



Your Data At Risk

Why you should be worried about
preserving electronic records

Foreword

You most probably own a personal computer. Consider the ever increasing tangle of processed words, spreadsheets, PowerPoint presentations, emails, websites, jpegs, and mpegs colliding inside it. Are you confident that you could convey a project case history that's stored on that computer to your successor? Not sure? Now imagine that uncertainty transferred to the whole organisation in which you work. Scary? Please read on...

1. Who the report is aimed at?

This report is aimed at those with responsibility for, or influence over, the strategic direction of an organisation's information management within the UK. Namely:

- Governing bodies
- Chief Executives
- Senior Administrators
- Information Managers and Professionals
- Funding agencies.

And that's because to tackle sustainable digital preservation of records successfully requires a strategic direction and policy commitment from the top of the organisation that translates into operational effectiveness below that level.

2. Why should I care?

A piece of paper can last for centuries left alone in a dry, dark room. Nothing created by a computer has that kind of inherent longevity - nothing like it in fact. Computers and their contents only survive by the active and ongoing help of human beings. Without this help, evidence that is essential to the smooth running of your organisation is lost forever and with it your organisation's ability to meet its business, legal and cultural needs over time, including your responsibilities under Freedom of Information and other legislation.

"Oceans of information get 'saved' on file servers, on personal hard drives, CD-ROMs and DVDs but none of these are true archives. We have no guarantee that these will be readable or even usable in 50 years time."

Bruce Sterling, The Telegraph, 14 March 2005

3. What is the report trying to do?

This report aims to provide, in layman's terms:

- A clear guide to what digital preservation is
- A clear and simple description of the issues surrounding digital preservation
- A demystification of the idea that the challenges and solutions are purely technological rather than being about strategies and processes developed by people
- Recommendations for action by the intended audiences
- Sources of further advice.

The master tapes for rock classics like the Eagles' "Hotel California" and REM's "Automatic for the People" are suffering from 'sticky shed syndrome', meaning the tape is literally sticking together as the chemicals in it degrade.

Source: BBC Radio 4

4. What is digital preservation?

Digital preservation is about a series of actions that need to be taken and managed to make sure there is continued access to digital materials for as long as is necessary. As long as is necessary could mean long term - into the indefinite future, or short-term – for a specific time limited business requirement.

4.1.1 Types of digital materials

Digital materials that need to be preserved can come in a variety of forms from very simple text documents to complex web-based resources combining sound and image. They can be either static, like PDF documents or dynamic, like web pages:

4.1.2 "Born digital" electronic records are created digitally in the day-to-day business of an organisation and refer to digital materials which were created digitally, were not intended ever to have an analogue equivalent and probably only exist in digital form. They are given formal status by the organisation that created them. They can include, for example, word processing documents, emails, databases, or intranet web pages. Some digital records are available in physical form e.g. have been printed to paper, although some may lose their meaning as records in context by this process.

"Last year marked the 30th anniversary of email. But it is salutary that we do not have the first email message, and no knowledge of its contents other than it was in upper case. Contrast this with how much we know about the first telegram (now digitised and on the web) or telephone message." Loyd Grossman, 2002.

Digital publications are “born digital” objects which have been expressly released for wider access to an audience and either made available or distributed free of charge or for a fee. They can be networked publications available over a communications network or physical format publications that are distributed on formats such as floppy or optical disks.

4.1.3 Digital surrogates are created as a result of converting material in analogue form to digital form. They are the products of digitisation e.g. scanned images of original documents or objects.

5. How does this relate to my organisation’s objectives?

5.1 Business objectives

Without active management of records, organisations are **failing to reap the rewards of investment in creation of digital materials**. For example, without proper preservation of digital assets, it is impossible to exploit those assets commercially (where possible) or make the most of Intellectual Property Rights and Copyright.

Organisations are leaving themselves exposed to **reputational risk**, as they will be increasingly **unable to find and make available corporate information** that is essential to maintain corporate standing.

This inability to retrieve and access information will **affect increasing demand by customers for information** (both current and historic) to be made available over the Internet or, at least, in an offline digital format.

The **consequences of doing nothing will also be felt internally** as potential efficiency gains in administration are lost and more and more resources are spent on locating information and wondering what on earth to you are going to do to make it readable or searchable again.

