



Report of the East of England Digital Preservation Regional Pilot Project



ISBN 0-9553410-0-0
ISBN 978-0-9553410-0-7



© MLA East of England and East of England Regional Archive Council, 2006

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the Copyright owners.

Published by MLA East of England and East of England Regional Archive Council

Produced by Shorthose Russell Ltd., Norwich www.shorthose-russell.co.uk

Contents

Foreword	3
Executive Summary	5
List of Abbreviations Used	7
1. Introduction	11
1.1 Background to the project	12
1.2 Scope of the project	12
1.3 Aims and objectives	13
1.4 Methodology	13
1.5 Scope of this report	14
1.6 A note on the terminology used in the report	15
2. The Participants	17
2.1 East of England Regional Archive Council	17
2.2 Museums, Libraries and Archives Council of East of England	17
2.3 UK Data Archive	17
2.4 Hertfordshire County Council	17
2.5 Bedfordshire County Council	18
3. Outline of Processes Relating to the Administration and Preservation of Digital Records	19
3.1 Acquisition	19
3.2 Reception	19
3.3 Processing	19
3.4 Preservation and Storage	20
3.5 Access	20
3.6 User Support	23
4. Findings	25
4.1 Issues identified by the project	25
4.1.1 Acquisition	25
4.1.2 Reception	27
4.1.3 Processing and Cataloguing	29
4.1.4 Preservation	30
4.1.5 Access and User support	30
4.2 Possible solutions	31
4.2.1 Costings	32
4.2.2 Cost Models	35
4.2.3 Creating a business case	38
5. Conclusions	41
6. Recommendations	43
7. Further Reading	45
Appendices	
A. Results of Readability of Obsolete Formats	47
B. Sample Deposit Form	53
Glossary	57
Acknowledgements	61

Foreword

As is becoming daily more obvious, all organizations need to be planning for how they will keep their 'born digital' records in such a way that they can still be accessed in future. Gone forever are the days when we simply made an extra paper copy to drop in the filing cabinet!

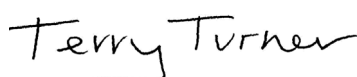
On behalf of MLA East of England and the East of England Regional Archive Council, we are today delighted to be able to publish the results of a groundbreaking study of one option for digital preservation.

The study reported on here has its origins in the Lord Chancellor's 2002 request that regional solutions for digital preservation should be investigated. Here, in the East of England, we have a head start, since the University of Essex at Colchester has over thirty years experience in this area and we knew – through Professor Kevin Schürer, Director of the UK Data Archive (UKDA) – that they were keen to help us pick up the Lord Chancellor's challenge. So, in 2004, we three partners called together representatives of all the county councils in the region to discuss future options and to identify who else might be able to come on board to test the water. In the event, it was agreed that Hertfordshire County Council and Bedfordshire County Council were best placed for a pilot investigation, which began in earnest in the winter of 2004/5.

Throughout 2005, the UKDA processed a range of obsolete or near obsolete electronic media selected from Hertfordshire's council records, to try to determine how much time and money it would take to process these records into a form in which they would remain readable and accessible for the foreseeable future. This exercise proved far more challenging than originally expected, but has, we believe, resulted in some robust cost estimates upon which we have been able to build a credible comparison, to which local authorities and other large organizations can refer to when deciding whether to 'go it alone' or to buy in their digital preservation from a third party.


So we present here a first stab at what it costs either to set up one's own repository, or to pay some other expert organization, such as the UKDA, to do it for you. We have been forced during this project to grapple with (and we hope resolve) many extraordinarily difficult issues, not least in finding a common language through which local government and higher education can understand each other in this ever-changing area, and we believe we have evolved the basis for a workable process model which will guide the next stage of our work.

We commend this report to anyone who – like us – sees no benefit in re-inventing wheels.



Terry Turner

Chief Executive,
MLA East of England



John Alban

Chairman,

East of England Regional Archive Council

Executive Summary

EERAC, MLA East of England and the UKDA are pleased to have worked together on this document, in partnership with Hertfordshire and Bedfordshire County Councils, in an attempt to promote a better understanding of the processes and costs of preserving digitalised material.

The need to preserve digital data is now widely acknowledged but due to the lack of clear working examples of digital preservation in practice, the processes and associated costs are not yet fully understood and recognised. This document reports on work funded by both EERAC and MLA East of England to support EERAC in meeting its goal of identifying collaborative solutions for the preservation of digital data. It is intended to be of interest to all archival organisations in the region, including those in the academic and private sectors, and provides guidance and recommendations for those organisations wanting to assure the future of digital records in their care.

The report aims to promote a better understanding of the processes and costs involved in digital preservation, and, to this end, is based on practical experience and a worked example involving Hertfordshire and Bedfordshire County Councils. By identifying and comparing the processes associated with the preservation of conventional and digital material, the document provides a useful base from which to inform consideration of the issues and associated costs of digital preservation.

Mindful of the solutions offered by the National Council on Archives' ***Your Data at Risk*** report, possibilities are discussed for the issues raised in relation to the different elements of the processes required for effective digital preservation. Cost models and indicative costs are included for the benefit of those concerned with budgeting for the preservation of digitised material and are set in the context of the desirability of a regional archive.

Finally, the report reaches a number of conclusions which are of relevance to all organisations with a requirement for digital preservation. It also makes a series of recommendations based on outcomes from the project, both in regard to further research and analysis, and in relation to lessons learnt.

We hope that you will find the document useful.

List of Abbreviations Used

BCC	Bedfordshire County Council
BLARS	Bedfordshire and Luton Archives and Records Service
CPA	Comprehensive Performance Assessment
CPU	Central Processing Unit
DARP	Digital Archive Regional Pilot
EDRMS	Electronic Document and Records Management System
EERAC	East of England Regional Archive Council
<i>erpanet</i>	Electronic Resource Preservation and Access Network
HALS	Hertfordshire Archives and Local Studies
HCC	Hertfordshire County Council
ICT	Information and Communication Technology
JISC	Joint Information Systems Committee
OAIS	Open Archival Information System
SFTP	Secure File Transfer Protocol
TASI	Technical Advisory Service for Images
TNA	The National Archives
UKDA	UK Data Archive
VPN	Virtual Private Network

‘We are faced at once with the necessity of choosing a nomenclature and fixing a definition. With regard to the name, we have a choice between *Records* and *Archives*.’

(Sir Hilary Jenkinson)

The definitions we have fixed upon for this report are:

Records: Information identified upon its creation as having value as evidence in ‘recording’ an activity or decision and so requiring a ‘lifespan’ for its management either through to destruction or appraisal and retention as part of an archive

Archives: Records created or accumulated by an individual, family or corporate body which are selected for permanent retention and preservation because of their value as information and evidence of some activity

Given the overlap between records requiring long-term preservation and archives requiring permanent preservation the words are often used interchangeably.

Preservation: The actions that need to be taken and managed to make sure that access is possible for as long as is necessary, whether that be indefinite or for a specific period limited by business requirement.

Introduction

Digital records have been part of society for over fifty years. In the last fifteen years, with the explosion in the use of computers in every walk of life and increased speed of change in both software and hardware, the particular preservation issues of records in digital form have been widely realised. It is no longer possible to simply keep the storage medium safe from water, fire and rodents and expect to be able to read the materials stored on it in a hundred years time. To preserve records in digital form, positive action must be carried out on a regular basis.

Accessing digital records relies on the intervention of machines capable of interpreting the way that the information has been formatted and stored. As software and hardware become obsolete, *either* examples of machines running the programmes need to be preserved or emulated *or* the record needs to be migrated to a currently supported format. Digital storage media are made of a mixture of complex materials including dyes, plastics and coatings; these may become chemically unstable or be damaged by scratches, dust or magnetism. Ensuring that records are stored on more than one medium acts as a safeguard against decay. These preservation measures need careful management if the records are to survive in a readable format.

Archives are those records that have been selected for permanent preservation. Even records that will not always be legally or historically valuable may have legal retention periods requiring that they be kept in useable condition for up to 75 years or even longer. Therefore it is important to provide for the preservation of both records and archives. However as a general rule the preservation measures that are required will depend on the length of time that records are likely to be kept.

Organisations such as government - local and national, businesses - large and small, charities, societies, and individuals all produce records and archives. Different organisations have reached various stages of development in the management of their digital records. For many, dealing with the preservation of their own digital material may appear a very technical and formidable task. Outsourcing in part or as a whole to specialist organisations is therefore a solution to be considered. The feasibility of providing this type of specialist service on a regional basis in the East of England was the focus of the Digital Archive Regional Pilot (DARP).

A considerable amount of research, training, advocacy and discussion on digital preservation issues is currently being undertaken by organisations such as the Digital Preservation Coalition (DPC), The National Archives (TNA), National Council on Archives (NCA), Joint Information Systems Committee (JISC), the Technical Advisory Service for Images (TASI) and the Electronic Resource

Preservation and Access Network (*erpanet*). Much of this work is theoretical with few concrete answers (*Report on the Digital Curation Centre and digital Preservation Coalition Workshop*, p5). It is hoped that the DARP will be able to contribute to this work and, with the involvement of UKDA, a long established digital repository, be able to add practical experience of procedures and costs to the debate.

1.1 Background to the Project

In *Eastern Promise: a Strategy for Archival Development in the East of England*, published in Spring 2003, the EERAC identified one of its ongoing activities as to 'seek to identify collaborative solutions for the preservation of digital data' (p33).

Following work already undertaken by the UKDA for EERAC, the MLA East of England agreed to incorporate a pilot project into its business plan.

In support of these goals, EERAC and MLA East of England co-financed the Digital Archive Regional Pilot.

It was felt that the project was likely to be of considerable interest to all organisations responsible for managing records and archives, but especially the larger local authorities who not only produce a large number of public records and archives themselves but also make the private archives of other people and organisations available to the public via the network of county record offices. On this basis, all county councils in the region were invited to attend an exploratory meeting in February 2004. In early summer 2004 they were sent a questionnaire devised by the UKDA to establish which of them were likely to be in a position to participate in a test bed pilot project: Bedfordshire and Hertfordshire County Councils were identified by this process.

1.2 Scope of the project

County councils potentially have two sources for digital records and archives:

- those produced by their own organisation as public authorities;
- those produced by other – mostly private - organisations and individuals and deposited with the county record office for long term preservation and access.

Records selected for long term or permanent preservation at the point of, or soon after, creation would probably *either* be managed in-house *or* be transferred to another, possibly regional, specialist repository while still in a current format. Those digital records whose archive potential had not been considered early on, or that had been produced before the issues of digital preservation had become apparent, could be in obsolete formats with little or no supporting information. Archives received from external organisations and

private individuals could be in formats never used by the receiving repository and therefore possibly inaccessible to them, so restricting or even denying the scope to appraise and catalogue.

