UK Data Archive

UK Data Archive
Preservation Policy

September 2005

UK Data Archive
University of Essex
Wivenhoe Park
Colchester, Essex CO4 3SQ
UK

Email: archive@essex.ac.uk
Tel: +44 (0) 1206 872001
Fax: +44 (0) 1206 872003

UK Data Archive Preservation Policy

Table of Contents

I. INTRODUCTION

1   EXECUTIVE SUMMARY

It is the mission of the UK Data Archive (UKDA) to support high quality research, teaching and learning in the social sciences and humanities by acquiring, developing and managing data and related digital resources, and by promoting and disseminating these resources as widely and effectively as possible. The responsibility to build and manage research collections carries with it the obligation to ensure that these collections are permanently accessible. The UK Data Archive, also here referred to as the Archive, is committed to the long-term preservation of its collections.

Since its inception in 1967, preservation has been a fundamental function of the UKDA. Preservation comprises a set of interrelated activities carried out by several sections of the Archive. The Information Systems & Preservation section has responsibility for preserving data and documentation to ensure they remain usable over time, including monitoring technological changes that will affect preservation and migration decisions. The Data & Support Services section is responsible for identifying and acquiring data and other resources according to the collections policy, validating the data and documentation and producing and enriching metadata about the data resources in order to populate finding aids. This section also provides user access to data and documentation, converts the preservation copy of data to distribution formats to meet user needs, and provides other value-added user support. The cross-sectional Information Development team interprets national and international standards for local implementation in its resource discovery metadata.

The purpose of this preservation policy is to provide a comprehensive statement about the preservation of the UKDA's data collections, dealing with all aspects of preservation and applying to all materials held by the Archive. The policy covers preserving data collections for which the Archive is a custodian and does not consider preservation of other materials such as the Archive web pages, internal administrative documents and correspondence, and the Archive's intranet. These materials are governed by the Archive's records management programme.

This policy provides the strategic direction required to initiate any measures that are necessary for the protection of the UKDA collections. It provides direction to Archive staff in carrying out their collection management responsibilities relating to preservation. It forms the basis of communication with colleagues outside the Archive concerning both internal preservation strategies, and national and international co-operation.

It also identifies related UKDA policies, and those documents that contain the detailed interpretations and procedures derived from these. Implementation of this policy is achieved through the normal channels of the Archive planning process, in particular through the operation of the Archive's strategic and operational plans. This policy defines organisational roles and responsibilities associated with the preservation of the UKDA collections.

The UKDA continues to organise and promote preservation as an integral part of the management of the Archive and to ensure both the rational use of resources and that the staff receive training and support for preservation activities.

Queries concerning the preservation of the digital collections in the UKDA, or digital preservation in general, should be directed to the UKDA Systems and Preservation group (preservation@essex.ac.uk).

Note

References of document titles or other information content given in italics are listed in section 21 (References), where URLS are given where the document and/or information content is available on the web.

2   MISSION OF THE UK DATA ARCHIVE

It is the mission of the UKDA to support high quality research, teaching and learning in the social sciences and humanities by acquiring, developing and managing data and related digital resources, and by promoting and disseminating these resources as widely and effectively as possible.

In fulfilling this mission, the Archive strives to ensure that the:

- materials it acquires and accessions are suitable for scholarly use;
- data are accompanied by adequate documentation to enable their use for secondary analysis;
- data are checked and validated according to strict data processing procedures;
- data holdings are professionally catalogued according to appropriate metadata standards;
- data holdings are indexed with keyword terms using an appropriate thesaurus created by the UKDA;
- data are effectively preserved for future use by converting them to several standardised formats and retaining multiple copies on different storage media;
- format of materials is changed as necessary to preserve access to their intellectual content, reducing the risk of losing access to them over time;
- materials are kept in conditions suitable for long-term archival storage.

This preservation policy is a codification of long-standing good archival practice at the UKDA. The Archive has been the primary repository for social science research data in the UK since 1967. As a 'national data collection service' the UKDA, originally called Data Bank, was created by the forward-thinking Social Science Research Council, now the Economic and Social Research Council (ESRC), to bring together 'social survey research materials for storage, retrieval and secondary analysis of the information in them'. For over three decades, preservation of these collections has been a core function of this enterprise. The ESRC *Datasets Policy* emphasises the importance and requirement of depositing ESRC data with the Archive.

The UKDA continues to facilitate secondary analysis in the scholarly community by preserving and sharing research material that may have served its initial purpose but has a continued value for re-use. By taking a strategic approach to long-term digital preservation, the Archive ensures that it is at the leading edge of technical advances, by monitoring hardware and software developments and migrating its collections accordingly. The Archive is committed to using its resources wisely, and adding value to data collections where it will most benefit the user community.

The formulation and biennial revision of a preservation policy for the UKDA is an essential step in fulfilling its strategic aims and responsibilities: it gives strategic direction both to initiate any measures which are necessary for the protection of its collections, and to meet, or extend, nationally and internationally agreed standards for the preservation of digital materials. A preservation policy helps the Archive meet accountability requirements and its user community expectations.

3 RATIONALE

The UKDA recognises that use of its collections is a prime motive for its existence. The main aim of the Archive's preservation strategy is to make information accessible, while ensuring its survival and usability in perpetuity.

The Systems and Preservation group can be seen as the central hub of the Archive: it safely preserves deposited data and makes them securely available for present and future user needs. This policy document is an outline of how the Systems and Preservation group fulfils these objectives for some of the largest collections of digital materials in the UK: the Economic and Social Data Service (ESDS), AHDS History, and in an advisory and consultancy capacity offers expertise in digital preservation to other bodies including the Medical Research Council (MRC). Details of the organisations and services for which the Archive provides preservation services are given in Section 15. This document also describes how the UKDA organises and promotes preservation as an integral part of the Archive's collections management, and how it ensures rational use of resources.

Digital preservation as covered by this document is concerned with the preservation of information on optical and magnetic media. It can be defined as the actions needed to ensure enduring access to the full content of digital resources over time.

The UKDA's comprehensive approach to digital preservation entails choosing the appropriate methods and formats for preservation of every resource. This is accomplished through: selection of the most suitable formats for acquisition and archiving, storage in optimum conditions, secure handling and housing that uses security systems designed to eliminate risks of corruption and theft, and data refreshment and migration.

A preservation policy aims to minimise the risks associated with technological changes and to ensure that the digital resources remain accessible. By adopting such a policy the UKDA has committed itself to preserving all formats of its digital resources for which it is custodian, as well as the descriptive, structural, and administrative metadata associated with them.

The preservation policy provides direction to Archive staff in carrying out their collection management responsibilities relating to digital preservation. It identifies related Archive policies and those documents that contain the detailed interpretations and procedures derived from them. It also promotes involvement in the wider arena of national and international best practice. Implementation of this policy is achieved through the regular channels of the Archive planning process, in particular through the operation of the UKDA's strategic plan.

4   SCOPE AND STRUCTURE OF THE PRESERVATION POLICY

The primary objective of the UKDA preservation policy is to organise and promote preservation as an integral part of the management of its collections and to ensure rational use of its resources by providing a framework for managing the preservation procedures. The specific aims of the preservation policy are to:

- instil good practice in preservation management;
- manage the Archive's collections so that they can be used as an effective information source;
- improve the speed and efficiency with which information is preserved and retrieved;
- develop and maintain systems of low-cost storage, with appropriate location and with regular review;
- ensure that digital resources are managed throughout their life cycle in the medium most appropriate for the task they perform;
- maintain the integrity and quality of the data collections;
- ensure that all data collections are protected and kept secure in a manner commensurate with their value and use;
- optimise the use of the Archive's space for storage purposes.

This policy outlines a preservation philosophy for the Archive. It provides a comprehensive statement on the preservation strategies of the Archive's data collections, dealing with all aspects of preservation and applying to all collections held by the Archive. The policy also defines organisational roles and responsibilities associated with the preservation of the Archive's collections.

The policy document covers the following topics:

- general introduction and context of preservation activities;
- information technology (IT) infrastructure for digital preservation;
- metadata required for effective preservation;
- preservation strategies used;
- storage facilities for preservation;
- security measures for physical safety of the collections and system security;
- disaster recovery procedures for files;
- wider context of digital preservation and co-operation with other professional groups;
- monitoring and periodic review of the preservation policy;
- responsibilities for implementing both the preservation policy and training.

This preservation policy has mainly developed from the Archive's recognition of its specific preservation needs and from a drive towards wider transparency of the Archive's working procedures. The policy has been accepted by all staff of the UKDA for implementation throughout the Archive.