For many organisations, there is a series of **legal imperatives that must be met**. Drivers such as the **Freedom of Information Act** (2000), the Data Protection Act (1998), Environmental Information Regulations (2004), and the increased demands of **Corporate Governance** required by the Sarbanes-Oxley Act of 2002, increasing the need to preserve business records in the long term, as well as the Government wide target for delivery of services electronically by 2005, compel organisations to actively manage their records properly over time. There are likely to be more pieces of legislation passed in the next few years to ensure public and private sector organisations make adequate provision for the protection and preservation of digital records such as personnel files, protection of identity and audit trails.

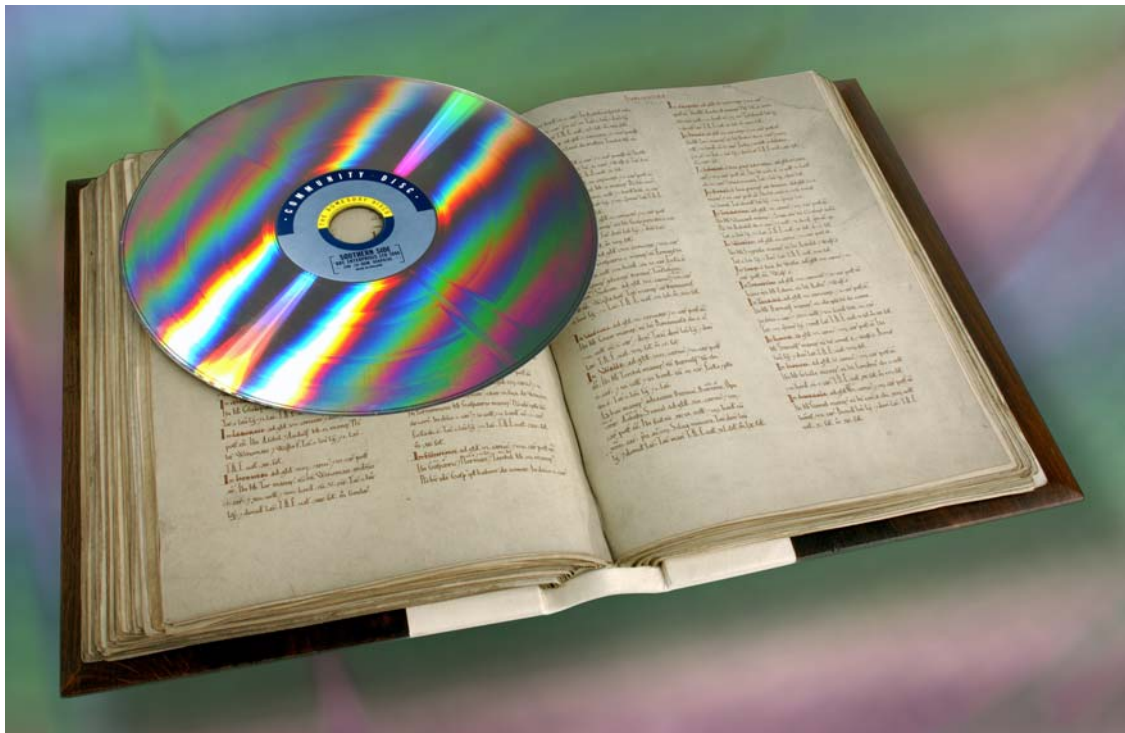
“The need to create and have widespread access to digital materials has raced ahead of the level of general awareness and understanding of what it takes to manage them effectively.” Neil Beagrie and Maggie Jones, digital preservation Handbook, Digital Preservation Coalition, 2001.

5.2 Cultural objectives

If we do not manage our digital assets well, we are heading for a **black hole in our and our organisation's and our country's collective memory.**

Digital assets are just as unique as any physical assets but they are **much, much more ephemeral than traditional materials.** The physical storage medium is mutable and vulnerable. Technological innovation moves ever more quickly and as it does so the means of storage becomes frailer and frailer. So we are ever more in danger of losing our intellectual and cultural heritage.

The Domesday Book of 1086 can still be seen and read by the public in The National Archives. In 1986, BBC Domesday was launched to celebrate the 900th anniversary of the original Domesday book with the idea of capturing a massive range of information on the social, environmental, cultural and economic make up of the UK. Contributions from researchers and thousands of school children from across the country were recorded onto two 12" videodiscs that could be viewed using a special BBC Microcomputer. The project was a landmark in terms of both its scale and its technological achievements, costing around £2.5 million. In a terrible irony, the problems of hardware and software dependence have now rendered the system obsolete. What was cutting edge in 1986 was virtually unreadable by 2000.



