservation needs to be supported. This should include the provision of continual professional development for existing individuals with relevant skills sets, e.g., archivists, librarians and IT staff.**

Need 10:	**Digital preservation training should form part of the professional training for conservators, archivists and librarians.**

A considerable challenge with digital preservation training is that the activity involves a number of members of staff within an organisation, ranging from the data creators to those responsible for running the digital repository itself. Training and awareness raising are therefore required throughout the organisation, making this a costly exercise.

Several respondents to the 2005 survey felt isolated in their digital preservation activities and suggested building up communication networks between digital preservation professionals in different organisations to promote exchanges of ideas, experiences and advice. This is clearly a very important step especially as training will not provide all the answers since solutions to all the issues are not yet known. The provision of such a support network is needed to ensure that the new discipline matures.

"There is a need for a coordinated sharing of best practice across all sectors employing digital media with a view to defining procedures and standards. In the heritage and cultural sector these should not only take account of what is possible/desirable, but also reflect what may be practicable given limited resources"

Source: MLA survey

In fact, there is already an active digital preservation community as evidenced by the numerous conferences and the activities of organisations such as the DPC. Hence, perhaps the real issue is greater awareness? The DPC makes considerable efforts to reach beyond the 'core' community of its members (which are mainly museums, libraries, archives and universities) but it is clear that this needs to continue.

---

[19] The Digital Preservation Training Programme: www.ulcc.ac.uk/dptp/   (last accessed on 03/01/2006)

> **Need 11:** There needs to be cross-disciplinary forums to allow both experienced individuals and organisations to exchange digital preservation best practice and others to turn to for advice.

6.6 Government and EU policy and programmes

As with all organisations, government is creating more and more information in digital form. Recent government policy has supported this increase as well as increasing the rights of the public to have access to this information through the Freedom of Information Act 2000. While this is of benefit to members of the public, it does place a burden on public organisations to be able to comply with the provision of this Act.

The UK Modernising Government Agenda promotes "cross-governmental co-ordination machinery and frameworks". The UKNA surveys have, however, found that coordination of government activities is a problem in the field of digital preservation. The DPC members' survey found that building "strategic partnerships at national level, including sharing of information and solutions" was a key issue. The MLA regional survey identified a need to clarify the link between e-government principles and the need for digital preservation. Several organisations complained in the 2005 DPC survey of a lack of coordination of government initiatives, particularly across different departments.

"Hiring staff with prior experience of digital preservation as opposed to some aspects of IT relevant to preservation is difficult"

Source: DPC members' survey

"Archivists at many levels feel inadequate when it comes to dealing with electronic records. We would like to see some national training initiative - at least to provide our profession with a base line."

"Training of staff in the issues and how to deal with them would be very useful. A lot of us are very unsure about where to start."

Source: MLA survey

"We also need more personnel with the right skills, or access to training and secondments."

Source: 2005 DPC survey

Some efforts are being made internationally to coordinate activities with projects such as the European Commission's European Digital Libraries project[20]. This includes digital preservation as one of its three key areas for action. It aims to bring together the fragmented initiatives in different Member States.

This lack of co-ordination and preparedness can be traced back to the impact of the policies not being considered from a digital preservation perspective. This is understandable given that there was not widespread awareness of the issues at the time of the creation of the relevant policies but there is clearly an opportunity to consider the implications of current policies for the future, and of future policies before implementation.

> **Need 12:** The impact of new legislation (e.g., the Freedom of Information Act 2000) and government initiatives (e.g., the Modernising Government agenda) on digital preservation requirements needs to be fully considered both before and after implementation.

In addition, policies and programmes need to be co-ordinated at regional, national, European and global levels to align research and funding arrangements, to allow business efficiencies and to ensure that economies of scale are delivered. This will also help to establish the network of repositories described above.

> **Need 13:** Policies and programmes need to be co-ordinated at regional, national, European and global levels.

---

[20] For more details, see press release:
europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1202&format=HTML&aged=0&language=EN&guiLanguage=en (last accessed on 03/01/2006)

6.7   Legal and regulatory frameworks

Legal admissibility

When digital records are being preserved to satisfy a legal requirement or as protection against litigation, it is particularly important that the legal admissibility of the records is clear. Demonstrating, for example, that a digital record is genuine and that it hasn't been modified since its creation are real challenges for which there are no clear solutions. This is often given as a reason for continuing to preserve paper records.

The 2005 DPC survey found that this is the most important barrier to digital preservation in the private sector (cited in 56% of cases). This is a growing concern, particularly as businesses rely more and more on electronic communication for contractual exchanges.