The scope of the project therefore was to consider the preservation of both current and obsolete formats.

It was agreed that the project should exclude moving image and sound files (although still images should be included). It was not possible to make any thorough investigation of the skills or training issues that are likely to arise, or of legal issues.

While the initial pilot project involved only local authorities, the results of the pilot will be of much wider interest and will be disseminated to all archive organisations in the region, including those in the academic and private sectors.

1.3 Project Aims and Objectives

The project aimed to investigate the issues which would need to be addressed in the establishment of a regional archive and records management service for digital records, and, more specifically, to assess the costs of providing such a service.

The objectives were to:

- identify problems and suggest solutions to sustainable management of digital records and archives;
- develop a strategic report as a base for a business case for a digital preservation strategy for the east of England.

1.4 Methodology

To achieve the project objectives the pilot:

- investigated the costs and options of migrating data from obsolete formats and media;
- discussed metadata standards;
- developed sample depositor forms for dealing with intake of digital records;
- Information and Communication Technology (ICT) infrastructure;
- investigated end user processes for retrieval;
- looked at the variations in processes and costs of dealing with conventional hard copy and digital intake.

1.0

Issues were discussed by members of the project via a series of meetings. To investigate the costs and options for migrating data from obsolete formats and media, a series of twelve examples of superseded media were provided by HCC. For the purposes of the project, the most obsolete materials possible were deliberately sought.

To aid the analysis and reading process, a series of questions designed to be understood by the non-specialist were drawn up by UKDA to determine as much information about the media as possible. These questions worked in a progressive form narrowing down to finer details. They included:

- When were the data written?
- Are you supplying the only copy?
- Are the same data available on a different medium?
- What was the environment like in which the media were stored?
- What is the exact medium type?
- Can you supply details of all labels on both sides of the medium (e.g. 3.5" Floppy: Maxell 2D HD Batch code 123456)?
- What type of PC were the data created on (e.g. MAC, IBM PC, BBC etc)?
- What format are the data in?
- What programs were used to create the data (ideally the exact version)?
- What are the data (e.g. text, database, spreadsheet)?
- Who created the data and are they still within the organisation?
- Is there any chance that the machine on which the data were created is still available?

Although not an exhaustive list of questions, they were felt to represent the types of information that need to be collected alongside the media to maximise the potential for efficient and effective subsequent recovery and transfer of data from old or obsolete formats.

The UKDA then attempted to recover the data from these media; writing them to a current device, creating a backup of the restored file on compact disc and preparing the new file for preservation by checking the file and migrating it to a suitable preservation format. Appendix A shows the results of this process for each medium. To produce detailed costings relating to the time spent by UKDA on preparing the data supplied, timesheets were used by all UKDA staff involved in the exercise.

1.5 Scope of this report

This is the final report for the DARP. Although the project attempted to address the aims and objectives set out in the Project Initiation Document of November 2004 a number of factors have influenced the scope of the final report:

- Some aspects of the project, for example the recovery of data from obsolete

formats, took longer than expected. Although useful lessons regarding the cost and time required were learnt, this did hold back progress on other issues.

- The project underestimated the differences in technical language and the resulting difficulties in communication between those specialising in management of digital archives and those administering archives produced on conventional media.
- Considering how to deal with digital records derived from both corporate records management and private sources has been a consistent theme throughout the project. The Office of the Deputy Prime Minister's ***Guidance on 'Proper Arrangements' For Archives*** advocates integrating archives and records management services. During the project, national developments have taken place and influenced thinking on this point. Within Comprehensive Performance Assessment (CPA), which has become a major driver for local government, archives services have become a defined part of the culture block, newly-formed as part of the CPA. Records management, however, is associated instead with the corporate resources block of CPA and has been the subject of discussions involving TNA and the Audit Commission based on Value for Money, a strong assessment criterion underpinning CPA. Presuming that local authorities would, by necessity, have to ensure the preservation and accessibility of their own records, not least because of the Freedom of Information Act and related legislation, the project began to consider private digital archives intake in a very broad sense as separate in terms of cost implications. This opened up issues about the possible spotlighting and scrutiny of the justification for collecting private material per se. The new backdrop is an emerging emphasis on providing evidence of outcomes, such as generic learning through access to archives documenting communities, linked to an increasing regional emphasis on striving to make cultural services contribute together towards the shared national priorities in CPA.
- During preparation of the report a number of other publications became available and where possible these were taken into account.

1.6 A Note on Terminology Used in the Report

In 2002 the Open Archival Information System (OAIS) reference model was approved as international standard ISO14721. Developed to provide a standard that would cut across domains, the model uses descriptions 'expressed in terms of a well-defined set of concepts and terminology transcending, yet mappable to, domain-specific vocabularies' (Lavoie, ***The Open Archival Information System Reference Model: Introductory Guide*** p5). During the DARP project it became apparent that whilst UKDA staff were familiar with the OAIS terminology other members of the project found it unfamiliar and therefore difficult to relate to

1.0

their understanding of the concepts involved in digital preservation. For this reason the OAIS terminology has not been adopted throughout this report. Instead a vocabulary that could be used comfortably by all parties has been applied. A glossary has been provided in which terms used in this report have been identified as OAIS terms or, if appropriate, have been mapped to OAIS equivalents.

The Participants

2.1 EERAC

EERAC is one of the network of English Regional Archive Councils, created by the Department of Culture, Media and Sport, in 1999 to provide archives with a regional voice. It provided the major financial support to the DARP project.

2.2 MLA East of England

MLA East of England, established on 1 April 2003, is the regional agency responsible for developing an effective regional infrastructure for museums, libraries and archives in the counties of Bedfordshire, Cambridgeshire, Essex, Hertfordshire, Norfolk and Suffolk and the unitary authorities of Luton, Peterborough, Southend on Sea and Thurrock. It acted as the sponsor of the project.

2.3 UKDA

Established in 1967, UKDA provides data and related information services, principally acting as a national service provider, acquiring, preserving and disseminating data for research and teaching to Higher and Further Education. It acquires digital data through deposit. Once a collection has been accepted for deposit, the integrity of the data and metadata are checked, a catalogue record is produced and the data and metadata are converted to appropriate preservation and dissemination formats. The UKDA has a preservation strategy based upon open and standardised file formats, data migration and media refreshment. Data are currently disseminated by one of four methods: direct data download; FTP download; online data analysis software tools; bespoke writing of hard physical media. User support is provided via the UKDA helpdesk. In 2005 the UKDA was designated a legal place for deposit by TNA, the first organisation to be specifically recognised as a place of deposit for digital records.

The UKDA provided specialist knowledge and skills to the project and offered to act as a 'test bed facility' to explore the feasibility of a regional digital archive.

2.4 Hertfordshire County Council

Hertfordshire County Council (HCC) was identified by the survey carried out in early summer 2004 as the local authority best placed to enter quickly into a pilot project. HCC were in the process of choosing an Electronic Document and Records Management System (EDRMS) for the county council's born digital documents and had identified the need for a service which provides digital preservation, legal admissibility, business continuity management, on demand recovery and remote access to records. HCC runs the Hertfordshire Archives and Local Studies service (HALS), which in future will be required to take in

2.0

archives from external bodies in digital form.

2.5 Bedfordshire County Council

Bedfordshire County Council (BCC) has worked with the Local Government Association and The National Archives in developing its strategic approach to Freedom of Information and information management. At the start of the project it was well advanced in developing a specification for an EDRMS by joint work between the records management service and the corporate information and communications technology strategy team. The Council is the parent body for Bedfordshire and Luton Archives and Records Service (BLARS). This service, like HALS, wished to explore ways in which archives in digital form, both generated by its own Council and deposited by external bodies, could be preserved and made accessible to the public.

Outline of Processes relating to the Administration and Preservation of Digital Records

The processes for acquiring and administering digital archives do not differ greatly from those for non-digital archives except potentially in:

- the order in which certain elements are tackled;
- the time scales allowed for the processes to take place;
- the need for a proactive rather than passive approach to preservation during storage;
- the need to be equipped to assist the user not only in understanding the contents but also in the technicalities of enabling the user to view the record in the first place.

3.1. Acquisition

The process whereby a repository negotiates for and acquires records for the purpose of long term preservation and to facilitate access to them. This includes:

- pre-deposit evaluation of the materials (evaluating their fit with the repository's collections policy, and the potential resource implications of taking the materials into the repository's custody including the need and scope for selection and appraisal);
- licensing and copyright agreements;
- completion of deposit forms (providing details on provenance, technical details of media and file formats, details of associated metadata and instruments);
- checking that all basic elements are in place and, where possible, liaising with the record producer to ensure capture of as much relevant background information as possible.

3.2. Reception

The initial process of gaining management control over deposited records includes:

- secure transfer of digital records to repository undertaking ingest work;
- checking the integrity of the data and metadata;
- risk analysis to assess vulnerability of data (for example risk arising from obsolescence or obscurity of software or media) and need for specialist user support;
- conversion of data and metadata to preservation formats;
- conversion of data and metadata to dissemination formats.

3.3. Processing

Achieving full understanding of the content of the material and creating resources to enable access to it includes:

- disclosure control checking (for compliance with the Data Protection Act etc.);
- production of a catalogue record and associated finding aids.

3.0

3.4. Preservation and Storage

The process required for maintaining records for the long-term includes:

- secure transfer of digital records to repository undertaking preservation (if different from that undertaking ingest work);
- storage of records and associated metadata across multiple media types;
- storage of records in multiple secure environments;
- on-going preservation 'watch' on fashions, developments and obsolescence in IT;
- refreshment of media types;
- fixity checks by periodically running programmes to check for losses and changes to data;
- migration of file formats (as appropriate).

3.5. Access

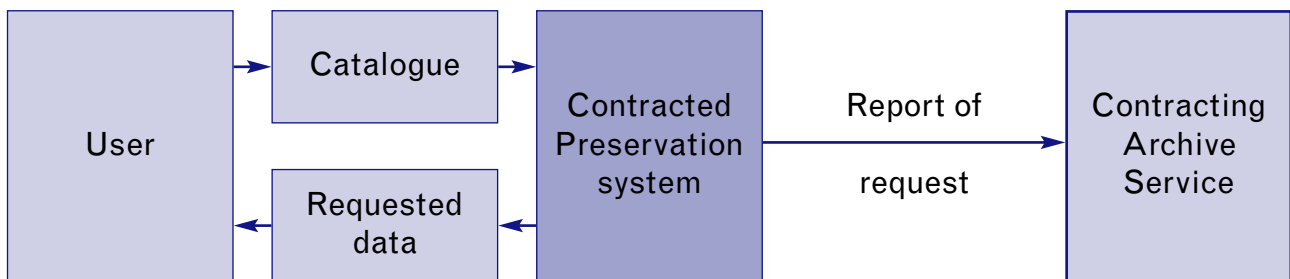
Identifying the sub-processes of this main process is not straightforward since a variety of models could be developed to enable users to access the records. Each of the possible options will have resource implications.