5   RELATIONSHIP WITH OTHER UK DATA ARCHIVE POLICIES

As a core activity in an archive, preservation does not exist in isolation. It needs to take account of:

- the aims and objectives of the archive;
- its strategic and operational plans;
- its collections development policy;
- the needs of the users of the archive;
- the significance of the individual resources;
- the place of the archive within local, national and international frameworks.

Preservation decisions have an impact on most areas of the UKDA and this preservation policy should be read in conjunction with the following related policy documents:

- *The UK Data Archive Strategic Plan*;
- *Assessment of UKDA and TNA Compliance with OASIS and METS Standards*
- *The UK Data Archive Collections Development Policy*;
- *The UK Data Archive Information Security Management Policy;*
- *Guidance on Data Management.*

The preservation policy and strategy are equally steered by a variety of external guidelines, manuals, and standards that represent an international body of knowledge and expertise pertaining to various issues within digital preservation. These are referenced in Section 21 of this policy document.

6   DEFINITIONS OF TERMS

For the purposes of this policy, some terms are used with specific meanings as defined below. Explanatory terms in bold are themselves defined in this section.

Code

This refers to uncompiled code in the command language of a **statistical package** or computer language, which occasionally forms part of a UKDA **data collection**. It may be important for code to be maintained as executable, in which case it is preserved in its own right and migrated as required, or it may be considered as more of a human readable resource and incorporated as part of the **documentation** (and cut and pasted back into an executable environment by the data user if desired). Code that shows how derived variables were created in a dataset might fall into this latter category.

Data

Data are all the material, regardless of format, which are intended to be analysed. They are the primary element of a data collection. More precise definitions of data vary according to context. Quantitative data may refer to just the matrices of numbers or words that comprise a data file, but may also cover other information (metadata) held within a **statistical package** data file, such as variable labels, code labels and missing value definitions. Qualitative data might include interview transcripts as well as audio and video recordings (analogue or digital).

Data collection

The UKDA organises **data** for preservation as data collections. A data collection is typically comprised of three components: **data**, **documentation** and **metadata**. Occasionally, a fourth component of **code** exists. Data collections are typically organised by reference to a particular survey or research topic and cover a specific geographic area and time period. There is normally a one-to-one relationship between a data collection and what the UKDA terms a **study**.

Data Documentation Initiative (DDI)

The DDI is a project of the social science community to establish an international criterion and methodology for describing the content, presentation, transport, and preservation of **metadata** about datasets in the social and behavioural sciences. The DDI is currently expressed as an XML Document Type Definition (DTD) and Schema. The UKDA has been centrally involved in the adoption, development and promotion of the DDI as an international standard for social science metadata.

Digital preservation

Digital preservation is a series of managed activities necessary to ensure continued access to digital materials for as long as necessary.

Documentation

Documentation is that portion of a **data collection** that is required in order to re-use **data**. It commonly covers the subjects of sampling design, methods of data collection, questionnaire/interview design, structure of the data files, lists of variables and coding schemes, details of weighting, confidentiality and anonymisation, and provenance of any secondary data used. It also includes licence arrangements and all materials obtained through the original negotiation and data deposit, as well as post-deposit information created during preservation and processing activities. The terms **metadata** and documentation are often used interchangeably and there is overlap between the two,

though documentation tends to have a structure that is specific to each **data collection**, whereas all UKDA metadata files have a common structure (the **DDI**). For the purposes of this policy, we extend documentation to include all print and digital materials that accompany the original data deposit or are created during its preservation and processing as a UKDA **study**.

File

A physical unit of storage on a computer disc or tape.

Fixity metadata

Fixity metadata documents the means by which a digital resource can be authenticated, and safeguarded from undocumented alteration.

Media refreshment

Refreshment is the act of transferring data from one storage medium to another, of the same or different type, without alteration to the format of the data.

Metadata

Information that describes significant aspects of a resource are called metadata. They are created for the purposes of resource discovery, managing access and ensuring efficient preservation of resources. Metadata, particularly **resource discovery** and **resource use metadata,** may exist at various levels, typically from that of the data collection through to the individual variables of each data file in that collection. Metadata standards such as the **DDI** typically have a hierarchical structure as a result. Various types of metadata are important for efficient digital preservation (see **fixity metadata**, **preservation metadata, and resource management metadata),** and data sharing (see **resource use metadata** and **resource discovery metadata).**

Migration

Data migration is a means of overcoming technological obsolescence by periodically transforming data from one hardware/software environment to another, for the purposes of preservation for future use.

OAIS Reference Model

The Open Archival Information System (OAIS) Reference Model is a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information. It addresses a full range of archival preservation functions including ingest, archival storage, data management, access and dissemination. It is not a metadata standard but rather it outlines a taxonomy that defines the information types deemed necessary for the understanding of digital content over an indefinite period of time. Details of how the UKDA and The National Archives (TNA) are compliant with OAIS standards are given in *Assessment of UKDA and TNA Compliance with OAIS and METS Standards.*

Preservation metadata
Preservation metadata are intended to support and facilitate the long-term retention of digital information. Preservation metadata are part of **resource management metadata**.

Preservation strategy
A digital preservation strategy is a particular technical approach to the preservation of digital materials. This document contains the strategy and policy of the UKDA.

Resource discovery metadata
The catalogue and index record component of a **data collection**.

Resource management metadata
Sometimes referred to as administrative metadata, resource management metadata comprise information that is required for management of a digital resource throughout its life cycle, including its preservation and processing history.

Resource use metadata
Information required to support the use of a resource, including details of its content, structure, any technical dependencies, and meanings of coded numbers.

Statistical Package
A package specifically designed for analysing and often structuring and storing social survey and other social science data. The three most widely used commercial packages are SPSS, SAS and STATA. These three packages store certain variable-level **metadata** in addition to the data.

Study
The study is the UKDA's basic content-tracking concept used for a **data collection** during the stages of acquisition, processing, preservation and dissemination. There is normally a one-to-one relationship between a study and a **data collection**. Similarly, there often will be a one-to-one relationship between a deposit and a study, but a deposit may, depending on circumstances, be divided into more than one study. Equally, more than one deposit from the same depositor may be combined into a single study, similar to accumulating an archival fonds.

II. POLICY STATEMENTS

7   REGULATORY FRAMEWORK

The functions and goals of the UKDA are defined in the *UK Data Archive Strategic Plan*. The regulatory framework for the management of the material accessioned into the Archive is composed of:

- a deposit agreement and licence form which confirms the rights and obligations of both parties and offers an opportunity for depositors to specify the conditions under which access may be given to third parties;
- assertion of copyright and intellectual property rights to ensure that the data creator has cleared all necessary copyright permissions;
- where necessary, negotiations for licence agreements with third parties to enable the Archive to distribute the material for the purposes of research and teaching.

In preserving its collections, the UKDA considers the following:
- *Current best practice*
  Preserved resources should be managed in accordance with relevant codes of practice for preservation management (see also Section 21);
- *Freedom of Information Act, 2000*
  Use of collections has to adhere to procedures under the Freedom of Information Act, 2000 and the associated *Lord Chancellor's Code of Practice on the Management of Records issued under Freedom of Information Act 2000*;
- Data Protection Act, 1998
  Collections need to be managed in accordance with procedures under the *Data Protection Act, 1998*;
- *Copyright, Design and Patents Act, 1988;*
  This act governs the creation, ownership, transfer and exploitation of copyright;
- *Commercial agreements and contract law;*
- *Information security management standards;*
  The IT architecture of the UKDA has to be secure and follow the requirements set by the international standard ISO/IEC 17799:2000;
- *Guidelines on copyright and confidentiality: legal issues for social science researchers*
  Researchers must be aware of copyright and confidentiality issues and seek advice accordingly to protect their own interests as well as the rights of their respondents.

Access to most of the UKDA's data collections is regulated through access agreements that are compulsory for all users. They are referenced in Section 21 of this document. Some data, nearly all the documentation, and all the resource discovery metadata are freely downloadable to users without registration or access agreement.

8   IT ARCHITECTURE

The preservation of the UKDA's collections relies on an IT infrastructure that is monitored and periodically reviewed to ensure timely upgrades in both hardware and software. Digital preservation needs are one of the priorities in developing and upgrading the Archive's IT architecture. The UKDA preservation system was most recently upgraded during summer 2003.

8.1   The UK Data Archive preservation hardware

The preservation system consists of on-site, near-site and off-site storage to provide the resiliency and redundancy required of a major digital archive. The four preservation servers are: on-site main server, on-site mirror server, near-site mirror server, and off-site mirror server. Whilst the main preservation server runs Unix, the others are Windows-based. Their combined storage capacity is 3.2TB.