The British Standards Institute publishes a code of practice on the admissibility of electronic records. The latest version of this, BIP0008, was published in 2004. Previous versions (BSI DISC PD0008) were found to be difficult to follow in practice, not least because it was not possible to provide a checklist of items that would guarantee that courts would accept documents as genuine. Courts tend to treat non-original evidence, such as a printout from an electronic record, as 'hearsay' and thus give it lower credence than oral testimonies or original documents. In addition, such 'hearsay' is subject to authenticity tests based on evidential weight. Documents that have been altered are unlikely to be accepted. However, digital preservation relies on active management of the document, which might include migration through a controlled process to alter the content in order to ensure it is readable into the future. At best, the situation with regard to the legally admissibility of well-preserved digital information can be described as confused.

Need 14:	**There needs to be a greater understanding of the steps organisations can perform to ensure that digital information remains legally admissible.**

Copyright

Both the MLA regional study and the 2005 DPC survey revealed that copyright can significantly impede digital preservation. It can add an extra burden onto the deposition process – for example by adding a manual step whereby an archivist must check for any copyright restrictions on all constituent parts of a record (e.g. images included in a presentation or a report). In some cases, copyright issues will mean that an organisation has to restrict access to its digital archive.

Need 15:	**Organisations that hold digital information need to be able to cope with the access restrictions imposed by copyright legislation.**

Regulated industries

Organisations working within regulated industries understand that they must comply with requirements that information on, for example pharmaceutical products, must be able to be made available a long time (in many cases many decades) into the future. This is clearly difficult for information in a digital form since this could span many generations of software and hardware. Since regulators could challenge the authenticity of records, organisations mainly get around this issue by printing both documents and data records, signing them by hand and filing them in a traditional archive. This is an expensive process, which may provide no better protection from fraud than a well designed digital system, and possibly even less. However, regulators have an established track record of acceptance of records from traditional, paper-based systems while it is not clear that regulators will accept information from a modern system. The impact of this is an unnecessary increase in business costs and the un-measurable cost caused by the reduction in the accessibility and reuse of information that storing information non-digitally entails (e.g., the ability the find new candidate drug compounds by data mining old experimental results).

Hence, regulators need to work with their respective industries to provide guidelines that make it easier for industries to adopt digital archiving solutions that both safeguard public safety by placing high demands on the authenticity of records and yet allow them to exploit the opportunities retaining digital information provides.

Need 16:	**Organisations working within regulated industries need the ability to perform digital preservation activities (e.g., following a given migration pathway) that can be certified as acceptable by regulatory authorities.**

6.8   Reducing risks

Research

The practical long-term preservation of digital materials is inherently difficult to plan because of the unknown future direction of digital preservation. Much research into digital preservation methods is still ongoing. The main solutions for tackling hardware and software obsolescence are migration and emulation.

Migration involves updating the digital files to newer formats when the old formats become obsolete. Migration tools are in their infancy, but are often quoted as the preferred strategy to cope with future changes in technology.

Emulation, on the other hand, involves leaving the original files unchanged, but viewing them with the help of specialist software, which simulates the hardware and operating system that were originally used to create the files. The use of emulation is very rare, and has only been used by 3% of the respondents to the 2005 DPC survey.

"The library is increasingly being offered archives that have a significant digital component. ... Much of this is the form of emails. Thus, we will need a system to 'accession' this and make it available - and recognizing that the volumes of digital objects are high, we need some way to automatically extract metadata.

For example, if we acquire thousands of emails - there is no way we can manually catalogue these. We need a system which automatically extracts metadata - such as sender, recipient, subject, date etc."

Source: DPC members' survey

A problem faced by many organisations which are planning to use either emulation or migration tools is that of the acceptability of the tool and hence of legal admissibility of the output. Again, this is a problem in its infancy, which drives many organisations to use the "print and file" approach as a more secure one for documents which have legal importance.

Some suitable tools and sources of information are available. Notably, The National Archives provides a Web service, PRONOM, which holds information about file formats and the software and hardware required to access them. These tools are found to be useful, and further suggestions have been made, including tools for migration, format validation and metadata extraction. These tools and services will also be essential if the levels of automation in digital preservation are to increase to reduce costs and help address the growing information deluge.

> **Need 17:**   There needs to be more technical tools to help organisations perform digital preservation activities such as performing format migrations, format validation and automated metadata extraction.