Requirements:

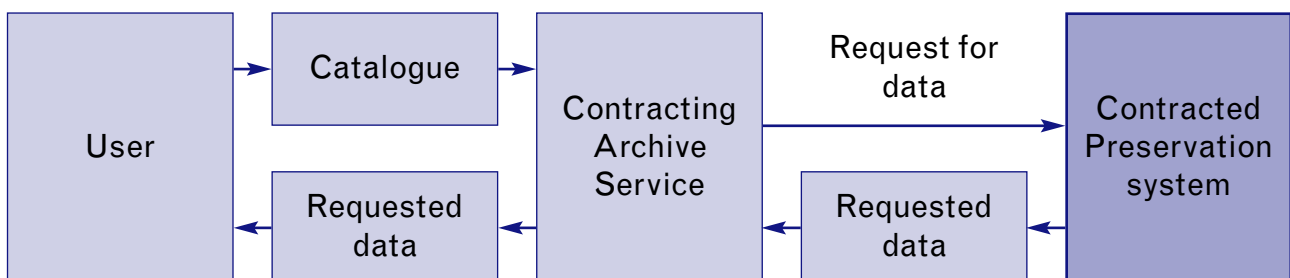
- ability for users to move directly from the catalogue entry to a dissemination version of the electronic record;
- ability to deliver digital records in multiple file formats;
- ability to deliver the digital records on multiple media;
- an important additional option may include restrictions on access - the need for an authentication front end to the access system (i.e. electronic records can only be accessed by 'registered' readers/users).

Options for Access Methods from a Contracted Out Preservation Service

1. Reader/user is routed straight from local (or remote) catalogue to external digital repository's preservation system front end, which delivers a dissemination version. Such a system could have an authentication front end and would need to record use so that usage reports could be passed back to the contracting archive service.

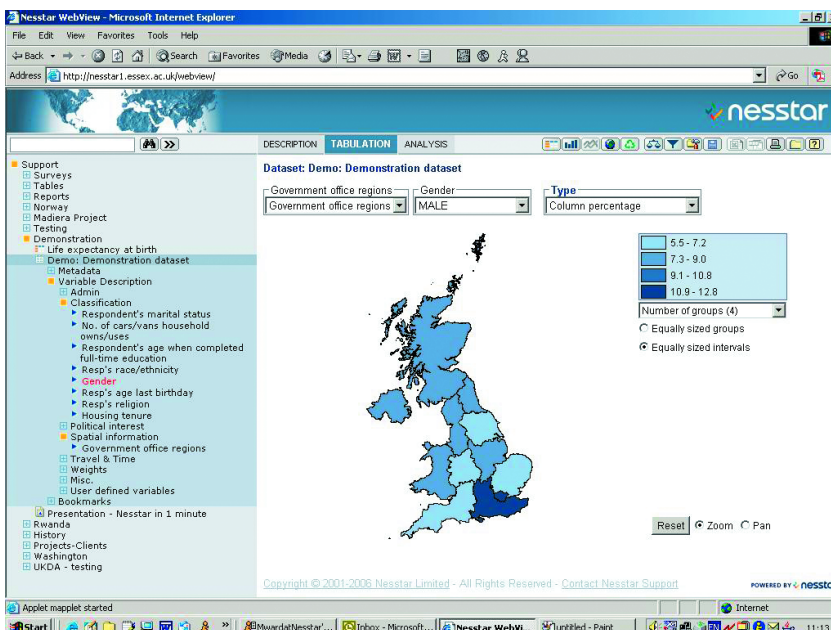
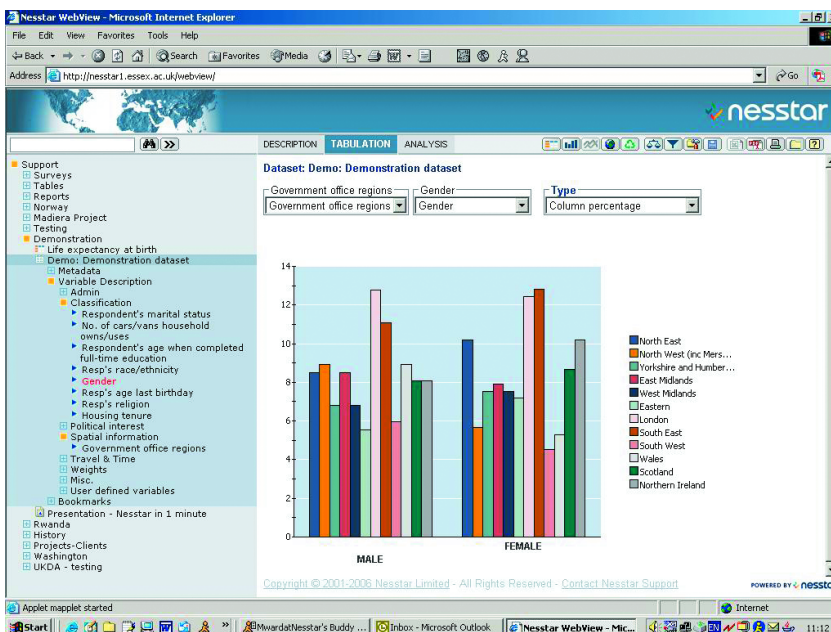


2. Request for copy of record is passed from the contracting archive service to the external digital repository. Copy of record is passed securely to requesting archive service (either electronically or on hard media), which passes this on to reader/user directly (via hard media) or indirectly (via provision of local terminals).

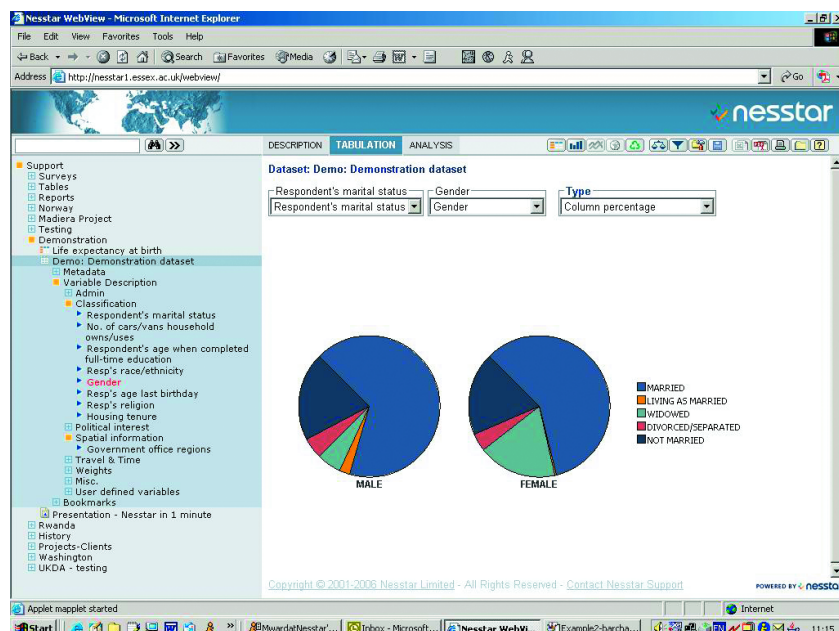
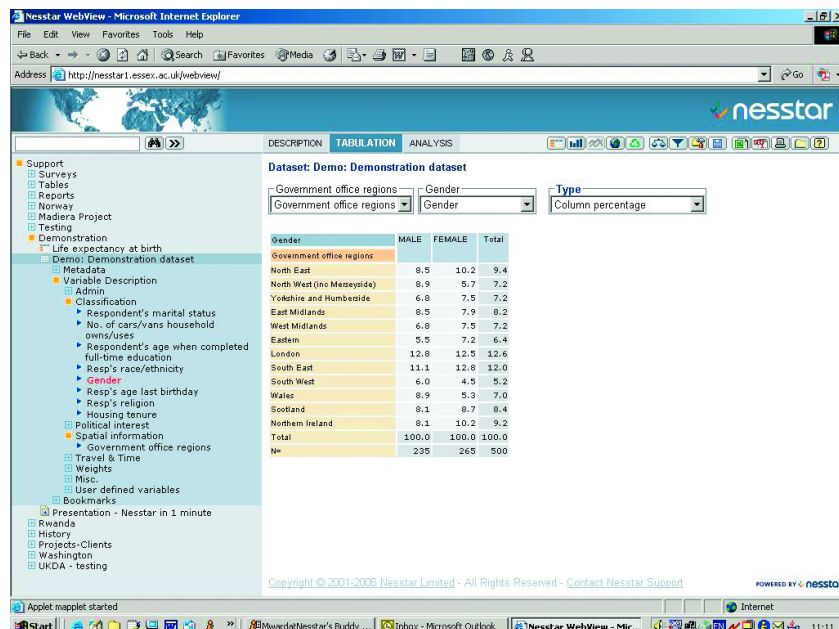


3.0

In thinking about these options (and costing them) it should be realised that many users will not want access to the digital record *per se*, but to the information, or part of the information contained within it. Thus in some instances, especially databases, access will mean providing 'views' of the data or records via online retrieval/browsing systems. In the screen shots below, each screen represents a different 'view', or means of presenting the same underlying data to the user.



3.0



'Views' of data all created from the same digital record but arranged by different parameters.

3.6. User Support

Providing mechanisms to help users access and understand the records includes:

- technical support (how to read the record on the user's computer system);
- content support (how to use and interpret the record).

3.0

Case study

A collecting repository is offered two collections of family correspondence – collection *A* consists of letters written in ink on paper, collection *B* of e-mail files. In both cases the repository needs to negotiate with the depositor, evaluate the collections against its collecting policy, consider copyright and data protection, complete deposit forms, arrange transfer, check the material on receipt, assess for conservation and cataloguing priorities, catalogue, store and possibly make surrogates for public use and dissemination.

Whilst not ideal *A* could be acquired, accessioned and then put into storage to be assessed for conservation and cataloguing at a later date. If this strategy was used with *B* the risks of increasing the cost of conservation and the risk of even losing the ability to catalogue and access the collection are higher. Therefore assessment of cataloguing and conservation priorities may need to be done earlier and more thoroughly in the processing of *B* and the time scales between acquiring, conservation and cataloguing may need to be much shorter than those permissible for *A*.

Provided the environmental conditions of the stores are monitored to ensure optimum temperature and humidity and security and maintenance checks are maintained, *A* will last for many years, possibly centuries, without needing further interference. *B* will need periodic checks to enable action to migrate or refresh the data as software and hardware become obsolete or unstable.

Access to *A* may be provided in a number of ways: for example, looking at the originals or surrogates in a public searchroom or, having copies sent to researchers by post or e-mail for them to study in their own homes. Access to *B* may similarly be given either via a terminal in a searchroom or by supplying a dissemination copy of the material to a user.