The *on-site main preservation server* provides near-line access, which means that files are not instantly available but are automatically retrieved from Super Digital Linear Tapes (SDLTs) from a 30-tape carousel without the help of a human operator.

The *on-site mirror preservation server* holds another copy of files to provide immediate, on-site redundancy. The mirror preservation server provides online storage which means that files are instantly available to the system via disc-based access on RAID 5.

The *near-site mirror preservation server* is located in a separate building from the main and mirror preservation servers. It too provides online storage via disc-based access on RAID 5.

The UKDA also operates an *off-site mirror preservation server* in another city in the UK that functions as the off-site security copy for all digital collections archived at the UKDA. Again, files are held online via disc-based access on RAID-5.

Technical details of the current configuration of the current UKDA's preservation servers can be provided upon written request to the Director.

8.1   Storage capacity

The main preservation file system is based on an AMASS® Hierarchical Storage Management (HSM) system where the files appear to be local to the user but are mainly based on tape. As each file is requested it is either brought back from the cache or automatically 'restored' from the required tape. The system provides 3.3 TB of native storage capacity with an expected near-line capacity of 7 TB. Additional unlimited capacity can be gained from external media which can be operator loaded on system request.

8.2   Network and communication equipment

The UKDA has been flood-wired with Category 5 UTP cabling. Three Gigabit switches and 10/100 full speed switches interconnect the network. Each one of the four server rack has its own dedicated Gigabit switch. There is also one standby 24-port 10 MB switch. This provides the following connectivity capacity:

- up to 2000 MB/s connectivity to each server and workstation;
- ability to restrict connectivity to valid Mac addresses;

- facility to segment the network for switched separated firewall connectivity.

8.4 Power surge protection systems

All servers in the UKDA are protected by power surge protection systems which are built in to the line interactive UPS system.

8.5 References

Further information on the IT architecture used at the Archive and the security measures to protect it are detailed in the *UK Data Archive Information Security Management Policy*.

8.6 IT responsibility

The UKDA Systems Administrator manages the on-site main preservation server. The Systems Manager takes responsibility for the on-site mirror and near-site mirror preservation servers. The off-site mirror preservation server is managed remotely by the UKDA Systems Manager and by a local Systems Manager at the remote site.

9 PRESERVATION METADATA

9.1 Documentation as metadata

Documentation is an essential aspect of preservation and collections management. It is required to support the following functions.

- Resource discovery
  Users must be able to locate resources and determine their usefulness. Information must also be available about any terms and conditions that attach to those resources, and the means by which they may be accessed.
- Resource use
  Information is required to support the use of a resource, including details of its content, structure, any technical dependencies, and meanings of coded numbers.
- Resource management
  The preservation and management of digital resources requires the maintenance of a wide variety of information, including their preservation and processing histories.

UKDA systems are compatible with the OAIS Reference Model .The OAIS taxonomy maps to most Archive processes but has not been used in this document. The model's main principles are consistent with the primary functions carried out by the Archive including those outlined here under 'Preservation Metadata' and 'Preservation Strategy' (Section 9).

The UKDA provides data creators with guidelines for documenting their data collection in (guidance is given at http://www.esds.ac.uk/aandp/create/research.asp) and requires a minimum set of metadata for each accessioned resource through its study descriptions as outlined in its *Deposit Forms*.

Documentation required for resource discovery and resource use is collected, created and held in the UKDA as metadata attached to each study. The study 'read file', 'user guide', 'depositor licence' and the 'study description' or catalogue entry, are the main components of such metadata. The study description is based on the DDI metadata standard for data documentation. The study description is also mapped to the Dublin Core metadata standard, and the Archive's online catalogue is Z39.50 compliant and compatible with the Open Archives Initiative (OAI). The catalogue is produced and maintained employing eXtensible Markup Language (XML) based on the DDI Data Type Definition (DTD). Resource discovery in the catalogue is further enhanced by the use of the Humanities And Social Science Electronic Thesaurus (HASSET) that was created and is maintained by the UKDA.

The resource management metadata are created by the Archive in the course of its preservation and processing activities and portions of these are recorded in the data processing database. Also there is a standardised UKDA study directory structure that reflects (and imposes) file formats and versions. This is maintained by the Data Standards Manager in association with the Systems Manager.

A special data processing directory on the preservation system contains standard outputs from processing including, most importantly, a list of file names and descriptions. The elements of the metadata recorded in the processing database during processing are shown in Table 1. This information is updated post-processing as required.

Element Name	Comments
Identification	
Study number	*UKDA study number*
Study title	*As specified in the catalogue record*
Resource content	
Details of study received	*Author(s), depositor(s) of the study and time of its deposit*
Details of data files received:	
File name	
File format	*Including version number*
Description of file content	
Details of documentation files received:	
File name	
File format	*Including version number*
Description of file content	
Details of hard copy documentation received:	
Content of hard copy texts	
Duplication with electronic files	*List of any file(s) of which hard copy is a printout*
Processing history	
File name changes	*Files for issue to users as well as files not for issue*
Level of processing	*One of the four levels of processing used at the UKDA*
Data file conversions:	
Conversion method	*Including what software was used for conversion*
File format created	*Including version number*
Validation	*Checks as required by each level of processing*
Documentation file conversions:	
Conversion method	*e.g. scanning of hard copy*
File format created	*e.g. PDF and TIFF*
Dataset validation	
Problems encountered	*Note of any confidentiality problems, data errors, wild-codes, incorrect labels, absent labels, etc.*
Solutions found	*Details of solutions to above problems*
Unsolved data or documentation problems	*Details of unsolved problems*
New edition information	*Details of why a new edition was triggered*
Data Delivery	
Notes to data delivery	*Additional comments that are helpful for data delivery to users*
Notes from data delivery	*Notes concerning any non-routine work performed to service a user*
Name of the member of staff who processed the note file and date	

Table 1. A portion of the resource management metadata as recorded in the processing database.

9.2   Administrative preservation database

The Systems and Preservation group is in the process of defining and implementing an administrative preservation database to manage all preservation decisions taken at the UKDA including those relating to storage media management.

9.3   Fixity metadata

The Systems and Preservation group is responsible for periodically generating a log with vital statistics and calculated checksum values for each study and preserving these alongside the other preservation metadata files. The log files (see Figure 1. below) for each directory associated with a study include: the study number, directory path, file name, size of file, date stamp including date created and modified, directory flag, and a checksum value, i.e. md5 sum value in 32 bit hexadecimal form.

```
3500\dat@0@858166912@858166912@D@NOCRC
3500\dat\fr7172.dat@3209769@991959008@858166850@ @134a3c87faf275aa950a113faa82a40d
3500\dat\fr7172.sps@971@991959019@858166850@ @ce26d1a014925017627e051dfd55438b
3500\dp\fr7172.des@5565@991959178@858166914@ @a179162fcf2ce4a1fe21b763214c638f
3500\dp\fr7172.lst@131161@991959178@858166914@ @e00524d402faa3899acf2f9bd3817b24
3500\exp\fr7172.exp@2759508@991959192@858166934@ @992d4bace55a839ea016170be6f71d70
3500\mrdoc\image@0@865345296@865345296@D@NOCRC
3500\mrdoc\image\3500img.zip@2615815@991959318@865345296@ @3089d663a39dc781772adfdba5bae7ce
3500\mrdoc\pdf@0@915635038@915635038@D@NOCRC
3500\mrdoc\pdf\a3500uab.pdf@5413472@991959319@915632856@ @0603e6fce7842f02a1ccffcc2ba39aab
3500\noissue@0@858166987@858166987@D@NOCRC
3500\noissue\dat@0@866023704@866023704@D@NOCRC
3500\noissue\dat\fr-7677@4510505@991959319@858166990@ @d51b6b5242ad3cdb5e37c7fd7dd261d7
3500\note3500.txt@4359@991959007@865431152@ @46f88e215c321110b5741709940f1e63
3500\read3500.txt@1082@991959008@865506394@ @26032b8fc90fa7db11d9257fb1a06979
3500\rf@0@867944847@867944847@D@NOCRC
3500\rf\3500rf.zip@189920@991959335@867944848@ @9016af95b2fd8ab01e0be04f4b18764d
```

Figure 1. A sample fixity metadata log for UKDA Study 3500

9.4   Storage media labelling

The spine of the storage media used is labelled with the UKDA study number. In the case of CD-Rs and DVD-Rs, the internal booklet is also filled in with details about the content of the disc with dates. Digital Linear Tapes (DLTs) are also labelled with a reference number to the system.