But this is not just about cultural memory. Without active management of our digital assets, citizens will not be able to exercise their democratic rights to access information about decisions made on their behalf and ensure transparency and accountability in public life. **It's about a healthy democracy and public engagement with it.**

"This is our intellectual and cultural heritage. As well as our books and our artworks, huge databases are the source of the aircraft in our skies, the cars in our streets, the wonder drugs in our bodies, and the list is endless."
Bruce Sterling, The Telegraph, 14 March 2005

6. What are the key business arguments?

The business arguments for digital preservation can be strategic (long term), tactical (medium to short term) and operational (short term or immediate):

6.1 Strategic

- The safeguarding of vital records is essential to business functions and decision making
- It is an essential component of good records management in the 21st century
- It is needed to ensure digital records are managed economically
- It contributes to sharing knowledge and efficient reuse of information
- It supports access to information for users, including allowing that information to be targeted
- It enables your organisation to meet its legislative obligations.

6.2 Tactical

- It ensures that the benefits of investing in electronic information management are delivered
- It ensures information is reliable and authentic
- It reduces access costs over time
- It supports efficient access to information which can build a stronger customer base
- It protects the significant investment in digital technology, both creating digital materials and the hardware and software to access them
- It supports improved security of information by controlling access issues.

6.3 Operational

- It reduces workloads in providing access to historic digital material
- It improves ability to plan for future volumes of data
- As more and more documents are created digital, it is cheaper and more practical to preserve and store them electronically than any other way.

The National Archives Digital Archive acquired 7.8 gigabytes between June 2003 and May 2004. In the following three years it expects to acquire over 10 terabytes, or over 43760% more material per annum.

7. What are the main organisational issues?

A survey carried out in December 2002, across local authorities in England, Wales and Scotland found that none had a facility for the storage and preservation of electronic records, and only 5.9 per cent had contracted support. Only 15.7 per cent of the local authorities interviewed or surveyed had decided which electronic records to keep and only 23.5 per cent had any historical electronic records held in their archives.

Source: Society of Information Technology Management

The fragility of digital assets means that they need careful management from the moment of creation and a proactive policy and strategic approach to their management to secure their preservation in the long term. What is key is that all those who have interest in the life of the digital resource are involved.

“Digital preservation has many parallels with traditional preservation in matters of broad principle but differs markedly at the operational level and never more so than in the wide range of decision makers who play a crucial role at various stages in the lifecycle of the digital resource.” Neil Beagrie and Maggie Jones, digital preservation Handbook, digital preservation Coalition, 2001

There are no easy technological solutions to obsolescence but this is not at the heart of the matter.

Neither does the existence of a functioning electronic document and records management system in your organisation mean you have digital preservation sorted.

The main issue is knowing what you’ve got, what of it to preserve and how to organize it. It’s about the classic principles of records management. You can’t preserve a record without managing it first.

“Increasingly the term ‘Information Age’ is being used to describe an era where it has been estimated we have created and stored 100 times as much information in the period since 1945 as in the whole of human history up to that point.”
Neil Beagrie, A Strategic Policy Framework for Creating and Preserving Digital Collections, July 2001

Any organisation needs to employ **the classic principles of good project management** to tackle digital preservation.

You need a **strategy, structure, processes and clear responsibilities across the whole organisation**. Put another way, the **creation of an infrastructure** to support what you are trying to do.

Clear communication and collaboration are essential as well as identifying the **skills and resources** you need. The cost of digital preservation is directly related to whether the issue has been considered from the point of creation.

Time is of the essence! Remember, the lifespan of a digital record can be much shorter than for traditional formats so there is less time to act before loss.

8. What strategies can be used?

There are three main strategies that can be used as a means of dealing with the challenges of digital preservation. Each approach has its own benefits and drawbacks.

8.1 Do it yourself

This strategy assumes that the organisation takes an “in house” approach.

Pros

- This is a good strategy if there is sufficient organisational expertise to support a cross-disciplinary team drawn from structurally distinct groups
- The strategy promotes skills sharing and team working which results in stronger retained organisational competence
- The strategy enables a greater degree of organisational control and ownership of results
- All of the above mean that this strategy delivers a good outcome in the long term.