While accessing *A* some users may need help with reading the handwriting or need to ask about the context to gain better understanding, for example 'Was Queen's Park, Bedford a desirable place to live?' While using *B*, the user may also need more technical support such as whether their own computer has the capability required to read the documents.

Findings

4.1 Issues identified by the project

A key task for the project was to consider the feasibility of a regional repository for both current and obsolete formats by investigating such issues as cost implications, methods of transfer and methods of public access and retrieval.

4.1.1 Acquisition

Issues relating to the acquisition of digital records that became apparent during the project were:

- difficulty in estimating both nature and volume of likely and expected intake;
- potential difficulty in evaluating material before deposit;
- need to gain the correct information and permissions by an appropriate deposit form.

The source of digital material might be one's own institution or, for collecting repositories such as UKDA, HALS and BLARS, both the parent institution and, much more unpredictably, a wide range of external sources. Whilst it is generally possible to predict the type and quantity of material from the former, it is virtually impossible for archive services such as HALS and BLARS to predict the quantities, quality or type of material that they may be offered, or that they may wish to acquire, from sources outside their parent organisations. No research has been done on this matter and any scaling up of estimation would depend on the relative level of 'activeness' with which collection was pursued, the breadth of the collecting remit, and the nature of the intake.

Traditionally, local authority archive services have collected material from the various tiers of local government, the established church, businesses, charities and the voluntary sector, landed estates and families, a wide range of community groups and individuals. All these sources are likely now to be producing records in digital form. So far, however, the amount of digital material offered to archive services has been negligible. For example, in the experience of BLARS, material has been limited to a very small number of individuals or local history groups offering collections of scanned images or databases and spreadsheets containing transcripts or catalogues of non-digital material. Many of these have been in proprietary, sometimes obsolete formats. Their creators or depositors have had little or no understanding of the issues of digital preservation.

Without any accuracy around estimation of the quantity, quality or format of potential deposits, it is difficult to predict what level of service or funding might be required for preservation and access; or whether the service will be possible through in-house provision or only through contracting out or a consortium solution.

4.0

The drive for e-government may lead parish councils to be one of the first major and common depositors to offer digital material for deposit. However, based on discussions between BLARS and a small random selection of parish clerks, it is believed that parish councils still rely heavily on printing out to paper and that they use products similar to those in use in the County Councils such as Microsoft Word, Access and Excel but only in a limited support role to paper record keeping. Therefore it may be some time before they wish to deposit material in digital form. When they do, County Council's may be able to manage the records offered to the archives service in the same way as their own digital records, possibly without significant additional cost, provided the material has been properly cared for in the time between creation and deposit.

In some cases repositories may be offered, or receive, material without supporting metadata. If the depositor is not the original creator the gathering of metadata may prove difficult if not impossible. It is clear that the longer the time lapse between record creation and creation of the metadata the greater the risk of not being able to capture all the necessary details.

If the material is held in an obsolete medium or was created in obsolete software it may not be possible to access the data. The problems of recovering the data from some of the obsolete formats supplied by HALS, without metadata about the systems and media used to produce and store them, demonstrated the need for adequate documentation. Neither HCC nor BCC could immediately and easily supply details of the systems and software used by their organisations over the last twenty years. Consequently, they gathered information from their ICT staff so as to construct timelines which would help in the provision of such metadata. However, these showed that data migration had only been consistently applied where mainframe and large PC network applications were involved; smaller and more specialist applications had tended to be overlooked and not picked up corporately.

It was felt that this form of would be an essential first step with any large deposit of poorly identified digital material received from an organisation. Without supporting information and/or without the facility to access and understand the material, it would be difficult to evaluate whether acquisition is appropriate. Moreover, the cost of recovering the data from the media purely to enable pre-deposit evaluation could be prohibitive and lead to material being rejected which, had it had sufficient supporting metadata, would have been accepted.

The need to gather metadata requires carefully drawn up deposit forms designed to ask the right questions. Much work has been undertaken by other projects to develop standardised metadata as a means of discovering online

resources (Jones & Beagrie, *Preservation Management of Digital Materials: A Handbook*, p115). Therefore, the project focused on the types of metadata required for transfer, preservation and access using a typical offer of private deposit of records to a local authority archives service.

UKDA already have deposit forms for transfer of digital data. HALS and BLARS also have deposit forms and written terms of deposit for the archives they collect. These existing forms were used to create a suggested deposit form for digital material received by a local authority, Appendix B. The form is designed to prompt the collection of metadata that not only aids the restoration and migration of data but also informs the selection process and the compilation of catalogue records – in general, the effective administration of the material. It includes similar questions to those used in assessing the obsolete media supplied by HALS (listed in 1.4 above).

Metadata would also need recording in relation to future access and use such as licensing agreements for the management of copyright and intellectual property rights since both preservation and access require copying of the data. UKDA already has template licensing agreements that could be used to aid the drawing up of agreements for use between a local authority and its depositors.

4.1.2 Reception

The issues were:

- specification of secure methods of transfer;
- risk assessment to identify priorities for processing due to data vulnerability and need for specialist support;
- difficulties of data recovery for conversion to preservation formats.

The transfer of digital archives between a depositor and a local authority archive service or an archive service and a specialist repository could take two possible forms:

- transport of the physical media;
- electronic transfer of the data.

Data received in obsolete formats or storage media would require the physical transfer of the actual media on which the data are stored. Suitable packaging to protect the media from a metre drop onto a hard surface and a secure courier service would need to be used. The examples of obsolete media in this project were delivered by hand and therefore packaging requirements and associated information were not tested or costed.

4.0

Electronic transfer depends on the availability of a network connection with sufficient bandwidth and security and with decryption facilities at the receiving end. The options for such a connection were discussed.

Electronic transfer would require a network connection of sufficient bandwidth to enable transfer in a reasonable time period which would be defined as part of the service level agreement. The connection could either be via a public or private network. The latter would be more secure but would also be more expensive and could exclude certain monitoring systems.

If using a public network, either a virtual private network (VPN) or a secure file transfer protocol (SFTP) could be used to make the link secure. VPN uses very robust encryption and monitors traffic to ensure packets do not get altered while travelling across the public network. However it is very CPU intensive, so serious use necessitates the use of VPN hardware. Where electronic transfer is possible a VPN over a public network using 128 bit encryption with appropriate monitoring systems at both ends is likely to be the most effective and cost efficient method assuming the appropriate technical infrastructure is already in place.

File metadata including details of checksum value for each file and the number of files expected would need to be created to accompany the data so that they could be validated at the receiving end. Applications are available to do this for small sets of files that do not change. UKDA has developed an in-house system that can handle small changes in a large number of files and this also acts as a validation system.

The reception process should include a risk assessment of the deposited material. Factors to consider include:

- the physical quality of the medium or media on which the data is written;
- the availability of hardware on which to read the medium/media;
- the currency of the software in which the files are written;
- the quality and appropriateness of accompanying metadata;
- the content and future value of the records;
- the existence of duplicate copies of the material (either on other media or in other file formats or paper records).

The attempts to recover data from the obsolete media provided by HALS demonstrated how the first four of the above list impact upon the successful and cost effective recovery of information. If media are damaged or obsolete, or data cannot be read easily because of either a lack of information about the system or software used for production, there will be additional resource implications in

continuing with the preservation process since specialist services are likely to be needed.

The last two points also need to be considered since duplicate copies on different media may mean that specialist services can be dispensed with or, if the content is not judged to be of sufficient long-term value taking account of the resources needed for preservation, the process can be halted at this stage.

Tools such as TNA's PRONOM can help in assessing the currency of software. IT departments will have some knowledge of software and hardware obsolescence but may not have the ability to read or assess the stability and completeness of data in many different forms. In addition, any repository that undertakes the reception of digital archives would need to demonstrate expertise in addressing reception issues for both non-numeric and numeric, structured and un-structured material. As in the example of the UKDA, this could be achieved by the application of procedures for the intake of material, security (both during and after transfer) and for the assessment and management of received media in preparation for preservation.

4.1.3 Processing

The issues identified were:

- how catalogue entries would be integrated with systems used by archive services for traditional archives;
- what level of detail would be required for finding aids.

Public access requires being able to find what is available via catalogues and other finding aids such as indexes. UKDA produces catalogues for digital records which, in addition to information regarding content, also require information on file size, structure, and methodology. See <http://www.data-archive.ac.uk/search/searchStart.asp> for example. Both HALS and BLARS are used to cataloguing archives in traditional media and would require any catalogue of digital media to be accessible via their existing finding aids for other collections, such as BLARS' online public access catalogue. Since international standards exist for cataloguing archives, this should not be a problem.

However digital archives may be received long after they were produced without accompanying information that enables cataloguing to take place. HALS and BLARS agreed that in principle they would like to produce the catalogue entries. However, it was acknowledged that in some cases cataloguing would not be possible until after the data had been transferred to an external repository for some combination of data recovery and conversion to preservation format.

4.0

Therefore a cost model that allowed for the variations in service needed from an external support repository was required.

Digital material is often created and designed with intrinsic search qualities and functionality and easily supports the production of more detailed finding aids than is possible with non-digital archives. For example cataloguing of a hardcopy minute book as part of the archives of an organisation would traditionally have described it very briefly with the details of the organisation which created it, the date range it covered, whether it was indexed, and possibly some examples of the subjects covered. A record of minutes in digital form could have keywords assigned to it or could be made fully freetext searchable. These possibilities may tempt archive services to improve finding and searching aids beyond what has been traditionally done, as is already happening with keyword search functionality with digital forms of catalogue. However, the potential cost implications need to take into account such factors as predicted audience requirements in order to make them justifiable.

4.1.4 Preservation

As part of the project, one collection was processed by the UKDA to the point at which it was ready for inclusion on its preservation system. Actual preservation was not undertaken but the work demonstrated the practicability of including local authority records within the UKDA preservation system.

Two key issues were identified by the project team in relation to digital preservation:

- the need for investment to establish a secure technical infrastructure for digital preservation;
- the need for a more proactive and interventionist approach to digital preservation than is necessary for the preservation of paper records.

Digital records will require a repository that can provide the technical infrastructure and environment suitable for their preservation. Meeting mandatory requirements of the OAIS Reference Model and having a formal preservation policy, as does UKDA, may become the types of base requirement for accreditation to manage digital records effectively. Local authorities holding public records would also need to consider whether the repository is a TNA designated place of deposit.

4.1.5 Access and user support

The issues identified were:

- the need to provide dissemination copies of the data to users;
- the need to provide technical support for the use of the dissemination copies.