9.5   Significance of preservation metadata

Preservation metadata serve as the blueprint of the preservation strategy and disaster recovery procedures. They document decisions and actions performed in processing, converting and migrating digital collections, forming what is essentially a preservation audit trail for each study. They also provide a record of the file location and properties required to authenticate and migrate files to new formats and media.

10   PRESERVATION STRATEGY

10.1   Preservation strategy overview

The UKDA has chosen to implement a preservation strategy based upon open and available file formats (described subsequently), data migration and media refreshment. Preservation decisions at the Archive must always be made within the context of its *Collections Development Policy*, balancing the constraints of cost, scholarly and historical value, and user accessibility. Different preservation techniques may be required for material with different levels of quality and significance. Data collections are assigned a processing level as outlined in the Archive's *Acquisition Review Process* document. Specifications for processing standards including validation and integrity checks are given in a series of Archive *Data Services Process Guides*.

The UKDA recognises that in principle no file format or physical storage media is going to last for ever. Indeed, it has seen movement from punch cards and paper tape, through 7 and 9 track tapes to optical media and high capacity magnetic tape cartridges. As a consequence a strategy has been adopted to store data on at least two and often three different storage media. These are reviewed regularly and data are copied onto new media when appropriate.

The minimum number of preservation formats that are necessary to manage the full range of data types in the Archive's collections through time has been identified; migration paths for these are carefully chosen. The Archive endeavours to follow international best practice in its choice of preservation formats and data migration procedures.

Defining, timing, testing and implementing migration pathways are the responsibility of the Systems and Preservation group. When new formats are created from data files either through migration into new file formats or through creating new file formats for dissemination, the old files are retained alongside.

The preservation strategy of the UKDA aims to maintain a flexible preservation system that evolves to meet the demands of changing technology and new and increasing user expectations.

10.2   Ingest formats

The UKDA accepts data and documentation files in a variety of formats. Preferred file formats are those that are not platform or software dependent or that can be easily transferred to a suitable preservation format; however, most formats can be handled by the Archive. The Data & Support Services group works with depositors to recommend a suitable ingest format and to assist with the transfer of data and documentation to the Archive. The success and cost effectiveness of the UKDA's preservation strategy are both in part dependent upon the formats that digital materials arrive in. The UKDA provides advice on suitable formats, which boils down to using widely used 'open' and 'available' formats (terms explained below). The following are the key points from the *Guidance on Data Management* document (also summarised as web content for ease of access by potential data depositors)*,* in which the list of formats is monitored and maintained by the Data Standards Manager.

The general advice is as follows:

<div style="border:1px solid">

Use non-proprietary standards
Data formats can be grouped into three broad categories relevant to the issues of sharing and preserving datasets.

• **Proprietary standards** - formats such as Microsoft Word that are owned by a company and not made generally available. Data held in these formats can only be accessed through software that has been licensed to read the format.

• *Available standards* (also known as 'somehow-open') - the specification of some proprietary standards, such as Microsoft Rich Text Format (RTF) and Adobe Portable Document Format (PDF) are made available to other software developers and the general public. Available standards are still proprietary and restrictions may be placed on their use in the future.
Some formats such as SPSS portable are effectively available since the structure is sufficiently transparent (tagged ASCII) to be easily understood and decoded by a non-specialist (and it has been a stable format since its inception over thirty years ago).

• *Open standards* - standards created by a co-operative group that are then made freely available to anybody to use without restriction. Open formats should be used whenever possible. Available formats, such as Microsoft RTF and Adobe PDF should be considered next. Proprietary formats should only be considered as a last resort.

Use widely-supported formats
Preference should be given to widely-supported data formats because these are more easily shared and less likely to become obsolete unexpectedly. Careful consideration should be given to the adoption of open standards, some of which may not be adequately supported by available software. The individual researcher usually has no more control over the evolution of open formats than proprietary ones.

Avoid specialist proprietary formats
Reliance on specialist proprietary formats should be avoided as this commits a project to using one set of software products, even if better options emerge, or the software company ceases to support the software.

</div>

When necessary, the Archive provides data conversion support for depositors. The list of recommended ingest formats for specific types of data is summarised here.

Type of Data	Recommended Ingest Formats
Tabular data with minimal metadata (i.e. a matrix of data with or without column headings/ variable names), but no other metadata - in contrast to below)	Delimited text of given character set[1], with SQL data definition statements where appropriate. Widely-used proprietary formats (e.g. Excel, Access, dBase) are also acceptable.
Tabular data with extensive metadata (e.g. a survey dataset with variable labels, code labels, and missing values, in addition to the matrix of data)	Delimited text and command file (SPSS, STATA, SAS, etc.) containing the metadata information. Binary formats of statistical packages (SPSS, STATA, SAS, etc.) are also acceptable.
GIS and CAD data (vector and raster)	ArcInfo export format (.e00) for vector data Mapinfo Interchange Format (MIF) for vector data TIFF for raster data Adobe Illustrator, DXF or SVG for CAD data GIS attribute data as per 'tabular data with minimal metadata' above Binary formats of GIS and CAD packages may be acceptable, but offer less long-term security
Qualitative (textual) data	Plain text, RTF or HTML, software specific formats such as NUD*IST, NVivo and Atlas.ti may be acceptable, but offer less long-term security.
Digital audio data	MS Waveform (.wav), MPEG-1 Audio Layer 3 (MP3)
Digital video data	MPEG-2
Digital image data	TIFF
Documentation	PDF, RTF or HTML. MS Word is also acceptable.

[1] all text should be encoded as ASCII or Unicode. When data may contain non-ASCII characters (generally, any non-Latin characters) it should always be encoded as Unicode. Newer versions of software are likely to use Unicode by default. Note that XML requires the use of Unicode.

Table 2. Ingest formats at the UKDA.

The UKDA Acquisitions Review Committee evaluates all new data acquisitions and submissions based on fulfilment of stated acquisition policy criteria, usability of the data and documentation for secondary use, and perceived user demand. A work plan is created for each 'accepted' resource that specifies:

- that all files will be preserved in their original format;
- that all files will be converted to the appropriate preservation format, if necessary;
- that additional dissemination formats for data and documentation will be created;
- the composition of the *user guide* that will be created for each resource;
- the level of checking and validation of the data, and any additional documentation that needs to be created.

10.3   Validation and checking

The acquired data and documentation are thoroughly checked for anything that might cause problems for users: chiefly, the data and documentation are compared to ensure they are

mutually consistent, and the data are checked for any confidentiality problems. The UKDA currently uses four levels of processing, graded from A*, A, B and C, to which slightly different checking/validation requirements apply. Levels of processing are assigned by the Archive's Acquisitions Review Committee in response to a number of criteria, by far the most important of which is the anticipated usage of the data collection over a foreseeable period. These levels of processing are described in detail in the Archive's *Data Services Process Guides.* They are also summarised in the 'Read file', an HTML (or for old studies plain text) file made available with each order or browsable from the UKDA catalogue, such that users know the level of checks carried out for each study.

10.4   Conversion

The acquired files are converted to standard preservation and dissemination formats (i.e. user-friendly formats) whilst ensuring that the integrity of the data is maintained. The procedures for creating and checking additional data formats are governed by the *Data Services Process Guides*. Procedures are in place to ensure that information is not lost in the course of conversion and file formats are chosen accordingly. All commonly used format conversions (between SPSS, STATA, SAS and delimited text; and between Word and RTF), are performed by script to ensure no selection of inconsistent options in export options or file translation software options, and to reduce repetitive manual tasks for data processing staff, which cause staff boredom and human error. These scripts execute some conversions using source code and for others shell to packages (chiefly StatTransfer and SPSS) to utilise their command processors. Where other packages are used within a script, the data file conversion results have always been pre-checked extensively. Standard manual checks are performed on all files after all conversions to verify the script was not interrupted.

10.5   Preservation formats

All deposited data and documentation files are retained and stored in the appropriate directory on the preservation system in their original formats. Files are also converted for preservation in their appropriate archival format for the specified data type. Alongside, the UKDA preserves files in a variety of dissemination formats and migrates them accordingly. The following table displays the current preservation formats used for different types of data at the Archive.