Cons

- This strategy is not a realistic option for organisations without the capacity to support staff with appropriate expertise, e.g. small charities
- Initial stages in developing this kind of strategy can be time consuming because of the issues involved in establishing successfully functioning cross-disciplinary teams
- Staff may have difficulty adjusting to a new cross-disciplinary role that this strategy requires
- This strategy may require more flexibility in organisational structure than is achievable from “in house” resources
- It can be prohibitively expensive for small organisations.

Case Study: Dspace @ Cambridge

Cambridge University Library, in conjunction with Cambridge University Computing Service, has joined the DSpace federation initiative which will develop methods to collect, preserve, and make accessible digital content used in undergraduate programmes and other activities; and will maximises the value of existing educational assets through the implementation of a long term digital preservation strategy. Strategies are being established for deployment costs, intellectual property rights issues, organisational change, institutional governance, and policy issues, based on the proposed implementation, and these lessons will be applied to establish a digital object preservation service in Cambridge University.

8.2 Consortia

This strategy assumes that an organisation takes a partnership approach, working with other external organisations with a similar interest.

Pros

- The process of establishing the partnership secures commitment from each organisation and a clear allocation of responsibility
- This strategy partnership approach can lever economies of scale
- This strategy's collective approach can mean that an organisation can carry a higher punching power than on its own
- This strategy can enable the pooling of expertise across organisations
- This strategy can lay the foundations for longer-term collaboration in other areas.

Cons

- The initial stages of developing the partnership approach may be time consuming if you do not work in partnerships already
- The implementation stage of this strategy can become difficult if the reality of divergent aims amongst partners becomes more apparent
- Communicating across different organisations can be difficult.



Case Study: The UK Web Archiving Consortium (UKWAC) aims to expand the lifespan of website materials to a century or more. Comprising six leading UK institutions - The British Library, Joint Information Systems Committee of the Higher and Further Education Councils (JISC), The National Archives, The National Library of Wales, the National Library of Scotland and the Wellcome Trust, the UKWAC is working, with the permission of rights holders, on an experimental system for archiving selected key UK websites – ensuring that invaluable scholarly, cultural and scientific resources remain available for future generations.

The project is running for an initial period of two years, during which approximately 6,000 websites will be collected and archived. UKWAC is obtaining the permission of website owners to archive selected sites whilst working collaboratively to explore how to develop compatible selection policies and to investigate the complex technical challenges involved in collecting and archiving web material.

Each consortium member will select and 'capture' content relevant to its organisational remit. Infrastructure costs, such as software, hardware, and ongoing technical development and support will be shared equally amongst the Consortium members. UKWAC is using HTTrack – the open source web crawler to acquire files for storage. The software to carry out the archiving processes – PANDORA Digital Archiving System (PANDAS) – has already been developed and tested by the National Library of Australia and its partners for archiving Australian websites and making them accessible through PANDORA the Australian national Web Archive (see: <http://pandora.nla.gov.au/index.html>). PANDAS can be set to automatically tag, gather and prepare pages for public display. If pages are not suitable for immediate public access, due to commercial, cultural or privacy reasons, PANDAS can manage appropriate access restrictions.

8.3 Contracting out

This strategy assumes that an organisation decides to contract out or outsource its digital preservation function.

Pros

- This strategy can be a sensible option where an organisation can benefit from external expertise and a new perspective

Cons

- This strategy still requires “in house” skills to define and manage what needs to be done in terms of any legal, organisational and contractual problems and in the long term
- This strategy does not necessarily promote in house ownership of the pan-organisational processes that need to be in place
- There are limited options for contracting out in an underdeveloped market.

Case Study: UK data archive

The UK Data Archive (UKDA) is an internationally-renowned centre of expertise in data acquisition, preservation, dissemination and promotion; and curator of the largest collection of digital data in the social sciences and humanities in the UK. The UKDA provides resource support for secondary use of quantitative and qualitative data in research, teaching and learning as a lead partner of the Economic and Social Data Service (ESDS). The UKDA houses AHDS History, provides preservation services for other data organisations and facilitates international data exchange. The responsibility to build and manage research collections carries with it the obligation to ensure that these collections are permanently accessible. The UKDA is committed to the long-term preservation of its collections.