Once someone has located archives in digital form that they wish to see, they need a way of viewing them via dissemination copies. UKDA currently disseminates data in one of four ways (see 2.3). None of these was tested during the project, although the data which were recovered from the obsolete media were migrated to present day media. It was considered that there are two possible models that could be used:

- a specialist repository could supply copies of the data on present-day media to the transferring archive who would then deal with further dissemination and user support;
- the transferring archive could contract the specialist repository to deal with all dissemination and user support as well as storage and preservation of the original data.

It was felt debatable as to whether it was cost effective for archive service staff to develop the technical knowledge to provide user support on accessing data although it was felt that they could provide support on its content and context, as they do with traditional archive materials. If storage, preservation and dissemination are provided in-house there is a question of the level of technical support available from local IT support which, as Deborah Tritton found, varies greatly from council to council (forthcoming *The archives of the Future: Planning for Electronic Records - a survey of South West Local Authority Archive Services*, 2006).

4.2 Possible solutions

The recent National Council for Archives publication *Your Data At Risk* identifies three options for organisations seeking solutions to digital preservation

- In-house
- Contracting out
- Consortia

Different organisations will have different in-house capabilities for dealing with the preservation of digital archives. They will also have varying levels of funding available to buy external services and different qualities and quantities of material to deal with.

Discussion of the options led the project to conclude that in-house or contracting out were the options most likely to be used and that organisations such as HALS and BLARS might find a combination of the two the most practical solution. It was felt that a consortium solution, even at regional level, was unlikely unless it was used as a bargaining tool for contracting out.

4.0

Large organisations, such as County Councils, particularly those running an EDRMS and with integrated archives and records management services, are most likely to be able to carry out many aspects of digital preservation in-house. Councils will have to make effective access provision for their own digital records as part of good corporate governance and compliance with legislation such as the Freedom of Information Act. However, they may wish to contract out some aspects of the management of their digital records. With the intake of privately deposited records, for example, data recovery from obsolete media may be a particular issue.

Some organisations, such as collecting repositories, have special requirements to make data from different sources available to the general public. They may find that contracting out not only data recovery but also other aspects of preservation, dissemination and user support is the best option for at least that proportion of material created in formats not supported by their own IT department.

Small organisations without in-house facilities may wish to enter into a consortium with other similar organisations. This could create economies of scale to enable the setting up of a joint digital preservation system but may also mean that an organisation 'can carry a higher punching power than on its own' when negotiating to contract out (*Your Data At Risk*, p8).

To provide an in-house or combined in-house/contracted out solution, it is necessary to understand what is required to carry out the processes involved and to have some idea of cost. These are outlined in 4.2.1 below.

Due to the different needs of the various archive services, a number of different cost models, from basic recovery to full preservation management and dissemination, would be needed if a regional repository were to provide services to the full range of archives across the region. Two possible cost models are outlined in 4.2.2.

4.2.1 Costings

Preservation System set up costs

The project sought to identify costs for setting up and running a regional digital repository based on facilities comparable to those of the UKDA.

To recreate the current UKDA preservation system of hardware, software and maintenance would cost in the region of £68,000 based on a five-year life-cycle renewal. Staff costs to keep the system working and infrastructure costs for establishing and running a secure machine room with adequate air conditioning, fire protection and security are estimated to bring the cost of establishing a

preservation system and running it for a five-year period to a figure in the region of £525,000.

Buying into a pre-existing system would reduce the set up costs. Collaborating with other county councils and organisations would help spread running costs. The UKDA undertook a theoretical exercise to determine how much it might cost for others to 'buy into' the existing UKDA preservation system and preservation infrastructure. To do this certain assumptions had to be made regarding the amount of data and the timescale involved. These assumptions were that initial contracts would be on a fixed three-year basis and that allowance would be made for up to one gigabyte of data to be stored on the system. Based on these assumptions the UKDA calculated that the total cost for 'renting' up to one gigabyte of space on its system for three years would be £35,500. This figure includes all hardware, software, media, maintenance, infrastructure and associated staff costs.

Whilst some costs can be calculated per unit of throughput based on the percentage of time spent on certain activities, such systems have certain fixed costs that have to be paid regardless of the amount of throughput.

Requirements for Individual Processes

The project considered cost implications of individual processes to give some idea of the resources required to provide these in house or by contracting out.

Acquisition

The requirements for acquisition are simple and straightforward. An on-line downloadable acquisition form with checks built into it can be used to ensure that metadata is collected which will assist cataloguing. A tracking system can be easily set up to show the progress of the data through the system.

Reception

Electronic transfer is a fixed cost, courier transfer is likely to be variable depending on quantity. Once received programmes can be used to assist with integrity checking.

Data recovery and migration based on the twelve examples of obsolete media used in the project were costed by the UKDA. Staff costs at face value and costs incurred by sending selected media to an external specialist media conversion company produced an average cost of £115. Adding related indirect costs incurred by the organisation to this figure, the total data recovery and migration costs could be expected to be in the range of £155 to £260 per media transfer, depending on whether more specialist media conversion was necessary. These costs are one off costs related to preparing data for preservation. The UKDA felt that this was comparable to its experience of

4.0

handling the main media types in general use from the early 1980s to mid 1990s and now generally considered obsolete. However, the UKDA also felt that unless a specialist repository undertook additional services to offset and spread the staff and related overhead costs these prices could not be maintained due to the lack of adequate economies of scale involved.

The main driver in the cost variations was not the medium in question but rather the physical state of the medium, the information available about it and the content and structure of the digital records they contained.

File size makes little difference to cost. Although the operation will take longer with a larger file, the 'hands on' staff time required is more or less the same. One exception to this is punch cards, as it is not recommended to leave the machine which is used for 'reading' these running unattended. The UKDA's experience suggests a fixed set up cost for reading punch cards of approximately £25 regardless of the number of cards read, in order to set up and test the reader, and an additional £8 per hundred cards to be read.

Where data are accessible or have been recovered, conversion scripts can be run to create preservation and dissemination copies. There are some commercially available scripts to do this.

Processing

It is possible to use automated tools, such as the Nesstar Publisher, for cataloguing, indexing and for looking at disclosure issues. Using the metadata collected at the time of acquisition and internal metadata from the data some of the cataloguing can be automated to the extent that standard data, received regularly, can be processed in five minutes rather than the two days it would have taken to do manually. However processing time can vary enormously, the UKDA estimate a variation of between two and eight hours, depending on the complexity of the data and whether it is a familiar format.

Preservation

It is possible to automate the preservation process so that tasks are carried out at different levels on a daily, weekly and monthly basis. At the UKDA bespoke scripts written internally have been developed to carry out the preservation tasks. Establishing and maintaining the system are the most important cost issues. The recent Digital Preservation Coalition's *Mind the Gap* report identifies the need for more technical tools to help perform digital preservation activities and recommends that organisations link up to engage software vendors in developing these tools (p38). It is important that any tools created for migration and emulation produce preservation and dissemination formats that will satisfy regulatory authorities and be legally admissible.

Access

Access requires an on-line ordering system with delivery in a range of formats. It is probably also advisable to have some form of user registration with automated authentication. The main cost implications are in the purchase or development and maintenance of such systems. The UKDA has developed its own systems for online use of textual material, see for example:

<http://www.esds.ac.uk/qualidata/online/explore/introduction.asp>.

It is also possible to buy off the shelf solutions such as Beyond 20-20 which allow online viewing of numeric material. The UKDA makes use of Nesstar software for online data browsing and delivery. Some tools may be linked to electronic service delivery and County Councils may be buying into them as part of their records management and EDRMS.

User Support

This needs staff with the technical knowledge to provide support. An off the peg database system can be used to log queries.

4.2.2 Cost Models

In order to consider the implications of different cost models the processes of digital preservation were broken down into cost elements. These were:

1. **Analysis** of the records and any catalogues which go with them by the specialist repository, perhaps involving a third party sub-contractor;
2. **Conversion** to preservation and access formats as appropriate;
3. Transfer of the data to a **preservation** system in its original form. Depending on the type of record and the length of time it needs to be preserved, this may also involve transfer or creation of hardware and software independent preservation and dissemination copies - for example an ASCII tab-delimited version of a Microsoft Excel file in addition to the original Excel version;
4. **Retrieval** of records for local delivery;
5. Content management, retrieval and **dissemination**.

Cost elements one and two could form Cost Model One. Records could be transferred back to the client at either stage. This model may be most attractive to large organisations such as County Councils who could then incorporate the recovered data into their own in-house systems, which would then deal with future preservation and dissemination.

Cost elements three to five would be Cost Model Two and may follow on from Cost Model One, if necessary. It is anticipated that buying into this model would be more attractive to clients with less well-developed in-house systems. Records can be transferred back to the client at any stage by agreement. See

