UK • DATA ARCHIVE

Preservation policy

Public

Version: 14.00

Table of contents

1. Scope

The scope of this policy is limited to the Archive's data collections. It deals with all aspects of preservation and applies to all materials held for long-term digital preservation by the Archive on behalf of the University of Essex and others. The policy only covers the preservation of data collections for which the Archive is the primary custodian. It does not consider preservation of other materials such as the Archive's web content, internal administrative documents and correspondence. Some of these materials, including third-party standards, legislation, policies and procedures, have a direct impact on the preservation process, but are governed by the Archive's *Records Management Policy.*

The Archive's preservation strategy is not formulated in a single document but rather is embedded across a range of business information related to data curation and preservation.

2. Definition of terms

Active preservation: Active preservation is a form of digital preservation which seeks to ensure continued accessibility and use of electronic information through the active intervention of format migration.

Authenticity: A record's authenticity can be said to rely on three significant provable properties. The three properties are: that the record is what it purports to be, that it was created by whomsoever it is purported to have been created by; and that it was created at the time when it is purported to have been created. (*Adapted from ISO 15489-1:2001.*)

Code: This refers to uncompiled code in the command language of a *statistical package* or computer language, which occasionally forms part of a UK Data Archive *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 needed, 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. As part of *datasets*, 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 make up 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 Archive organises *data* for preservation as data collections. A data collection is typically made up of three components: *dataset(s), 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 Archive terms a *study*.

Data Documentation Initiative (DDI): The DDI is a project of the social science community to establish an international standard 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 Schema. The Archive has been centrally involved in the adoption, development and promotion of the DDI as an international standard for social science *metadata*.

Data migration: Data migration is the process of converting data from one structure to another structure to counter software obsolescence.

Dataset: A dataset is defined in this policy as any computer file (or set of files) which is organised under a single title and is capable of being described as a coherent unit within the Archive's catalogue.

Depositor: A depositor is an individual who is named on a deposit agreement as having sufficient responsibility to grant particular rights to the Archive on behalf of a data collection. The depositor may be the instigator, creator or the copyright owner of a data collection, but does not have to be. In OAIS terminology the term producer is used in a similar sense.

Designated user community: The UK Data Archive's designated user community is made up of social science and related data users within HE and FE in the UK, though best efforts are made for all users. (For the purpose of this definition, quantitative and qualitative data collections created by or for historians are considered to be 'social scientific'). All users are expected to have a basic understanding of social science methods and techniques relevant to the data collections being accessed. Additional requirements will be expected in the case of users of data subject to secure access.

Digital preservation: Digital preservation is a series of managed activities necessary to ensure enduring access to authentic versions of the content of 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 ingest 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 Archive metadata files have a common structure (the *DDI*). For the purposes of this policy, documentation is extended to include all print and digital materials that accompany the original data deposit or are created during its preservation and ingest as an Archive *study*.

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

Integrity: The integrity of a record refers to its completeness and to a continued state of unalteration. (*Adapted from ISO 15489-1:2001*). The Archive uses the term logical integrity to highlight the fact that a digital object within its system may change, but that the content information (data) remains unaltered. Similarly providing explicit documentation for any alteration to the content allows for completeness.

Media refreshment: Refreshment is the process of transferring data from one type of storage medium to another to ensure continued access to the information, without alteration to the format of the data.

Metadata: Information that describes significant aspects (e.g. content, context and structure of information) of a resource is called metadata. Metadata 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: Migration is a means of overcoming technological obsolescence by periodically transferring, and where appropriate transforming, digital material from one hardware/software environment to another, for the purposes of preservation for future use. Migration must maintain the authenticity, integrity, reliability and usability of any digital object.

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 UK Data Archive and TNA are compliant with OAIS standards are given in *Assessment of UKDA and TNA Compliance with OAIS and METS Standards*. The OAIS is an international standard (ISO 14721).

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 Archive.

Producer: See *depositor*.

Record: The UK Data Archive generally uses this term with its accepted computing meaning as a collection of related data. The word record is used in this glossary in its archival sense as information that provides evidence for the future. The content of that information is understood to be the primary element of the record, but both structure and context are necessary for integrity and usability.

Redaction: Redaction is the separation of disclosable from non-disclosable information prior to the release of a document. When related to datasets this is usually carried out by the removal of information from a record while retaining the structure of the record in the version being released.

Reformatting: Altering the underlying bit-stream of a dataset (by moving it between different physical carriers or transforming the character code), but making no changes to the physical representation or substantive content (*Adapted from ISO/TR 18492*). If this form of reformatting takes place it is documented.

Refreshment: See media refreshment.