Type of Data	Usual Preservation Formats
Tabular data with minimal metadata (i.e. a matrix of data with or without column headings/variable names), but no other metadata - in contrast to below)	Delimited text of given character set with SQL data definition statements where appropriate
Tabular data with extensive metadata (e.g. a survey dataset with variable labels, code labels, and missing values, in addition to the matrix of data)	Delimited text and command file (SPSS, STATA, SAS, etc.) or other structured text/markup file containing the metadata information (e.g. DDI XML file). [in the future an XML standard for statistical datasets may be developed][1]
GIS and CAD data (vector and raster)	ArcInfo export format for vector data Mapinfo Interchange Format (MIF) for vector data TIFF (version 6) for raster data DXF or SVG for CAD data GIS attribute data as per 'tabular data with minimal metadata' above
Qualitative (textual) data	XML marked-up text according to an appropriate DTD or schema, RTF
Digital audio data	Microsoft Waveform, Audio Interchange File Format (.aiff)
Digital video data	MPEG-2
Digital image data	TIFF (version 6)
Documentation	Plain text, RTF, HTML, PDF or XML marked-up text according to an appropriate DTD or schema (e.g. XHMTL 1.0)

Notes

1 All preservation formats are subject to change over time as new archival and interchange formats are developed. In particular an XML schema for statistical datasets (tabular data with extensive metadata) would be extremely useful. The closest development so far, though still not able to store all the internal metadata and variable format information of a typical SPSS, SAS or STATA file, is the Triple-S data model: www.triple-s.org/sssdown.htm. Recent versions of SAS, SPSS and STATA have their own XML data models, which may become useful intermediate formats for conversion (via XSLT) to a new common XML standard, as well as generating DDI metadata files via XSLT (which the UKDA already does).

Table 3. Preservation formats used at the UKDA

The UKDA prefers to store files in formats that are based on stable, open and available standards, or more simply still, structured text files (tab-delimited or fixed width). Table 3 shows formats used in current practice. Previously, SPSS portable as an effectively

'available' tagged ASCII format was used as a preservation format for tabular data with extensive metadata, but since this cannot store variable names with more than 8 characters, the default preservation formats were changed in 2005 to those shown in table 3. Older preservation formats are migrated when they cease to be easily convertible to the current preservation format(s).

The UKDA currently also publishes key datasets from its collection with their associated documentation and study description into the Nesstar system allowing online access to data via the *ESDS Nesstar Catalogue*. Originally named NESSTAR for Networked Social Science Tools And Resources, the Nesstar system is based on the DDI DTD standard and employs XML. An XML generator is used to create an XML file that contains the study description and documentation for the study. These files are also archived in the preservation system.

The UKDA routinely uses the PDF format to store documentation for dissemination to users. Adobe Systems Incorporated offers an Adobe PDF to HTML conversion tool on its web site. The Archive is monitoring this development as well as the viability of PDF over time.

The UKDA also generates its own data dictionaries/codebooks and certain other conversion information files in Rich Text Format (RTF). While for dissemination only purposes, the RTF tags are generated at source to ensure the file will open in any RTF-enabled application.

All paper correspondence and any other materials not intended for dissemination are scanned (to TIFFs) and copied to a special subdirectory on the preservation system to create a complete digital copy of all the study materials.

10.6   Dissemination formats and version control

Dissemination data formats are created via script as described for preservation formats at Section 10.5. These formats are also stored on the preservation system, but to allow immediate access are automatically copied (overnight) to the Archive's download server, where each study is automatically assembled into zipped file bundles, one for each dissemination format. Each night an automated check is made that file names and time and date stamps match between the two servers, such that updated versions of files on the preservation server are copied across to the zipped download file bundles on the download server. In this way instant access to data can be given without compromising the security or performance of the preservation server, but version control of files between the two servers is guaranteed.

10.7   Compression

The UKDA uses industry standard lossless compression tools, which are outlined in the *UK Data Archive Information Security Management Policy*. Compression software is used under tightly controlled conditions on files stored on CD-Rs and as a concatenation tool for coherent collections of multiple TIFF files. The Archive uses sufficient redundancy for its long-term storage to warrant use of these tools. Archive staff monitor the technology continuously to ensure timely updates of compression software to enable the migration from older versions.

10.8   Transfer media

The UKDA accepts data and documentation on a variety of media, the most popular being the following:

- FTP
- CD-ROM/DVD
- Floppy disc
- Email attachment

- ZIP/JAZ cartridge
- DAT
- Exabyte
- 9-track tapes

10.9   Storage media

The UKDA uses the following storage media for preservation:

- Hard Disc Drive (HDD)/RAID 5
- Super Digital Linear Tape (SDLT)

- CD-R
- DVD-R

The main preservation system stores data on hard disc and on SDLTs. As a practical measure, the Archive also uses CD-R for short- to medium-term storage and an ease-of-use alternative to its primary storage media. This redundancy of duplicating files on an off-line medium, with increasingly widespread use, provides a convenient copy of all the files accessible on its robust preservation system. While the CD-Rs may not yet be considered suitable for long-term archival storage, they do have proven short-term viability.

In addition, the Archive is testing DVDs for their archival and long-term value. Copies of very large data collections are now being stored on DVD-R in addition to the other copies held on the main and mirror preservation systems. This ensures that in case of DVD failure or corruption in the future, no files will have been lost. It also provides the Archive with hands-on experience with this new medium. The value of DVDs as an archival storage media is being monitored.

The UKDA endeavours to follow international best practice and recommendations in its choice of preservation storage media and short-term storage media. The storage media currently in use conforms with the following international standards:

Storage Media	Standard
SDLT1 tape cartridge	ISO/IEC 22051:2002
CD-R	ISO 9660:1988
DVD-R	ISO/IEC 20563:2001

Table 4. Standards for storage media used in the UKDA

10.10   Media monitoring and refreshing strategy

The UKDA operates a media monitoring procedure as part of its AMASS® preservation system. This allows it to check for potential upcoming problems of wear and tear on media and act before the problems become severe.

DLTs that are used for preservation are re-tensioned every six months and each full tape in the system is also copied every year onto a new tape. This is scheduled annually and a log of all actions is kept for checking. The Systems Administrator is responsible for performing the media refreshing procedure.

Idle tape media are automatically ejected from the DLT drives and placed in the carousel at set regular intervals to prevent excessive wear of both tapes and the drive.

The CD-R/DVD-R media are checked on a scheduled basis, every two years. If any media have either recoverable or non-recoverable errors then they are regenerated from the on-site mirror preservation server. A log is kept of all refreshment results and all storage media are provided with a date stamp indicating the time they were written and the next renewal date. CD-Rs are used within three months of purchase to ensure a short time period between when they are acquired and when they are written.

10.11   Consistent organised directory structure and standardized file extensions

Every study within the Archive follows a consistent directory structure for storage, and this is enforced by automated checks when files are copied onto the preservation system. This has many benefits, such as the ability to locate set types of information and also to allow automated tasks (e.g. migration of file formats) to be run without the need for complicated locator scripts. In addition to this structure, file label details are kept in an in-house system to provide extra information about a file in addition to its filename. Further, file extensions are always standardised, with a single extension allowable for each type of file.

11   STORAGE

In accordance with various British and international standards the UKDA provides a stable and secure environment for its collections. This is achieved through the establishment and maintenance of specified levels of temperature, relative humidity, lighting, air quality and vibration in the repositories. Data are stored with a view to their physical protection and ease of retrieval.

11.1   Environmental conditions for storage

In providing safe and secure storage for its collections, the UKDA meets British standards for environmental conditions for storage media (BS 4783) and for the storage of archival materials (BS 5454).

11.1.1   Recommended environmental conditions for storage media

Each device, component and storage medium used in the UKDA has both a manufacturers' recommended range as well as a British and/or international standard range for environmental conditions. The Archive has focused on complying with the strictest standard; in most instances this is the British Standard.

Comparative figures for suggested environmental conditions are given in the following two tables.

Storage media	Operating	Non-operating	Long-term Storage
SDLT1 tape cartridge	10 to 40°C 20 to 80% RH	16 to 32°C 20 to 80% RH	16 to 32°C 20 to 80% RH
CD-R	-5 to 55°C 10 to 90% RH	5 to 25°C 8 to 60% RH	18 to 22°C 35 to 45% RH

Table 5. Manufacturers' recommended environmental conditions for storage media used in the UKDA

Storage media	Operating	Non-operating	Long-term Storage
SDLT1 tape cartridge	10 to 40°C 45 to 55% RH	16 to 32°C 20 to 80% RH	18 to 22°C 35 to 45% RH
CD-R	10 to 45°C 20 to 80% RH	5 to 45°C 20 to 80% RH	18 to 22°C 35 to 45% RH

Table 6. Summary of environmental conditions recommended in BS 4783-5:1991 and BS 4783-7:1993

11.1 Environmental conditions in the UKDA repository

The temperature in archival storage areas is maintained at a reduced level compared with that in areas where staff or users may need to work continuously. Appropriate action is taken to ensure that a stable relative humidity is maintained with a total variation of no more than 5%. These environmental conditions comply with the recommendations of the BS 5454:2000 Chapter 11.