Since its inception in 1967, preservation has been a fundamental function of the UKDA. Preservation comprises a set of inter-related activities carried out by several sections of the Archive. The Systems and Preservation group has responsibility for preserving data and documentation to ensure they remain usable over time, including monitoring technological changes that will affect preservation and migration decisions.

9. What skills does my organisation need to tackle this?

Archives are matter of conscious, human choice and formal organisation. Archives can be created and maintained only by thorough, controlled, curatorial processes that take humanity's dank heaps of general blether, and enhance them. Archives add quality, trustworthiness, usability, and the power to last into the indefinite future....We would like to create automated methods to decide what data mean...but these are value judgments. Our machines lack that ability.
Bruce Sterling, The Telegraph, 14 March 2005

All those involved in the lifecycle of a record need to be involved at some stage and this requires an **interdisciplinary approach, skills exchange and understanding** most particularly between record managers and IT experts.

The speed of technological advance is such that, although **it is possible to define and advise on a minimum skills set required this will undoubtedly change over time** and, therefore require review.

This means that, although larger organisations are likely to have many of the necessary skills in house, your organisation's staff are likely to benefit from **training, re-skilling and continuing professional development** if your organisation is going to be in good shape to preserve its digital assets well.

The **core skills of records managers and archivists in selecting, identifying and managing records** are absolutely critical to an organisation's competency to tackle digital preservation issues. To do so effectively **such skills are needed at the point of creation of a record** and can be much less effective if brought in later on.

10. Resources

We do not know for certain whether the resources required to preserve a record digitally are greater than for traditional formats. What we do know is that they are significant and for the foreseeable future are strictly in addition to those required for other formats. There are a raft of factors that affect the resources needed for digital preservation in any organisation. These include:

- The **quantity** of digital records that require preservation
- The **range and complexity** of formats
- The **extent of control over the digital materials**
- The **standard of access** that your organisation needs to those records to meet its business needs
- The **degree of standardization** in the digital materials themselves
- The need to **manage technological changes** over time
- The need to **balance preservation costs of traditional and digital materials**.

It is important to recognize that the costs of preserving the records cannot be separated neatly from the cost of providing access to that record, and that those access costs can rise quickly because of the pace of technological change, and user expectations associated with that change, which may also be increased by legal requirements.

Despite the uncertainties of costs, what can be said with certainty is that the initial outlay in investment early on in the process at the selection and identification stage will save resources later on and avoid waste. **It's the principle of invest to save. The longer you leave it the more it'll set you back.**



11. Recommendations

The life of an average website is estimated to be around 44 days - about the same lifespan as a housefly. Source: UK Web Archiving Consortium

If your organisation has not already done so, as soon as possible, it should consider:

11.1 Developing a Digital Preservation Strategy

Consider the capacity of your organisation and its needs. Choose one of the strategies described in Section 8 (starting on p. 6). Put together a Digital Preservation Strategy for your organisation. Allocate an adequate budget for its implementation.

11.2 Identifying a senior management team champion

Allocate the responsibility for the implementation of the Digital Preservation Strategy to one of the organisation's senior managers, perhaps the person who already has responsibility for records management, or the senior lawyer.

11.3 Creating and maintaining an IT history for your organisation

The implementation of the Digital Preservation Strategy should start with the creation and maintenance of an IT history for your organisation, covering both content and media. This should be maintained and preserved within the structure chosen in 11.1.

11.4 How digital preservation will be considered within the context of your Electronic Document and Records Management System

Once your organisation has electronic records dealt with in a systematic way, they will be added to regularly under the terms of your organisation's records management and archiving policies. However, due to technological advances and obsolescences, your organisation must review its policies and strategies on a regular basis to ensure that its electronic records management is serving the business objectives of your organisation in the most efficient and effective way.

An important consideration is what the standards are to which your organisation is required to conform. For instance, BS ISO 15489 standard defines the characteristics of a record's authenticity, reliability and integrity, and is required by those public organisations that deal with legal documentation, such as local authorities.

11.5 If you are a funding agency, how you will ensure sustainability of projects by requiring grantees to have a digital preservation strategy

If you are distributing public money, you have a legal requirement to ensure that recipients protect their records, regardless of the media in which they are kept. Add a question to your grant documentation asking applicants whether they have a digital preservation strategy, and make sure that it is up to standard. Make one available as a template, and signpost applicants to the sources of help and advice in Section 12.