Reliability: The reliability of a *record* relies on having trusted and dependable contents. (*Adapted from ISO 15489-1:2001*).

Resource discovery metadata: In the Archive, resource discovery metadata are the catalogue and index record component of a data collection.

Resource management metadata: Generally referred to as administrative metadata, resource management metadata comprise information that is required for management of a digital resource throughout its lifecycle, including its preservation and ingest history.

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

Significant properties: Significant properties are those elements of a digital object that need to be preserved in order for it to be used by the designated user community. They will almost always include information content and a level of functionality, but might also include formatting and look and feel.

Statistical package: A statistical package is a (software) 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 Archive's basic content-tracking concept used for a data collection during the stages of acquisition, ingest, preservation and dissemination. There is normally a one-to-one relationship between a study and a data collection. Similarly, there is often 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 archival fonds.

Usability: The usability of data relies on its ability to be located, retrieved, presented and interpreted. (*Adapted from ISO 15489-1:2001*).

3. Purpose

This document defines the UK Data Archive's (the Archive) policy on preservation, including for the designated community of users. This policy generally conforms to the OAIS reference model, with additions and alterations that are specific to the materials held within the Archive.

Terms in italics are defined in the 'Definition of terms' section above or refer to related controlled documentation.

The UK Data Archive's Mission is to support high quality social and economic research, teaching and learning through ensuring long-term access to quality economic and social data, supporting and promoting their use, value and impact. To achieve this for current and future data depositors and other beneficiaries, the Archive offers long-term preservation services for the digital objects in its collection.

The Archive is the lead partner in, and provides Trustworthy Digital Repository (TDR) functions to, the UK Data Service (the Service)[1]. The Archive supports the Service's mission "to support high quality social and economic research, teaching and learning through assuring long-term access to quality economic and social data, supporting and promoting their use, value and impact". It also helps it meet the Service's vision "to continue to be a critical part of the UK's research infrastructure where the exceptional economic and social data we make available are central to the achievement of excellence in research, teaching and in the realisation of public benefit."[2]

The Archive takes as its primary user community social science data users within the higher education (HE) and further education (FE) sector(s) in the UK, though best efforts are made to meet the needs of all users.[3] To ensure the continued accessibility and use of these resources the Archive follows a policy of active preservation with the aim of ensuring the authenticity, reliability and logical integrity of all resources entrusted to it for long-term digital preservation while providing up-to-date formats suitable for research, teaching or learning, in perpetuity.

This policy codifies long-standing good archival practice at the Archive. The Archive has been

---

[1] https://ukdataservice.ac.uk/
[2] https://ukdataservice.ac.uk/about/strategic-focus/vision-and-mission/
[3] The vast majority of the UK Data Archive's current users are staff and students at UK HE/FE institutions, though central and local government staff, commercial users and personal researchers also figure.

the primary repository for social science research data in the UK since 1967. As a 'national data collection service' the Archive, originally called Data Bank, was created by the 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. At the time of writing the UKRI through the ESRC is the primary funder of the UK Data Archive through its resourcing of the UK Data Service. In 2012, the UK Data Service integrated the Economic and Social Data Service, the Secure Data Service (now UK Data Service SecureLab), ESRC Data Store (now UK Data Service ReShare) and parts of the Census Programme. The responsibilities for long-term preservation of assets may change during the lifetime of this service.

The ESRC Research Data Policy[4] emphasises the importance and requirement of depositing data generated from ESRC funded research with the Archive; The National Archives' Records Collection Policy[5] regulates the disposition of datasets created by government departments.

The formulation and annual review of a preservation policy for the UK Data Archive are essential steps in fulfilling its strategic aims and responsibilities: it gives strategic direction both to initiate any measures that are necessary for the protection of its collections, and to meet, or extend, nationally and internationally agreed standards for the preservation of digital materials including the CoreTrustSeal[6]. A preservation policy helps the Archive meet legislative and accountability requirements and its user communities' expectations. The Archive ensures that it is at the leading edge of technical advances by taking a strategic approach to long-term digital preservation, which includes the monitoring of hardware and software developments, and obsolescence and migrating its collections accordingly.

The Archive also aims to continually improve all aspects of the preservation-related workflow by embedding an awareness of quality in all processes.

3.1 Objectives

The primary objective of all archives is to select, preserve and make available for use documents or information that have permanent or continuing value. The Archive is no exception; however, it differs from some archives in that it provides archival facilities only for digital material, whether born digital or digitised. Consequently, specific digital preservation activities are essential in ensuring the continued use of material. This policy outlines the key actions and rationale behind the actions necessary to ensure that the data collections of the Archive are permanently accessible in a form that is fit for purpose for all end users of the service provided by the Archive.