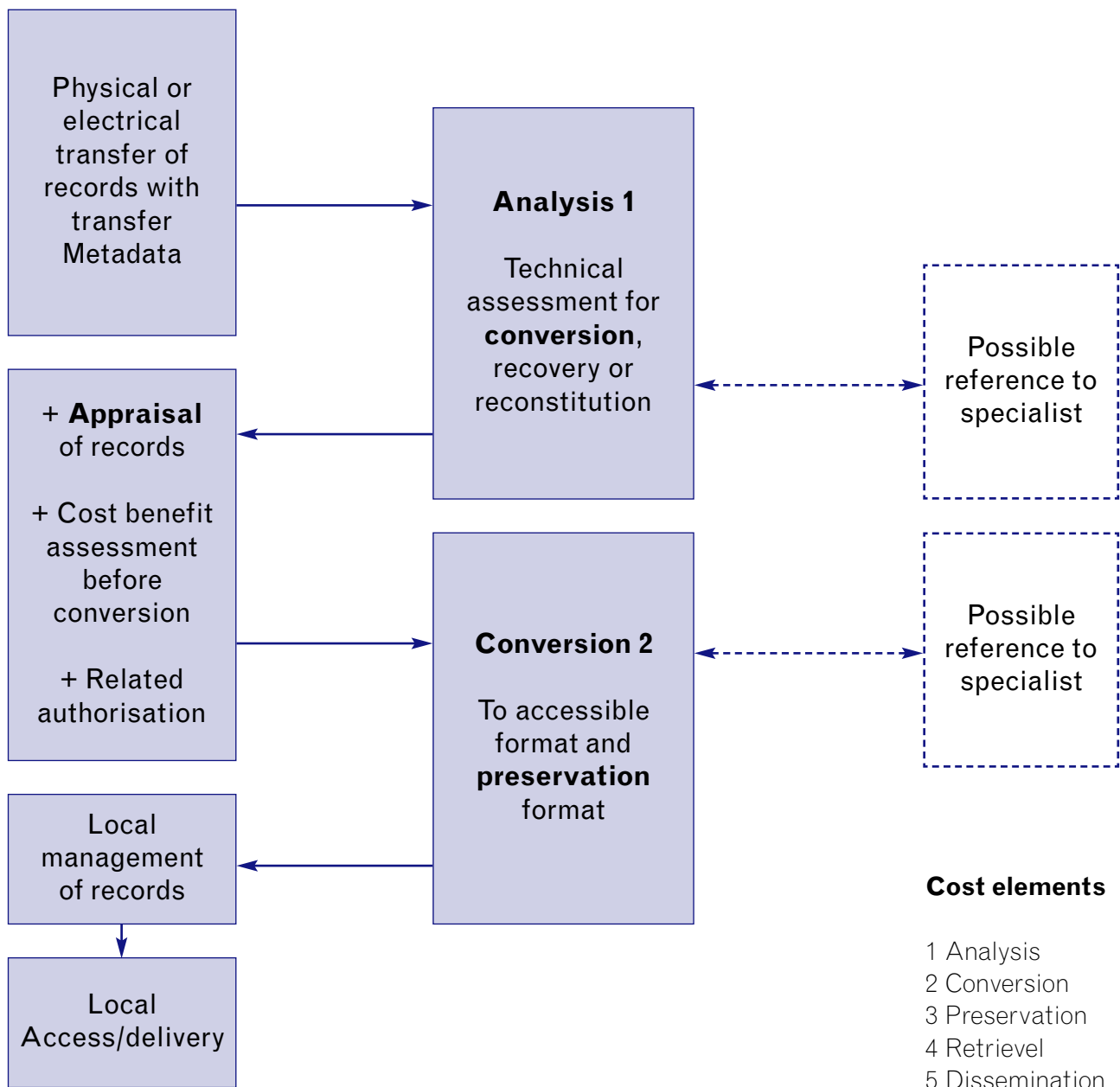
4.0

Figure 1. Other cost models are being developed and made available (*Report on the Digital Curation Centre and Digital Preservation Coalition Workshop*, p5) and these may be used in developing business cases and in the procurement process.

Cost Model 1

Contract for analysis and/or conversion

Transferring archive



Cost Model 2

Contract for long term preservation management and optional access delivery

Transferring archive

Physical or electrical
transfer of readable records
(possibly following analysis
and conversion via
cost model 1)
with transfer Metadata

Specialist repository

Remote 'hosted'
management of records

Preservation 3 and Storage Contract

Refresh periodic
remastering

Local
Access/delivery

Retrieval 4
In original form

Optional
Content management
retrieval and

Dissemination 5

4.0

Traditional archives require similar processes to those required by digital archives. Even where processes, and therefore costs, may appear different, as in the migration and refreshing needed for a series of e-mail letters, it should be remembered that there is a comparable equivalent for traditional paper based letters. These will need packaging replaced on a regular basis and may in time also need interventive conservation such as de-acidification or physical repair. The cost of maintaining a media store can be equated with the costs of maintaining an archive storage area to the correct environmental standard to ensure long term preservation of archives in traditional formats, which may require an air conditioning plant running 24 hours a day 7 days a week. The supporting technical infrastructure which is integral to digital record keeping achieves similar objectives to those undertaken by archivists working traditionally to produce finding aids, deal with enquiries etc. Similarly where, with digital archives, archivists may need to turn to the support of an IT specialist, with traditional archives they may currently need the technical knowledge of an archive conservator to explain the physical characteristics of the materials from which a letter is constructed. One additional unique cost involved in managing digital archives is the provision of back-up copies of the data and mirrors of the systems that provide the catalogue and dissemination services to ensure continuity of service provision.

4.2.3 Creating a business case

One of the objectives of the Digital Archive Regional Pilot Project was to develop a basis for a business case for a regional digital strategy. To this end the project considered the options of in-house, contracting out and consortia for the delivery of digital archive services. The likelihood of needing to contract out at least parts of the process and the potential advantages of economies of scale made a regional archive one of the possible viable options. Although the possible solutions and options for providing remote access make the geographical location of the repository less important, the presence of the UKDA within the region and the lack of similar repositories within the field make the development of the UKDA into a specialist regional repository for digital material a possible strategy.

County Councils and other organisations are required to follow set procurement procedures, including tendering, designed to ensure value for money and these will impact on the choices made by these bodies when deciding whether to provide digital preservation in-house, by consortia or by contracting out. The market is currently underdeveloped and options for contracting out are limited. However, if the market develops, as is very likely, and automated systems and collaboration bring competition in pricing, the business case for preferring one provider over another and the costs and feasibility of changing provider would also need to be considered by these organisations.

As responsible records managers, authorities will have to make effective arrangements for their own records to comply with Freedom of Information, related legislation, and corporate retention requirements. Building a business case for the investment required for this will be comparatively easy. However, building a business case for all digital holdings may not be straightforward and may pose a new and wide series of questions about collecting policy and intake that could even go to the heart of the *raison d'être* for local archive services, challenging assumptions for long tacitly accepted. *Mind the Gap* found that '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...' (p36). It recognised that there is currently little information available to help with this.

The management of external digital archives will necessarily have cost consequences for base revenue funding as it will involve new costs. The long-term preservation needs will require intervention, the non standard nature of the material (multiple formats with unstructured data), the programming required to convert material to preservation formats and the need to invest in ways of ensuring fundamental issues such as preservation and access would seem to make this inevitable.

Analysis of the experience of the UKDA in measuring the costs associated with the intake and management of digital records poses challenging issues for local authorities in dealing with deposits of digital archives from private sources. Collecting repositories such as HALS and BLARS face the question of whether the additional spend required for digital archives from external sources can be justified to the parent authority. Those authorities are likely to question how this spend will help contribute to their strategic objectives and the national shared priorities set by government, which are increasingly the most powerful drivers of policy decision making. This will probably influence the approach local archives services adopt to deal with private intake of digital archives and may pose interesting questions about the relative value of intake against the backdrop of national issues such as social inclusion and community engagement, which are likely to underpin bidding regimes that could also be critical to any new investment.

In some cases traditional archives services have employed contracting-out or consortia to reduce costs, particularly where specialist skills are required such as with conservation. Collecting archives have also often sought external short-term grant funding to pay for processing, cataloguing, and preservation of large deposits of privately or publicly owned material. However, in general, most processes involved in archives administration have largely been provided in-house by archivists and conservators employed by the local authority and whose prime responsibility is to take care of the parent organisation's own archives.

4.0

Given the specialist requirements of digital material, the probable scope and value of contracting out some or all of the processes for digital archives is higher than for traditional archives and this would make the costs involved (both revenue and capital) more visible to the funding authority.

This issue of costs may open up broader questions about the value of privately deposited archives to a council's strategic priorities and the willingness of authorities to subsidise new spend to preserve archives for demonstrable community benefit. It may also pose questions about whether dowries need to be associated with some intake as part of terms of deposit. Many of the tacit certainties around intake of paper records could come under question in the context of the requirements of dealing with digital intake. Questions about economies of scale around preservation and access may well be raised once more cost modelling is undertaken.

The challenge of putting together a business case for the acquisition, management and preservation of digital and traditional material from external sources will, as the *Mind the Gap* report suggests, rest on the ability of the sector to prove the value of such archives to the authority and to the community it serves. To this end the Community Archives Development Group, an affiliate of the National Council on Archives with its partners, TNA and the Museums, Libraries and Archives Council are seeking to commission a systematic and critical evaluation of the impact of community archives: 'to assemble a body of evidence that can be used for advocacy and to leverage funding, demonstrating the contribution of community archives to the agendas of Central Government, especially DCMS agendas around liveability, civility and inclusion; Local Government; Funders, particularly Heritage Lottery Fund and Big Lottery Fund; and Senior executives/parent organisations' (Invitation to tender announcement, and details of project http://www.ncaonline.org.uk/materials/cadg_impact_itt.pdf available 12/4/06)

Conclusions

The project revealed the relative inexperience of local authority services in dealing with intake of digital records when compared with the UKDA to which this work is central. It showed the extent to which the management of digital records has been developed with limited consideration of the terms and processes used in the archives administration and records management of hard copy media.

This was evident from the difficulty in applying the terms used in the OAIS model which was used for reference but not considered adoptable. It was also clear from the time spent in sharing and explaining terminology, unravelling confusion around terms, and drawing comparisons about the various processes and relative costs involved in all aspects of managing digital records. Mutual understanding was achieved but took time and showed the relative immaturity of much of the terminology applied to digital records.

There was significant value in going through a stage by stage comparison of the processes and costs associated with the preservation of hard copy and digital records and the management of access to them. This revealed (see section 3) both similarities and differences. There were two particularly significant issues which were different with digital records. Firstly, the manager needs to be much more active in dealing with preservation. Secondly, there is a clear need to be able to assist the user not only in understanding a record but also the technicalities, including perhaps registration of access, in being able to view it in the first place.

Preservation was agreed to be a critical issue with an important distinction between issues associated with intake and those linked to long term management and maintenance. Distinctions were drawn between digital preservation of corporately generated records (the 'corporate memory' of the Councils involved in the project) and those that might be offered privately by the range of donors and depositors from which a local authority archives service takes records into custody.

In both cases the project team recognised the importance of looking at the technical history of systems used by their organisations and, likewise, the hardware and software that had been used to create private digital records. Capturing key metadata was recognised as a critical element in being able to manage all digital records. There was more expectation of conscious management with the corporate records and some anxiety about the range of options that might emerge from private offerings. With private records the need to gain correct information and permissions at the time of deposit was recognised and it is hoped that the sample depositor form which was created to deal with privately deposited digital intake is a useful development.

5.0

It was felt that this project was perhaps the first effort to quantify the costs associated with the various processes involved in managing digital records by a non-specialist repository. It revealed both the sizable capital sums required for some processes and the revenue requirements. There remains significant uncertainty about the latter because of the limited knowledge about what might be offered from private sources.

The project explored the challenge of trying to turn the information about costs, much of which was derived from the UKDA's operational experience, into a business case but there was a limit to how far this could be taken at this stage. From the perspective of the local authority services involved, the importance of using a specialist repository – which might be linked to a consortium – was shown for some of the processes required both in preserving and making accessible digital records. To give a simple example, the work on 'obsolete' data revealed both the cost and expertise need in rescuing, recovering and transferring which required some use of specialist sub-contractors by the UKDA.

Costing out a business case on a consortium basis, whether regional or wider, would require a more precise assessment of the range and extent of processes that the array of different repositories might want to procure. This was not possible with any certainty even for the local authority archives services involved in relation both to their corporate archives and the private records which they administer and manage.

It was agreed that digital preservation needs to have a corporate profile in organisations and to involve ICT staff working in partnership with archive and perhaps contracted third party services. For the network of local authority archives services it was felt that the private intake of digital records should be considered as part of performing an important service to local communities.

Recommendations

These fall into two categories: recognising and building on the lessons and learning from the project and, secondly, recognising the need for further research and analysis in some areas.