Open Standards for file formats

File format obsolescence is one of the major challenges of digital preservation. Software vendors drive the rapid changes in file formats with their continual push to include new features. As digital preservation issues gain greater prominence, software vendors will respond. However, appropriate advocacy would speed up the process of vendors committing to supporting a set of open file formats.

> **Need 18:**   Promote initiatives to define a set of open file formats to which major software vendors can be encouraged to commit to providing long term support.

7   Needs and Recommendations

The following table brings together the needs that have been brought out in this report. For each need, the current state of progress in meeting this need is summarised and a recommendation for further action is made (where appropriate).

No.	Need	Current State of progress	Recommendations
Growing Awareness			
1	Awareness of the impact of digital preservation needs to be more commonplace and spread beyond the current core of informed individuals and organisations. In particular, awareness needs to be greater amongst data creators.	This work is currently well supported by organisations such as the DPC and DCC. They are planning to extend their influence, especially within the commercial sectors.	Given the results of the survey, this will be an ongoing need for the foreseeable future.
Translating awareness into action			
2	Organisations need to go beyond awareness of the problem and create a proactive plan for dealing with the issues. This should include appointing clear responsibilities for each part of the plan.	Most organisations have no plan and responsibility for dealing with digital preservation issues is usually very unclear.	Organisations should create a long-term proactive information/knowledge management plan. This should start with an information audit and then, once the problem has been appropriately sized, the task of creating a solution should be properly resourced. This should include the assignment of clear responsibilities.
3	Organisations need to take stock of the digital materials they hold and are currently creating along with the time periods over which they should be preserved.	Most organisations do not know the extent of the problem they face.	Organisations should perform regular information audits to measure (and continue to measure) their digital preservation needs, and to ensure that these are being met.
Funding Justification			
4	Project based funding needs to take into account the long-term value of the information produced by the project and the cost of its retention.	Very few projects take a long-term view of the information.	Organisations should include long-term data retention as a criterion in the cost-benefit analysis of any project that will lead to the creation of a substantial amount of digital information.

5	It needs to be easier to build a business case for the long-term preservation of digital material, e.g., a cost-benefit analysis. In particular, there needs to be the ability to measure digital asset value taking into account any long-term commercial value (e.g., by enabling future decision-making) and the risks of losing information (e.g., risks to public safety or health).	Reasonable estimates of costs can be obtained through standard estimation methods and via some examples from previous archiving projects. However, there is very little information on value as most current repositories have been created by organisations that are mandated to retain information.	Funding bodies should support research into the long-term value of digital information and models of how that value may change with time.
			Organisations should consider the long-term value of digital material when putting together plans and budgets.
Repositories for all			
6	Internationally-agreed standards are needed to help guide digital preservation projects and thus reduce the risks involved.	OAIS and PREMIS are major steps but still leave organisations with issues of interpretation and/or a concern of how much of this needs to apply to them. Standards can create a barrier to implementation if it is perceived as requiring an organisation to make large up-front commitments before getting started.	The UK Government and funding bodies should promote further development of collaborative standards and methodologies in digital preservation for all parts of the digital preservation lifecycle. In particular, a roadmap of how to scale up a system while meeting standards is needed.
7	Organisations need to be able to place digital material into a repository (whether in-house or provided by another party) with measurable confidence.	RLG/NARA have established an Audit Checklist for the Certification of Trusted Digital Repositories. Additional funding has been provided to test this checklist on four existing repositories.	This important work needs to be closely monitored and mechanisms for certification need to be developed by the digital preservation community.
8	All public organisations need to have an identified repository where their material can be sent to be preserved at all times in the lifetime of that information. This needs to include those organisations that are unable to justify the capital expenditure involved in performing digital preservation (e.g., because they have only a small volume of digital material to store).	Some information from some public organisations is covered such as records of long-term value within government departments (will be retained by TNA) and some records created through academic scientific research. However, there are currently a lot of gaps.	The UK Government and funding bodies should promote, through seed funding, the creation of more digital archives across relevant sectors and organisations. Promote collaborative regional or national repositories to meet the needs of small and medium size organisations. These can then be promoted as exemplars for other organisations to follow. The overall aim should be to set up a network of trusted repositories.