---

[4] ESRC Research Data Policy
[5] Records Collection Policy (nationalarchives.gov.uk)
[6] UK-Data-Archive.pdf (coretrustseal.org)

In the language of *digital preservation*, the Archive is a digital repository. The Archive assumes responsibility for the long-term preservation and accessibility of digital objects. For all practical purposes 'long-term' means beyond the next round of technical change (as monitored by the Archive's technology watch functions).

An archival organisation is also responsible, when selecting which material to archive, for ensuring the reliability and logical integrity of the *data collection*. For paper-based documents, the reliability and integrity of the various elements of a data collection are usually assured by a simple audit trail; for digital material these requirements cannot be guaranteed in such a straightforward fashion. This is because some of the *significant properties* of a data collection may have to be altered in order to ensure a level of software independence and guarantee long-term *usability*.

Any strategy for the long-term preservation of any electronic information must address the issue of software dependence. For most electronic information it is generally possible to eliminate software dependence by producing a more abstracted version of the information independent of the software 'wrapper', but the end products of these transformations are not always authentic versions of the original. In these cases the authenticity needs to be re-established through the documentation of the actions taken and validation that the substantive content has not been altered.

As an example, a relational database could be migrated to a number of flat files. However, for each table the relational links need to be expressed in *documentation* in terms of the type of the link and the manner in which the keys can be identified.

Thus, the primary goal of the Archive's preservation policy is to ensure the long-term accessibility of electronic information while ensuring the highest level of authenticity of any formats disseminated. In effect this means that all the essential qualities of the electronic information upon which their authenticity depends are preserved.

Most of the data currently archived at the Archive does not rely on presentation characteristics, such as font or colour, to ensure they are 'understandable'.

This policy is also designed to promote preservation as an integral part of the management of the Archive's collections and to ensure the best use of resources by providing a framework for managing the curation and preservation procedures. The specific aims of the preservation policy are to:

- Provide authentic, reliable instances of data collections to the designated user community.

- Be a 'trusted digital repository' within the generally accepted scope of the term[7].

- Maintain the integrity and quality of the data collections.

- Ensure that digital resources are managed throughout their lifecycle in the medium that is most appropriate for the task they perform.

- Ensure that all data collections are protected.

- Ensure that the relevant level of information security is applied to each data collection.

- Instil good practice in active preservation management.

- Improve the speed and efficiency with which information is preserved and retrieved.

- Develop and maintain systems of cost-efficient storage, with appropriate location and with regular review.

- Optimise the use of the Archive's space for storage purposes.

4. Requirements

As the main functions of the UK Data Archive are to acquire, develop and manage data and the related digital resources that are of value to social scientists, and to promote and disseminate these resources as widely and effectively as possible, the Archive has developed a series of requirements which it strives to ensure are followed as closely as possible:

- The data collections it accessions are accompanied by adequate *documentation* to enable their use for analytical and research purposes.

- The data collections are checked and validated according to strict data and documentation ingest procedures.

- The data collections are professionally catalogued according to appropriate metadata standards.

- The data collections are indexed with keyword terms using an appropriate thesaurus.

- The datasets, documentation, metadata and other representation information that comprise each data collection are kept in conditions suitable for long-term archival storage.

- The authenticity, integrity and reliability of data collections preserved for future use are

---

[7] A definition of a TDR is available in Research Libraries Group / OCLC, Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report (2002): http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf.

retained.

- The basic preservation actions undertaken by the Archive are uniform regardless of the perceived value of any data collection. (Events within the preservation process may differ from data collection to data collection but the actions specific to data preservation must not.).

As a core activity in the 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

- the UK Data Service *Collections Development Policy* and referenced requirements

- the UK Data Archive's *Information Security Policy* and referenced requirements

- the needs of the users of the Archive

- archival theory and practice

- the place of the Archive within local, national and international frameworks.

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.

4.1 Legal and regulatory framework

The legal and regulatory frameworks for the management of the data collections accessioned by the UK Data Archive are complex. The University of Essex is the legal entity under which the Archive functions. The Archive is a department within the University of Essex and has no legal status.

The relationship between the *depositor* of a data collection and the Archive is based on:

- A legally-binding deposit agreement and licence 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.[8]

- An assertion of copyright and intellectual property rights to ensure that the data creator/depositor has cleared all necessary permissions.

---

[8] Since 2008, the UK Data Archive promotes 'open data' with the use of a de facto access condition of open to all comers, for all uses.

- Where necessary, negotiations for licence agreements with third parties to enable the Archive explicitly to distribute the material to particular user communities.