1. Lessons and learning

- need for corporate recognition of the importance and urgency of preserving digital records and the costs associated with managing them;
- need to extend this recognition to private intake as part of the community role performed by local authority services;
- need to consider further modelling of costs and benefits linked with the three scenarios of: in-house provision, working through consortia and contracting out;
- importance of considering the technical histories of systems and the importance and quality of metadata, especially with regard to private intake;
- need to develop and clarify the OAIS model to make it more intelligible and better aligned, with more conventional terminology applied to archive administration and records management.

2. Areas for further research and analysis

- accuracy around predictions on intake – the quantity, quality and format of potential deposits; and clarity about the likely nature and extent of obsolescence in records offered as private deposits;
- revenue cost consequences of intake and management of digital records – clarity about the many variables and assumptions applied because of limited experience of digital records outside the UKDA, including possible differences in dealing with structured and unstructured data;
- variations in the contracting out of the range of processes for managing digital records across the different repositories.

Further Reading

Report on the Digital Curation Centre and Digital Preservation Coalition Workshop on cost Models for Preserving Digital Assets, July 26th 2005, British Library, London. Available at <<http://www.dcc.ac.uk/training/cm-2005>> 12 September 2005>

Digital Preservation Coalition. *Mind the Gap*. Available at: <<http://www.dpconline.org/docs/reports/uknamindthegap.pdf>> 15 February 2006>

Jones, Maggie. and Beagrie, Neil. *Preservation Management of Digital Materials: A Handbook*. The British Library. Available at <<http://www.dpconline.org/graphics/handbook/>> 24 August 2005 >

Lavoie, Brian F. *The Open Archival Information System Reference Model: Introductory Guide*. DCP Technology Watch Report 04-01 2004. Available at <http://dpconline.org/docs/lavoie_OAIS.pdf> 29 March 2006>.

National Council on Archives. *Your Data at Risk: Why you should be worried about preserving electronic records*. NCA, 2005

Office of the Deputy Prime Minister *Guidance on 'Proper Arrangements' for Archives*. Available at <<http://www.odpm.gov.uk>> 25 April 2006

Pickford, C. *Eastern Promise: a Strategy for Archival Development in the East of England*. EERAC, 2003. Available at <<http://www.ncaonline.org.uk/materials/easternpromise.pdf>>; 24 August 2005

Tritton, Deborah. *The Archives of the Future: Planning for Electronic Records - a Survey of South West Local Authority Archive Services*, (forthcoming) 2006

Appendix A: Results of Readability of Obsolete Formats

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC1	Full Intranet 5/7/96 First Intranet of HCC. March 1997 T Kent Member data – Confidential	Year of creation indicated it would most probably be Windows based OS, created with a standard Windows backup package	DAT Tape Maxwell HS-4/120s. Supplied in protective plastic case in good condition	Analysis confirmed that media was written under Windows OS using Veritas backup software. All data restored using Veritas 9.1 on Windows 2003 using a SONY DDS3 DAT drive. All possible files written to CD – Contents determined to be a full backup of HCC's Oracle Intranet database with a web-based front end from July 1996. Seems intact and complete. Further analysis possible but would require reconstruction of Oracle database.	Restoration of data and migration to present-day media successful. Work on checking materials and preparing them for preservation system completed.
HCC2		Year of creation indicated that it would most probably be Windows based OS, created with a standard Windows backup package.	DAT Tape Maxwell HS-4/90s Supplied in protective plastic case	Analysis reported that the media had been written under Netware OS using Veritas backup software All data restored using Veritas 9.1 on Windows 2003 using a SONY DDS3 DAT drive. All possible files written to CD. Media contained daily backup of HCC intranet on a Netware server. Tape contains two backups.	Restoration of data and migration to present-day media successful. Work on checking materials and preparing them for preservation system completed.
HCC3		Type of media indicated that it would be Windows based OS, backed up using the standard backup software that would have come with the tape drive.	Imation 3m Travan TR3 Supplied in protective plastic case.	Analysis using media's native drive, IBM Travan3, confirmed that it was written under Windows using the supplied backup software. All data restored using Iomega Ditto on Windows 2003 using an IBM Travan TR3 parallel port drive. The media was also successfully restored using a SCSI based TR5 drive.	Restoration of data and migration to present-day media successful. Work on checking materials and preparing them for preservation system completed.

48 **Appendix A: Results of Readability of Obsolete Formats**

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC4	Tuesday 89/90	Media type and age indicated that it would be either Unix or early Windows based OS.	3M DC 600A Data cartridge Supplied in protective plastic case	<p>Analysis under UNIX indicated that either media was faulty or could not be read in available UKDA tape drive.</p> <p>Due to age of media and possible issues with tracking, a second drive was sourced to try and confirm if media faulty. Still couldn't be read in second tape drive.</p> <p>Additional media required to test further using our existing drive.</p> <p>As media was deemed as most probably being faulty, it was decided not to send the tape off to a specialist media conversion company.</p>	Restoration of data not successful.
HCC5	'Sound Stuff'	WAV file	<p>Disk Cartridge MD?HD 12 MB</p> <p>Cartridge looked undamaged</p>	<p>Media investigated confirmed that UKDA do not have anything in-house that could possibly read it.</p> <p>Requested additional media of the same type from HCC.</p> <p>Media sent off to specialist media conversion company for further analysis.</p>	<p>Specialist media conversion company has been unable to read the media.</p> <p>Restoration of data not successful.</p>

Appendix A: Results of Readability of Obsolete Formats

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC6	Capital Monitoring System Reset		Punch cards x 9	<p>Visual analysis of the media indicated that they were encoded in standard Hollerith format with the value printed on top of each card.</p> <p>Read successfully using punched card reader. As cards had the values at the top the media could just as easily be transcribed from the cards directly, if a suitable device was unavailable of the cards were in a poor condition.</p> <p>All data restored. However, on their own the information on the cards is quite meaningless since the cards are job control cards relating to a dataset that would have been stored elsewhere within the system, maybe on additional punch cards. Thus for completeness the additional data media would also need to be restored</p>	Media analysis complete, but of little to no value.
HCC7	East Ga Lotus 2.0 2 K	Lotus V2 would indicate time period of 1985 – 87 and Operating System of CP/M or MS-DOS.	5_ inch floppy. Dysan Double density	<p>Media investigated determined that the disc had a standard MS-DOS FAT table.</p> <p>Media read successfully on a Windows 2003 server using a 5 inch floppy drive.</p> <p>File EASTGA.WK1 read and assessed: contains Education Committee budget details for East Herts College for 1986-1989. Lotus 123 file opens in Excel, copy made in Excel and archival version in tab-delimited ASCII created.</p>	<p>Restoration of data and migration to present-day media successful.</p> <p>Work on checking materials and preparing them for preservation system completed.</p>

50 **Appendix A: Results of Readability of Obsolete Formats**

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC8	J Alleyne Works order invoiced DBASE II.	Dbase II would indicate 1981-2 for age of media.	5 inch floppy BASF flexydisc.	<p>Media analysis indicated that it was not a standard DOS FAT table, so would most likely be one of the variants of CP/M.</p> <p>Analysed disc to try and determine what type of CP/M format, but without any success.</p> <p>Help was requested from HCC to try and determine what the originating machine may have been. Informed that it was created using a Xerox CP/M machine.</p> <p>Attempting to read again using Xerox CP/M mode, but still gave problems indicating that media is faulty.</p> <p>Media sent to specialist media conversion company for further analysis.</p>	<p>Specialist media conversion company confirmed that the media were faulty and could not be read.</p> <p>Restoration of data not successful.</p>

Appendix A: Results of Readability of Obsolete Formats

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC9	EW9969 1600 EBDIC Back up data Environment planning	Type of media and character encoding indicated IBM mainframe as the server.	EPOCH 480 9 track tape reel	<p>Attempted to analyse under UNIX but media format unknown.</p> <p>Help requested from HCC in order to try and determine exact host machine on which the tape was written. Necessary details no longer available.</p> <p>Attempted to read media using format analyser tool, but unknown format returned.</p> <p>Media read in raw data stream format and written to CD-ROM for possible later analysis and restoration. 25MB of data were converted.</p> <p>Visual analysis after conversion to EBCDIC character encoding reveals details that it was probably created on 3/3/88 on from a DEC mainframe machine.</p> <p>The tape had a few soft read errors which were re-read successfully.</p> <p>Further analysis could be done on the file in order to try and interpret the contents and structure of the data. Although possible, this would be very time-consuming as it would require scanning the data line by line and looking for file header signatures of the time. Even if this were done, it would not guarantee that the data structure could be interpreted satisfactorily.</p>	<p>Media analysis partially complete.</p> <p>Resulting data of little value without significant extra effort in trying to unravel the data structure.</p> <p>No guarantee of success.</p>

Appendix A: Results of Readability of Obsolete Formats

Reference	Title/presumed contents	System	Media Type and condition	Progress	Outcome
HCC10	'Info Disc 2, Rebecca Phillips Images of bricks'		Phillips Laser disc Floppy disc	<p>Media could not initially be read using our standard 12" laser disk.</p> <p>Media sent off to TNA and played successfully.</p> <p>The material on the disk contains hundreds of pictures and a video titled "Design For Living With Brick".</p> <p>However, the player used does not have the ability to offload the contents so restoration of contents would require sending the disc off to specialist media conversion company.</p>	Media analysis partially complete
HCC11	D/P 4 30/8/91		Maxwell MF 2 - DD Floppy disc	<p>Help requested from HCC in order to try and determine exact host machine but no further information available.</p> <p>Various formats attempted include Macintosh and OS/2 but without success.</p> <p>Media sent to media conversion specialist for further analysis.</p>	<p>Specialist media conversion company confirmed that the media were faulty and could not be read.</p> <p>Restoration of data not successful.</p>
HCC12	BGAPT+C.RFT SWAPT+C.RFT Disc2		£M DS DD 1.0 MB	<p>Help requested from HCC in order to determine exact host machine but no further information available – maybe similar format to HCC11.</p> <p>Various formats attempted include Macintosh and OS/2 but without success.</p> <p>Media sent to media conversion specialist for further analysis.</p>	<p>Specialist media conversion company confirmed that the media were faulty and could not be read.</p> <p>Restoration of data not successful.</p>

Appendix B: Sample Deposit Form

BEDFORDSHIRE AND LUTON

Archives and
Records
Service

Riverside Building
County Hall
Cauldwell Street
Bedford
MK42 9AP
Tel: 01234 228833/228777
e-mail: archive@bedscc.gov.uk
Website: <http://www.bedfordshire.gov.uk/archive>
Kevin Ward, County Archivist

Interim receipt for deposited records

Name of depositor (BLOCK CAPITALS)

(i.e. the person or organisation in whose name the records are to be held)

Name of person making delivery (if not depositor)

Position (where applicable)

Address for receipt

Telephone No:

e-mail address

Category of material (if institutional)

e.g. Parish, Parish Council, Hospital, Charity

Brief description of documents (including quantity)

Covering

FOR OFFICE USE ONLY

Received the above listed documents on behalf
of the Archives & Records service

Signature: _____

Position: _____

Date: _____

Accession Number: _____

Processing reference: _____

Digital material included? ☐ Yes ☐ No
(if yes complete digital deposit form)

Terms of deposit ☐ Gift
☐ Indefinite Loan
☐ Purchase

Authority to weed? ☐ Yes ☐ No

Disposal of weeded material _____

Power to transfer to
other repositories? ☐ Yes ☐ No ☐ N/A

Special conditions (e.g. embargoes) _____

I have seen and agreed the appropriate terms of deposit and agree
this is a correct record of my deposit

Signature _____

Initial this space if no further acknowledgement is required

Appendix B

BEDFORDSHIRE AND LUTON

Archives and
Records
Service

Riverside Building
County Hall
Cauldwell Street
Bedford
MK42 9AP
Tel: 01234 228833/228777
e-mail: archive@bedscc.gov.uk
Website: <http://www.bedfordshire.gov.uk/archive>
Kevin Ward, County Archivist

ACCESSION No.