12. Sources of advice

Further help and advice can be accessed through the following organisations:

- **Digital Preservation Coalition**

The Digital Preservation Coalition (DPC) fosters joint action to address the urgent challenges of preserving digital resources in the UK and works with others internationally to secure our global digital memory and knowledge base.

Contact: www.dpconline.org

+44 (0)1904 435362

- **The National Archives**

The National Archives aims to make sure that records in a wide variety of media, from medieval parchment to digital tapes, remain in a good condition for present and future use. They advise other British institutions on best practice.

Contact: www.nationalarchives.gov.uk

+44 (0)20 8876 3444

- **National Council on Archives**

The National Council on Archives brings together the major bodies and organisations across the UK concerned with archives and their use. Copies of this document and a more detailed "Handbook of Digital Preservation" are available from the website.

Contact: www.ncaonline.org.uk

+44 (0)20 8392 5376

- **Joint Information Systems Committee (JISC)**

The Joint Information Systems Committee (JISC) supports education by providing strategic guidance, advice and opportunities to use Information and Communications Technology (ICT). It provides advice and publishes research on digital preservation including a handbook.

Contact: www.jisc.ac.uk

+44 (0)117 954 5083

- **Technical Advisory Service for Images (TASI)**

The Technical Advisory Service for Images is a JISC funded service providing advice and guidance on creating digital images, delivering digital images to users and managing both small and large scale digitisation projects.

Contact: www.tasi.ac.uk

+44 (0)117 928 7091

- **Electronic Resource Preservation and Access Network (ERPANET)**

The European Commission funded ERPANET Project has established a European Consortium, which makes viable and visible information, best practice and skills development in the area of digital preservation of cultural heritage and scientific objects.

Contact: www.erpanet.org

+44 (0)141 330 4568

- **Records Management Society**

The Records Management Society is the foremost professional association for all those involved in the process by which a company manages all the elements of records in any format or media type, from their inception/receipt, all the way through to their disposal.

Contact: www.rms-gb.org.uk

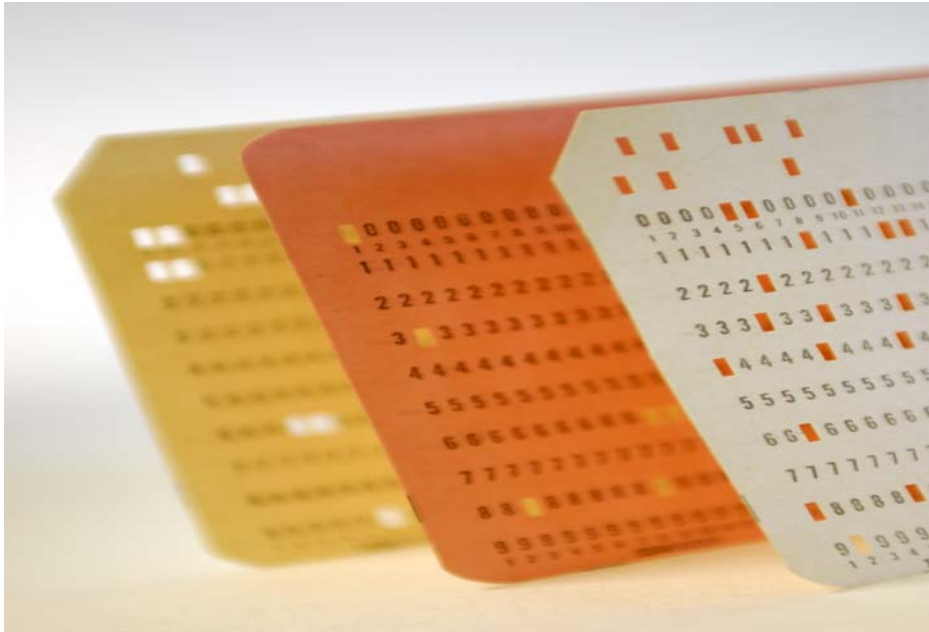
+44 (0)1494 488566

- **UK Data Archive**

See Case Study on p.10

Contact: www.data-archive.ac.uk

+44 (0)1206 873560



The National Council on Archives
C/o The National Archives
Kew
Richmond
Surrey
TW9 4DU

www.ncaonline.org.uk

September 2005



the
**NATIONAL
COUNCIL ON
ARCHIVES**