Temperature
Air temperature in the repository is maintained at 20°C by a circulating airflow within the repository room. The blow setting is kept low to prevent it acting as a fanning agent in the case of a fire.

Humidity
The relative humidity is monitored and ranges from a lower limit of 35% to an upper limit of 45% RH. A dehumidifier with a dust filter operates to make adjustments if necessary.

Other elements
The avoidance of dust and dirt of all kinds is essential and as a minimum the air is filtered and circulated on a regular basis. No smoking, food or drink is permitted in the storage areas.

11.2 Physical data preservation and storage

Preservation hardware
The configuration of the UKDA preservation hardware is outlined in Section 8 of this policy.

Multiple copy resilience
The UKDA now has over 4,600 studies, comprising over 266,000 individual files. Preservation file formats and dissemination file formats may in some cases overlap. However, each file is held on three different media – SDLT, HDD, CD-R/DVD-R, and across several copies.

Main near-line copy (on main preservation server)
This copy is held on the main area on the Hierarchical Storage Management (HSM) system and is presently accessed only by the dedicated preservation user. The storage media used for this copy is SDLT and disc cache area.

Shadow copy (on main preservation server)
A shadow copy is made of every file on the main preservation system and previous versions of files are kept and maintained. All files that are updated are shadowed onto a separate SDLT in the main HSM system.

Access online copy (on mirror preservation server)
A mirror of the files stored on the main preservation server is held on the mirror preservation server in a RAID 5 disc system. Copies are generated for user access and dissemination.

Near-site online copy (near-site mirror preservation server)

A mirror of the files stored on the main preservation server is also kept on a RAID 5 disc system on a server located in another building.

Off-site online copy
An off-site, online copy is kept in case of a major disaster at Essex. The same environment and security rules are applied at the off-site repository as are effective in the repository in the UKDA building.

CD-R off-line copy
A CD-R copy is created for each study as part of the preservation procedure. This allows convenient access to an alternative local copy in the case of downtime of the main preservation system and serves as an alternative short- to medium-term storage medium. For each study all of the files are compressed and stored as a single zip file and written on to a CD-R. Subsequent updates to a study are created as complete zip files with a different filename and appended to the existing CD-R for that study. As an experiment to test a new storage medium, the Archive has begun using DVD-Rs selectively as replacement storage media for the CD-R copy.

Storage requirements for magnetic and optical media
The magnetic tapes are stored safe from the effects of magnetic fields. The containers of tapes are close fitting to exclude dust.

CD-Rs and DVD-Rs are stored in a locked cabinet and in manufacturers' plastic cases in a climate-controlled room.

11.3   Storage responsibility

Monitoring storage conditions and storage media, and initiating remedial action when necessary, is the responsibility of the Systems and Preservation section and the Archive's Systems Administrator.

12 SECURITY

The UKDA is committed to taking all necessary precautions to ensure the physical safety and security of all data collections that it preserves.

12.1 Fire prevention and protection system

Smoking is not allowed in the repository or anywhere in the whole building. The Archive and the University of Essex authorities display fire action notices throughout the buildings as required. Repository staff are instructed in the location of fire extinguishing appliances and in their use.

The repository room is protected by an Argon gas-based, fire-extinguishing system. This includes a four-zone fire detection system with automatic release that is activated when fire or smoke is detected within two of the four zones. A manual override for both system isolation and full immediate release is available. This system is serviced and tested twice yearly by a qualified fire protection system engineer. A fire door that gives at least one hour of fire protection is in place to prevent an external fire entering the room. All furniture material conforms to the required fire retardant British Standards. The network cabling used in the Archive is compliant with standard requirements.

An overview of the compliance of the Archive's fire-extinguishing systems and other security measures is presented in Table 7 below.

12.2 Physical intruder prevention and detection systems

The repository door is equipped with key entry and a code-protected keypad linked to an on-site alarm system and to the University Security Office. The surrounding area is locked between 7 p.m. and 7 a.m. and all weekend, and is regularly patrolled by University of Essex security staff. Access to this room is restricted to Archive key personnel only, or visiting personnel accompanied by key authorised Archive personnel. Further, access to the entire UKDA building is protected by keypad.

All machine room computer systems are locked by a logon password system to prevent unauthorised access in the case of a security breach of the room, as required by the ISO/IEC 17799:2000.

12.3 Environmental control systems

Areas and rooms designated for storage of Archive collections are structurally sound and practically free from the risk of flood and as far as possible from the risk of fire.

Environmental tracking
A seven-day thermohygrograph is used to keep track of the temperature and the relative humidity in the repository room. The wheels of the thermohygrograph are checked once a week by the Systems Administrator to ensure that the system maintains the environment within the required parameters. Additional electronic tracking devices are being investigated to allow an electronic audit trail to be maintained.

Environmental reporting
An environmental alarm system is part of the intruder detection system. This will activate an alarm in the event of out-of-range parameters.

British Standard	Description of Standard
BS 3621:1998	Specification for thief resistant locks.
BS 5306-4:2001	Specification for carbon dioxide systems.
BS 5454:2000, ch. 5.3	Protection against damage by fire. Doors and other openings.
BS 5454:2000, ch. 5.7	Security.
BS 5588-9:1999	Fire precautions in the design and construction of buildings.
BS 5839-1:1988	Fire detection and alarm systems in buildings. Compliance with set-up of alarm servicing and regular testing.
BS 5839-2:1983	Specification for manual call points
BS 5839-3:1988	Specification for automatic release mechanisms for certain fire protection equipment.
BS 5852:1990	Fire tests for furniture.
BS 6266:1992	Code of practice for fire protection for electronic data processing installations. Contingency plan for restoring a service after disaster recovery.
BS 7083:1996	Computer rooms – Guide for accommodation and operating environment of information technical equipment.
BS 7671:2001	Requirements for electrical installations.

Table 7. Compliance of the UKDA repository environment and security systems with the standard requirements for the storage and protection of archived data collections

12.4 Preservation system security

The UKDA preservation system is protected by the general University of Essex security systems. The preservation system is based on industry-standard operating systems and is configured not to leave any security holes. Information contained in this section is more thoroughly covered in the *UK Data Archive Information Security Management Policy* that complies with the ISO/IEC 17799:2000 standard.

Power surge protection
All servers in the Archive are protected by power surge protection systems through a line interactive UPS system.

Firewall
The preservation system is protected by the University of Essex firewall systems.

Intruder detection
The preservation system is equipped with the TripWire software detection system that ensures the security and integrity of files on the main preservation servers by reporting if, when, and how files have changed.

System access control
The Archive is part of the University of Essex centrally controlled staff user account system, which is maintained by the University Computing Services.

Password changing

Standard staff user account passwords must be changed every four months whilst administrative accounts are changed every two months. The full rules for this are documented in the *UK Data Archive Information Security Management Policy*.

Operating system security

The operating systems on the preservation servers have been trimmed down for increased security and have a TripWire monitoring system in place to ensure that operating systems are not breached.

Data synchronisation checking

When saving files to the main preservation system and mirroring servers, all copies are compared for completeness. Secondary preservations systems check the MD5 sum values, file size and date to ensure the integrity of the files. The off-site server is synchronised by secure FTP mirroring software.

Preservation account

The preservation account is a special account that is used only for preservation purposes and is only used on servers that are dedicated to preservation of data.

Main preservation system

File read access to the main preservation system is tightly controlled and restricted to the preservation account. All external access is provided through separate servers. Write access is only available to one user account — the preservation account.

Mirror preservation system

The mirror preservation system is the main Archive access server to the data collections. It reduces the workload of the main preservation system and has a complete, up-to-date copy of the holdings on the main system. Archive staff have read-only access to the studies stored here. Write access is not allowed on this server; instead changes are mirrored directly from the main preservation server.

13   DISASTER RECOVERY

13.1   Principles and procedures for disaster planning

The UKDA Systems and Preservation group staff constitute the Archive's IT disaster recovery team. Team members are responsible for ensuring that contingency plans and procedures are in place to prevent, react to and recover from emergency situations that may have an adverse effect on the Archive's collections. A general overview of responsibilities for disaster recovery and business continuity planning is outlined in the *UK Data Archive Information Security Management Policy,* an internal document that is modelled after and complies with the Information Security Management standard ISO/IEC 17799:2000. Details of the Archive's preservation infrastructure and IT architecture are provided in Sections 8 and 11 of this document.

The UKDA's IT architecture for safeguarding and managing its data collections is built on these basic preservation principles:

- physical reliability of storage media and servers;
- security of the data from unauthorised users;
- usability of data formats in the future.