REFERENCE No.

DIGITAL MATERIAL DEPOSIT FORM

This form records the information necessary for the cataloguing and preservation of digital material. Please complete it as fully as possible. If you are unable to answer a question please leave it blank.

CONTACT DETAILS (BLOCK CAPITALS)

In case we have difficulty reading or understanding the material supplied please give the name, address and telephone number of someone who can be contacted for further details.

1. PROVENANCE

1.1 Creator

(i.e. person or organisation responsible for creation. Please be specific e.g. to department level)

1.2 Other contributors

1.3 Copyright holder/s (if known)

2. QUANTITY & TYPE

1. 2. Medium for transfer to the archive

(Please state number and type) Disk (e.g. 1.4Mb 3" floppy disk in DOS or Apple Mac format, Amstrad 'Amsoft' discs etc)

- | | | | |
|---|--|--|---|
| <input type="checkbox"/> Cartridges: | <input type="checkbox"/> Dat (4mm) | <input type="checkbox"/> Exabyte (8mm) | <input type="checkbox"/> 6150 "QIC" |
| <input type="checkbox"/> CD Roms | <input type="checkbox"/> CD-R | <input type="checkbox"/> CD-RW | |
| <input type="checkbox"/> DVD | <input type="checkbox"/> DVD-R | <input type="checkbox"/> DVD-RW | <input type="checkbox"/> DVD+R |
| <input type="checkbox"/> DVD+RAM | <input type="checkbox"/> Magneto-optical | <input type="checkbox"/> Tape-streamer | <input type="checkbox"/> Zip drive/disk |
| <input type="checkbox"/> Jaz drive/disk | <input type="checkbox"/> External hard drive | <input type="checkbox"/> Microdrive | <input type="checkbox"/> Memory stick |

Other (please describe)

3. SCOPE & CONTENT

3.1 Title

(When the file bears a formal title it should be transcribed exactly as to wording, order and spelling. If appropriate, abridge a long formal title, but only if this can be done without loss of essential information.)

3.2 Description of content

(This should be used to identify the chief distinguishing feature/s of the file, including subject matter, location or theme etc. Where possible indicate the source of the data, its completeness and geographical coverage, and data collection and processing methods. or contact details for any one who may be able to provide this information).

3.4 Covering dates:

3.5 Creation date/s:

Appendix B

DIGITAL MATERIAL DEPOSIT FORM (continued)

3.6 Kind of data

(Please specify the nature of the data materials)

- ☐ Text (e.g. MS Word, Wordperfect, RTF, XML, HTML, SGML etc.) ☐ Database (e.g. MS Access, dBase, XML, Filemaker Pro etc.)
- ☐ Spreadsheet (e.g. MS Excel, Lotus, Quattro Pro, delimited text) ☐ Image/s (e.g. TIFF, JPG, GIF, BMP, PDF)
- ☐ Video (e.g. MPEG-1, MPEG-2, MPEG-4, Apple Quicktime) ☐ Audio (e.g. MS Wave, MPEG-1, Audio Layer 3 (MP3))
- ☐ Statistics (e.g. delimited text, SPSS.por, SPSS.sav, STATA, SAS)
- ☐ Web pages

Other: _____

3.7 Nature of material (Please specify details where appropriate)

- ☐ One-off ☐ Repeated ☐ Follow-up ☐ Continuous
- ☐ Aggregate statistics ☐ Other _____

3.8 Related material (Please specify any related material e.g. where data continues previous series in hard copy)

4. FORMAT (Please be as specific as possible)

4.1 Created on: (Please specify details where appropriate)

Main frame	Personal computer	
Model	Platform	Operating System

4.2 Software (give version where possible e.g. Microsoft Word 6.0)

4.3 Coding formats (Please indicate whether the data are compressed. If compressed please indicate the format and if possible include a copy of the decompression programme.)

- ☐ Uncompressed ☐ UNIX TAR file ☐ UNIX compressed file ☐ pkZIP file 00
- ☐ Gzip file ☐ Other _____

Software (give version where possible e.g. Microsoft Word 6.0)

5. RIGHTS MANAGEMENT

5.1 Data Protection Act

Does the material contain information which identifies living individuals? YES/NO

5.2 Freedom of Information Act

Is the creating body subject to the Freedom of Information Act? YES/NO

5.3 Environmental Information Regulations?

Does the material contain information which is subject to the EIR? YES/NO

5.4 Other intellectual property rights (Please specify anyone who may have rights in the material e.g. funding bodies)

6. Return of material

If you wish to have your disks or tapes returned please tick here ☐

Important! Please note:

Records received by the Service in digital form may be transferred to a specialist repository for secure hosting and/or any migration or copying considered necessary for their long-term preservation.

FOR OFFICE USE ONLY

Requirements for access and delivery

Glossary

*Term used in OAIS Reference model

Term	General
Acquisition/accession/authoring/ingest*	Set of processes by which archives are submitted to and prepared for inclusion in an archival store*
Access*	Mechanisms and services to support users' needs and requirements for preserved material*
Administration*	Management of day to day operations of a digital repository: for OAIS purposes, the term relates to an OAIS, i.e. an Open Archival Information System*
Appraisal/Collections/Selection policy	Organisational policy under which material is accepted into a repository
An Archive	Record or collection of records which is deemed to be worthy of permanent preservation
Archive Service	Department or organisation providing and/or administering the processes required to acquire, catalogue, preserve and make available archives
Authentication	Verification of the identity of one requesting access to records. This is usually achieved via a challenge response scenario i.e. username/password
1. Catalogue	1. Written description of the contents and format of an archive collection
2. To catalogue	2. The act of creating a catalogue
Checksum	Analogous with fixity check: an algorithm based method for determining digital data integrity by determining whether errors or alterations have occurred during the transmission or storage of digital data
Collection	A group of archives created or received from a single source. Equivalent to the French archive word <i>fonds</i> .
1. Deposit	1. Archive placed in a repository on long-term loan, as a gift or under licence: in county record offices deposits are not gifts they are long-term loans as ownership of the material is retained by the depositor; similarly with the UKDA where, deposits are lodged under licence

Glossary

*Term used in OAIS Reference model

2. To deposit	2. To place an archive in a repository
Disclosure control	Methods by which individual information held in electronic files is protected from misuse
Dissemination (format)	File prepared for immediate, non-specialist use
Emulation	The mimicking of one system by another; this may relate to both hardware and software
Encryption	Encryption is the process of converting data into a form that cannot be easily read by unauthorised users. The converse is decryption.
Finding aids	Means by which records and accompanying information can be discovered e.g. by indexes, keyword tagging etc
Fixity checks	Analogous with checksum: an algorithm based method for determining digital data integrity by determining whether errors or alterations have occurred during the transmission or storage of digital data.
Format	Either the structure or composition of a data file or the physical form of information as opposed to the content.
Front end	Computer screens designed for users
Interoperability	The ability of two or more systems or components to exchange information and to productively use the exchanged information
Licence & copyright agreements	Contract under which the author of an archive passes preservation and usage rights to the repository
Medium/media	Physical object on which digitised files are stored (e.g. tapes or discs)
Metadata	Structured information to describe data
Metadata standard	Formal systems for structuring metadata (e.g. e-GMS, DDI)

Glossary

*Term used in OAIS Reference model

Migration	Transfer of data from one medium to another, or of software from one hardware platform to another, or of servers from one operating system to another
Online browsing	Viewing web pages via the Internet
Online retrieval	Recovering records via the Internet
Preservation (format)	File prepared for long-term preservation
Preservation/technology* watch	Constant review of developments that affect the long-term maintenance and functioning of a digital archive
Processing	Preparation of files for preservation and dissemination
Producer*/Depositor/Author	Individuals, organisations or systems that transfer information for long-term preservation*
Provenance	Information about the origins of a collection
Reception	Process under which an archive is transferred for long-term preservation
Records	Information identified upon its creation as having value as evidence in 'recording' an activity or decision and so requiring a 'lifespan' for its management either through to destruction or appraisal and retention as part of an archive
Refreshment	Systematic copying of files between used and unused media
Repository	Site at which digital records are stored: can be a data warehouse for records or a digital archive for archives
Retrieval	Recovery of files for use
Risk analysis	Assessment of effects of adverse events on long-term preservation
Secure (transfer/environment)	Protected from interference by non-authorised parties
Transfer	Movement of files between organisations

Acknowledgements

The project team `comprised:

Pamela Birch	(Bedfordshire County Council)
Kevin Ward	(Bedfordshire County Council)
Christine Shearman	(Hertfordshire County Council)*
Gordon Chancellor	(MLA East of England)
Hilary Beedham	(UK Data Archive)
Kevin Schürer	(UK Data Archive)

The project team would like to acknowledge the individual contributions to this project made by:

Carol Fuller	(Bedfordshire County Council)
Rebecca Phillip	(Bedfordshire County Council)
Tony May	(Hertfordshire County Council)
Tim Heathcote	(MLA East of England)
Terry Turner	(MLA East of England)
Richard Blake	(The National Archives)
John Alban	(Norfolk County Council and EERAC)
Mus Ahmet	(UK Data Archive)
Kim Berry	(UK Data Archive)
Sharon Bolton	(UK Data Archive)
Susan Cadogan	(UK Data Archive)
Mike King	(UK Data Archive)
Ken Miller	(UK Data Archive)
Pam Miller	(UK Data Archive)

*Now at Cambridgeshire County Council



ISBN 0-9553410-0-0
ISBN 978-0-9553410-0-7



© MLA East of England and East of England Regional Archive Council, 2006

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the Copyright owners.

Published by MLA East of England and East of England Regional Archive Council

Produced by Shorthose Russell Ltd., Norwich www.shorthose-russell.co.uk