	New discipline		
9	The new discipline of digital preservation needs to be supported. This should include the provision of continual professional development for existing individuals with relevant skills sets, e.g., archivists, librarians and IT staff.	The JISC-funded Digital Preservation Training Programme (DPTP) has now been piloted and additional programmes are planned for 2006.	Organisations should set aside a budget for training staff.
			Options for subsidising intensive programmes such as DPTP should also be explored in order to help smaller organisations with a limited training budget.
10	Digital preservation training should form part of the professional training for conservators, archivists and librarians	Professional conservators, archivists and librarians have to pick up digital preservation skills when these are needed.	Higher education courses for conservators, archivists and librarians should include digital preservation training.
11	There needs to be cross-disciplinary forums to allow both experienced individuals and organisations to exchange digital preservation best practice and others to turn to for advice.	There is an active digital preservation community in the UK with regular conferences etc. There is increasing take-up of existing services supported by the DPC and others.	Promote wider awareness and use of relevant resources available from the DPC, the DCC and others.
	Government Policy		
12	The impact of new legislation (e.g., the Freedom of Information Act 2000) and government initiatives (e.g., the Modernising Government agenda) on digital preservation requirements needs to be fully considered both before and after implementation.	The digital preservation impact of new policies are usually only considered as an afterthought.	In formulating policies, the UK Government should take into account their impact on the digital information lifecycle. In particular, the sooner digital preservation activities can be carried out, the more effective and economical they can be.
13	Policies and programmes need to be co-ordinated at regional, national, European and global levels.	Some communities, e.g., the bioinformatics community, do have co-ordinated, global policies and programmes but this is unusual rather than the norm.	The UK Government should ensure that programmes and policies are co-ordinated nationally and on a wider scale (especially within the EU).
	Legal and regulatory Frameworks		
14	There needs to be a greater understanding of the steps organisations can perform to ensure that digital information remains legally admissible.	Organisations are not confident that they can take actions that will guarantee legal admissibility of information.	The UK Government needs to work with the digital preservation community and the legal profession to establish clearer guidelines for ensuring that digital information is legally admissible.
15	Organisations that hold digital information need to be able to cope with the access restrictions imposed by copyright legislation.	Copyright is a barrier to wider access which (since this is one of the main benefits from the preservation of information) provides a barrier to investment in long-term information retention.	The UK Government needs to ensure that an appropriate balance is struck between the legitimate rights of information creators and the desires of others to gain enhanced benefits from that information.

16	Organisations working within regulated industries need the ability to perform digital preservation activities (e.g., following a given migration pathway) that can be certified as acceptable by regulatory authorities.	Current regulations can be seen as a barrier to using digital systems and leads to a reliance of 'print and file'.	Regulatory bodies need to work within their sectors and the digital preservation community to map out a framework, which will allow organisations to store information in an exploitable form while retaining the ability to satisfy regulatory concerns.
Reducing Risks			
17	There needs to be more technical tools to help organisations perform digital preservation activities such as performing format migrations, format validation and automated metadata extraction.	A few tools have been created.	All organisations need to encourage an international 'market' for digital preservation tools by linking up with other projects around the world and engaging with software vendors. This would deliver economies of scale and reduce risk for individual institutions.
18	The formats used by data creators need to be influenced by the long-term preservation needs of the information they will create.	Data creators usually use formats that suit their immediate needs without worrying about their long-term preservation characteristics.	Organisations should consider the long-term preservation characteristics of the formats they use.
			Organisations should work together and with software vendors to encourage the development of open file format standards.

Table 3: Needs and Recommendations

Acronyms

The following acronyms are used throughout the report.

AHDS	The Arts and Humanities Data Service
BL	The British Library
DCC	Digital Curation Centre
DPC	Digital Preservation Coalition
EDRMS	Electronic Document and Record Management System
JISC	Joint Information Systems Committee
MLA	Museums, Libraries and Archives Council
NERC	Natural Environment Research Council
TNA	The National Archives
UKDA	The UK Data Archive
UKNA	UK Digital Preservation Needs Assessment study
ULCC	University of London Computer Centre

Steering Group

Dr Juan Bicarregui
e-Information Technology Sector Manager,
Council for the Central Laboratory of the
Research Councils

Hugh Campbell
Head of Information Systems,
Public Record Office of Northern Ireland

Adam Farquhar
Head of Architecture, eIS,
The British Library

Helen Hockx-Yu
Programme Manager,
Joint Information Systems Committee of the
Higher and Further Education Funding Councils

Jim Jamieson
Senior Archivist,
University of London Computer Centre

Maggie Jones
Executive Secretary,
Digital Preservation Coalition

Dr William Kilbride
Assistant Director,
Archaeology Data Service, AHDS

Martin Nail
Information Management Adviser,
Museums, Libraries and Archives Council

Dave Thompson
Web Archiving Project Officer,
The Wellcome Library

Stella Wisdom
Research Assistant, Strategic Policy Division,
National Library of Scotland

Acknowledgements