The Archive IT systems employ various disaster prevention measures, including restricting preservation server access to key personnel, routinely monitoring storage media, and implementing a high degree of server redundancy to ensure the isolation of server malfunction or loss of data should they occur. There are at least five copies of files archived on four separate preservation servers as outlined in Section 10.2.

The Archive's disaster recovery procedures predominantly focus on reacting to scenarios of hardware and software failure, corruption or degradation of storage media, and potential problems caused by human error. The following three fundamental questions are addressed by the IT disaster recovery procedures.

How often are backups created?
Backups of archived files on the main preservation server are made every day.

How often are these checked for readability?
Every six months.

How long will it take to recreate the system from these backups?
Depending on the problem that occurs, there is little to no downtime of the server expected since all data are mirrored onto two on-site servers, one near-site server and one off-site server. If the main preservation server malfunctioned, it would be possible to immediately rely on the next functioning server in the sequence – the on-site mirror, near-site mirror or off-site mirror.

13.2   Disaster recovery procedures for files

A range of recovery measures designed to meet consequences of disasters that result with a loss of stored data files have been put in place.

Unreadable/corrupt file on the main preservation server

Problem: A single file is unreadable from the media (e.g. due to a bad block on a tape) or damaged.

Solution: a. The data carrier is checked to make certain that this is an isolated problem. If it is found to affect the complete tape the corrupt media disaster recovery procedure is activated.

b. If the problem is isolated then the problematic file is recreated from the mirror preservation server.

Corrupt file on the on-site mirror preservation server as well as the main preservation server

Problem: In this situation both the main and on-site mirror preservation areas cannot be read, nor any or the updated tapes.

Solution: The near-site preservation server would be utilised or, as an alternative, the off-site preservation server.

Corrupt media

Problem: A complete tape or CD-R/DVD-R is damaged or cannot be reliably read.

Solution: A new tape or CD-R/DVD-R can be regenerated from the mirror preservation server.

13.3   Disaster recovery procedures for complete loss of data at the UKDA in Essex

Problem: In this major disaster scenario, all of the data held at the UKDA in Essex are unreadable and all of the systems are damaged beyond repair.

Solution: The main HSM systems would be built and data would be retrieved from the off-site holdings.

13.4   Disaster recovery responsibility

The responsibility for initiating and taking charge of the disaster recovery procedures lies with Systems and Preservation group and its head. A detailed list of contacts and responsibilities is part of these procedures.

14   ACCESS TO COLLECTIONS

The UKDA recognises that use of its collections is a prime motive for its existence, as service provider for both the ESDS and AHDS History.

In providing access to the collections it preserves, the Archive is regulated by the deposit agreements that establish use conditions for every data collection (cf. Section 6) and the access agreements which its users have to accept.

The UKDA distributes and provides access to data from its collections via:

- Secure HTTP download;
- online access (via the *ESDS Nesstar Catalogue*);
- guest FTP;
- CD-R and DVD-R;
- other media by special request, e.g., DAT, Exabyte, Zip disc.

Dissemination or access copies of files are made from copies of the preservation files that reside on the mirror preservation server.

The Archive's secure HTTP-based download service provides a quick and reliable means of gaining access to the most heavily used collections held at the Archive. Quantitative data are typically offered in SPSS, STATA and tab-delimited text formats, qualitative data in RTF, PDF or XML formats, documentation in PDF format, and code in plain text format (with other formats used for all these materials where this is advantageous for the user). Other commonly used formats are made available on request. The Archive also provides online access to quantitative data that have been enhanced and published in the Nesstar system. Nesstar provides the capability for data discovery, browsing, subsetting, visualisation and downloading in a large variety of formats via the internet. The system is based on the DDI standard.

For further information on access to UKDA data collections see the *Ordering/Downloading data* and *ESDS Nesstar Catalogue* documents on the UKDA web site.

15   PRESERVATION RESEARCH

The UKDA is committed to reviewing and remaining up-to-date with technological advances. It endeavours to be capable of adapting or initiating new preservation standards and procedures that are suitable for its collections. The Archive supports and participates actively, where appropriate, in research, development and implementation of new practices for preservation of digital resources.

The Systems and Preservation group of the Archive is responsible for monitoring developments and advances in the digital preservation area through a technology watch scheme.

The UKDA is involved in several European and international projects to develop new metadata standards, and better facilities for sharing data, e.g. the Data Documentation Initiative (DDI) and the Digital Preservation Coalition. Archive staff participate in discussion groups and email lists that consider digital preservation issues.

16   CO-OPERATION

The UKDA has established productive working relationships with other institutions and organisations in order to address the Archive's preservation needs. The Archive recognises the need for communication with groups active in formulating national preservation policies and programmes. It also acknowledges the need to participate in activities and programmes in the area of digital preservation.

The UKDA hosts two specialised digital resource collection and distribution services and provides preservation facilities and services for them (web links are provided in Section 21):

- The *Economic and Social Data Service* (ESDS) provides preservation, dissemination, user and training for an extensive range of key economic and social data, both quantitative and qualitative, spanning many disciplines and themes. ESDS provides an integrated service offering enhanced support for the secondary use of data across the research, learning and teaching communities, covering a collection of several thousand datasets. ESDS also operates four specialist services targeted at specific user communities:
  - o *ESDS Government* facilitates more effective use of large scale government surveys, such as the General Household Survey, Expenditure and Food Survey, Family Resources Survey and the Labour Force Survey;
  - o *ESDS International* provides access and support for a range of international macro and micro data sources including National Statistics Time Series Data, major international databases from UNIDO, and OECD Main Economic Indicators which contain data on national accounts, industrial production, employment, prices, business trends, finance and trade for the major world economies;
  - o *ESDS Longitudinal* supports key UK longitudinal data collections including the British Household Panel Survey (BHPS), the British Cohort Study 1970 (BCS70), the National Child Development Survey (NCDS) and the Millennium Cohort Study (MCS);
  - o *ESDS Qualidata* provides enhanced access and support for a range of multimedia qualitative data sources. Data supported include: in-depth and semi-structured interviews, focus groups, field notes and observations, personal documents and photographs. ESDS Qualidata also provides advice on the archiving and re-use of qualitative data.
- The AHDS History (formerly the History Data Service), one of five Subject Centres of the Arts and Humanities Data Service (AHDS), is a national data archiving service jointly funded by the Joint Information Systems Committee and the Arts and Humanities Research Board.

The UKDA is a member of these national and international organisations that promote the collection, use and preservation of digital data and that offer opportunities for collaboration:

- Digital Preservation Coalition (DPC);
- DDI Alliance;
- Council of European Social Science Data Archives (CESSDA);
- International Federation of Data Organisations for the Social Sciences (IFDO);
- Inter-university Consortium of Political and Social Research (ICPSR).

Individual Archive staff are members of professional groups and hold offices in organisations promoting digital preservation and standards including:

- East of England Regional Archives Council;
- International Association for Social Science Information Service and Technology (IASSIST).

The UKDA shares its expertise, built up over thirty years of digital preservation practice, with institutions and research networks on both the national and international level. Recent activities and co-operative partners seeking expert advice from the Archive include:

- DLM-Forum of the European Commission;
- Medical Research Council (MRC);
- Royal Statistical Society (RSS).

The Archive frequently hosts visitors and in-house training for individuals from organisations from around the world seeking to establish new data archives or to implement best practices in data archiving at their own institution. The Archive also hosts European visiting fellows who wish to learn about data archiving through the University of Essex-based European Centre for Analysis in the Social Sciences (ECASS).

17   MONITORING AND REVIEW

The preservation policy of the UKDA is monitored and reviewed in the light of changing technologies on a biannual basis to ensure timely updates.

The Head of Systems and Preservation group or the Director of the UKDA initiates the review process.

Implementation of the preservation policy is monitored and there are regular planned audits to assess how the policy is executed including an annual benchmarking of the UKDA preservation agenda.

18   FUNDING AND RESOURCE PLANNING

The UKDA is committed to supporting continued funding for all of the operations relating to preservation management. Resource management for preservation of digital resources includes:

- technical infrastructure, including equipment purchases, maintenance and upgrades, software/hardware obsolescence monitoring, network connectivity, etc.;
- financial plan including strategy and methods for financing the digital preservation programmes and commitment to long-term funding;
- staffing infrastructure including hiring and ongoing training.

The Archive has established a rolling planning scheme for lifetimes of computer equipment and storage media to facilitate better forward planning for the necessary upgrades.

As stated in the *UK Data Archive Strategic Plan* the preservation of data and documentation to ensure they remain usable over time is a core activity of the Archive. The Archive will, therefore, make every effort to remain up-to-date with any relevant technological advances to ensure continued access to its collections.

19 ROLES AND RESPONSIBILITIES

All staff of the UKDA have a responsibility to implement the preservation policy and directives approved by the Archive's senior management. Systems and Preservation group staff take an active role in assisting and encouraging the implementation of the preservation policy Archive-wide. Accountability is shared amongst the Archive staff.

Director
The Director has a duty to ensure that the UKDA complies with the requirements of legislation affecting management of the collections and their preservation, and with supporting regulations and codes.

Head of the Systems and Preservation group
The Head of the Systems and Preservation group works closely with Heads of Sections to ensure that there is consistency in the management of preservation and collections and that advice and guidance on good preservation management practice is provided.

Systems Manager
The Systems Manager ensures the effective running of all the Archive's IT systems, software and hardware and is responsible for the day-to-day running of the Archive's preservation services, under the line management of the Head of the Systems and Preservation group.

Data Standards Manager
The Data Standards Manager establishes standards, procedures and tools for ensuring optimal conversion of materials from ingest to preservation and dissemination formats. Although outside the Systems and Preservation group, this post is involved in deciding what format digital materials are preserved in (in association with that section).

All staff
All UKDA and project staff are accountable to their line managers for compliance with this policy and with related policies, standards and guidelines.

20   STAFF TRAINING

The UKDA encourages its preservation staff to maintain a high level of awareness of developments in preservation practices, technological developments and procedures. Staff in the Systems and Preservation group provide preservation training for Archive staff, other archives and users as appropriate.

The Archive promotes and encourages good practice in resource creation and sharing among potential depositors of data collections to ensure that resources selected for preservation are in a stable and good condition when deposited.

The Archive will continue to encourage the development of co-operative programmes and workshops related to all aspects of digital preservation.

21   DISTRIBUTION OF THIS DOCUMENT

As one component of the UKDA collections management policy, this document will be communicated to all line managers and staff as appropriate. It will also be published and made available on the Archive's web site, and shared with other archives and digital preservation associations.

22   REFERENCES

Legal acts

Data Protection Act 1998
> http://www.opsi.gov.uk/acts/acts1998/19980029.htm

Freedom of Information Act 2000
> http://www.opsi.gov.uk/acts/acts2000/20000036.htm

Lord Chancellor's Code of Practice on the Management of Records issued under Freedom of Information Act 2000
> http://www.dca.gov.uk/foi/codemanrec.htm

Explanatory Notes to Freedom of Information Act 2000
> http://www.opsi.gov.uk/acts/en2000/2000en36.htm

Draft Code of Practice on the discharge of public authorities functions under Part I of Freedom of Information Act 2000
> http://www.dca.gov.uk/foi/codepafunc.htm

Copyright, Design and Patents Act 1988
> http://www.legislation.hmso.gov.uk/acts/acts1988/Ukpga_19880048_en_1.htm

UK Data Archive policy documents and web content

Access Agreement for Individuals (End user license)
> http://www.esds.ac.uk/aandp/access/licence.asp

Access Agreement for Teaching (Academic Sector)
> http://www.data-archive.ac.uk/orderingData/agreements/accessTeaching.pdf

Acquisition Review Process [Internal document]

Assessment of UKDA and TNA Compliance with OAIS and METS Standards
Report for JISC by Hilary Beedham (UKDA), Julie Missen (UKDA), Matt Palmer (TNA), and Raivo Ruusalepp (Estonian Business Archives Ltd.), published 2005.

Deposit Forms
> http://www.esds.ac.uk/aandp/create/depform.asp

*Data Services Process Guides***:** [Internal documents]
1. Translating data files into different formats
2. Data processing
3. Documentation processing
4. Standards and procedures for indexing and cataloguing
5. Nesstar procedures

Guidance on Data Management
> Draft document, due for uploading October 2005.
> Will be located under http://www.esds.ac.uk/aandp/

Simplified Web content is give at http://www.esds.ac.uk/aandp/create/data.asp

Submitting data
http://www.esds.ac.uk/aandp/create/depproc.asp

ESDS Nesstar Catalogue
http://nesstar.esds.ac.uk/webview/index.jsp

Ordering/Downloading Data
http://www.data-archive.ac.uk/orderingData/introduction.asp

UK Data Archive Collections Development Policy
http://www.esds.ac.uk/aandp/create/policy.asp

UK Data Archive Information Security Management Policy [Internal Document]

UK Data Archive Strategic Plan
http://www.data-archive.ac.uk/news/publications/strategicplan0409.pdf

Standards

BS 3621:1998, *Specification for thief resistant locks*

BS 5306-4:2001, *Fire extinguishing installations and equipment on premises. Specification for carbon dioxide systems*

BS 5454:2000, *Recommendations for the storage and exhibition of archival documents*

BS 5588-9:1999, *Fire precautions in the design, construction and use of buildings. Code of practice for ventilation and air conditioning ductwork*

BS 5839-1:1988, *Fire detection and alarm systems for buildings. Code of practice for system design, installation and servicing*

BS 5839-2:1983, *Fire detection and alarm systems for buildings. Specification for manual call points*

BS 5839-3:1988, *Fire detection and alarm systems for buildings. Specification for automatic release mechanisms for certain fire protection equipment*

BS 5852:1990, *Methods of test for assessment of the ignitability of upholstered seating by smouldering and flaming ignition sources*

BS 6266:1992, *Code of practice for fire protection for electronic data processing installations*

BS 7083:1996, *Guide to the accommodation and operating environment for information technology (IT) equipment*

BS 7671:2001, *Requirements for electrical installations. IEE Wiring Regulations*

BS ISO/IEC 17799-1:2000, *Information Technology — code of practice for information security management*

BS 7799-2:1999, *Information security management — specification for information security management systems*

Data Documentation Initiative (DDI)
http://www.icpsr.umich.edu/DDI/index.html

ISO/DIS 14721.2 Reference Model for an Open Archival Information System (OAIS)
DISC PD 0008:1999 Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically

Other policy and guidance texts

Council of European Social Science Data Archives, *CESSDA Goals and Tasks*
http://www.nsd.uib.no/cessda/

Economic and Social Research Council (ESRC), *Datasets Policy*
http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Annex%20C_tcm6-9738.pdf

International Association for Social Science Information Service and Technology (IASSIST), *IASSIST Constitution - Objectives*
http://www.iassistdata.org/membership/constitution.html#3

Guides to best practice and recommendations

DLM-Forum, *Guidelines on Best Practices for Using Electronic Information* (INSAR Supplement II, 1997)
http://europa.eu.int/ISPO/dlm/documents/guidelines.html

Charles M. Dollar, *Authentic Electronic Records: strategies for long-term access* (Cohasset Associates, Inc., 1999)

Ann Green, et al, *Preserving the Whole: a two-track approach to rescuing data and metadata* (Council on Library and Information Resources, 1999)
http://www.clir.org/pubs/reports/pub83/contents.html

Inter-university Consortium for Political and Social Research (ICPSR), *Guide to Social Science Data Preparation and Archiving,* 3rd edition (ICPSR, 2002)
http://www.icpsr.umich.edu/ACCESS/dpm.html

Chris Pickford, et al., *Best Practice Guideline 4: Preservation and Conservation* (Society of Archivists, 1997)
http://www.archives.org.uk/publications/inprint.asp#bestpractice

The Royal Statistical Society and the UK Data Archive, *Preserving and Sharing Statistical Material* (UK Data Archive, University of Essex, 2002)
http://www.data-archive.ac.uk/news/publications/PreservingSharing.pdf

Other organisations

Cedars Project (CURL Exemplars in Digital Archives)
http://www.leeds.ac.uk/cedars/

CESSDA – Council of European Social Science Data Archives
http://www.nsd.uib.no/cessda/l_eng.shtml

Digital Curation Centre
http://www.dcc.ac.uk

Digital Preservation Coalition
http://www.dpconline.org/graphics/index.html

DLM-Forum
http://europa.eu.int/historical_archives/dlm_forum/index.htm

ERPANET – Electronic Resource Preservation and Access Network
http://www.erpanet.org/

AHDS History
http://ahds.ac.uk/history

IASSIST – International Association for Social Science Information Service and Technology
http://www.iassistdata.org/iassist/

ICPSR – Inter-university Consortium for Political and Social Research
http://www.icpsr.umich.edu/

IFDO – International Federation of Data Organisations for the Social Sciences
http://www.ifdo.org

The Royal Statistical Society
http://www.rss.org.uk