The Archive will not ingest materials that have unclear ownership or unresolved rights issues.

In preserving its collections, the Archive follows:

- Copyright, Design and Patents Act, 1988 and amendments to this Act

- Data Protection Act, 2018

- The UK-GDPR (General Data Protection Regulation)

- Statistics and Registration Services Act, 2007

- Freedom of Information Act, 2000

- EU Copyright Directive, 2001

- Environmental Information Regulations, 2004

- English or UK law for commercial agreements and contract law

- Current best practice.

In terms of national standards for the management of information security, the Archive is certified to *BS ISO/IEC 27001: 2013 - Information technology -- Security techniques -- Information security management systems -- Requirements*.[9]

Depositors are responsible for reviewing any ethical issues relating to data collections they wish to deposit including those surrounding the potential for risk of harm to any participants in making data available to third parties. The ESRC's *Research Ethics Framework (REF)* provides guidance to both the Archive and to its researchers.

5. Roles and responsibilities

Since its inception in 1967, preservation has been a fundamental function of the UK Data Archive. It has always been recognised that the preservation function is made up of a series of interrelated activities that have traditionally been carried out by different sections of the Archive. Digital preservation requirements are relevant to all activities within the Archive's workflow, hence the centrality of this policy to the undertakings of the UK Data Archive.

The Digital Preservation Systems and Security (DPSS) section, part of Technical Services, has responsibility for preserving data and metadata in all forms to ensure they remain usable over time, including monitoring technological changes that will affect preservation and

---

[9] The UK Data Archive has been certified to BS ISO/IEC 27001 since June 2010.

migration decisions. It is also responsible for the creation of *preservation metadata*.

The Pre-Ingest team is responsible for identifying, selecting and acquiring data and other resources in line with the UK Data Service *Collections Development Policy*. It is also responsible for the initial validation of the data.

The Ingest team is responsible for the validation of the data and *documentation*, including checking of consistency, accuracy and suitability for preservation and secondary analysis, and producing and enriching metadata about the data resources in order to facilitate resource discovery. This section also converts the archival copy of data to current dissemination formats to meet user needs, and provides other value-added user support.

Archive staff develop and maintain the Archive's thesaurus products and interpret national and international standards for local implementation in *resource discovery metadata*, and establish, maintain and develop cataloguing and indexing metadata standards, metadata offerings and resource discovery systems across all Archive data services.

The Application Development and Maintenance (ADM) section, part of Technical Services, provides access to data collections through the internet and maintains the supporting application infrastructure for metadata.

All UK Data Archive staff have a role to play in the implementation of the preservation policy. Digital Preservation, Systems and Security staff take an active role in assisting and encouraging the implementation of the preservation policy Archive-wide.

6. Model

The Open Archival Information System (OAIS) reference model is an international standard which proposes common terms and concepts and a framework for entities and relationships between entities in digital preservation environments. The Archive recognises the benefits of the OAIS as a conceptual framework for defining and aligning archival activities and not a concrete implementation plan. The Archive follows the broad guidance given in the OAIS reference model.

When, in 2005, the Archive assessed its conformance with the OAIS model, the main divergence between model and practice was the strict separation of Archival Information Packages (AIPs) from Dissemination Information Packages (DIPs). There were also a number of activities within the data management function relating to monitoring and management that were not appropriate to the stated objectives of the Archive. This holds true today.

The UK Data Archive believes that, as digital preservation activities are an integral part of the mission of the organisation, all areas of work are influenced by the preservation policy. Periodically reviewed guidance, reference information and procedures document the implementation of this preservation policy across the key areas of Archive activity.

6.1 Pre-ingest function

The pre-ingest function is not explicitly specified in the OAIS model. However, experience has shown that inclusion of this function within the preservation model has considerable benefits. Most notably it ensures quality, comprehensibility and accessibility of all information packages by enforcing quality assurance and minimum standards at the 'Producer-Archive interface'. Successful application of the pre-ingest process also helps ensure that data collections are submitted at a standard which requires a lower level of processing at the ingest stage and also helps ensure that increased levels of usability are enabled through the provision of adequate documentation. This function also ensures that the Archive involves the depositors in any decision-making process about what information properties of a digital object shall be retained in the preservation process.

Proactive pre-ingest activity also addresses various issues which might impact on preservation activities (relating to consent, confidentiality, ethics, legal issues and data formats) to be considered and addressed before data deposit starts.

6.2 Ingest function

Ingest is the first functional component of the OAIS reference model. It includes the receipt of information from a *producer* and the validation that the information supplied is uncorrupted and complete. This process also identifies the specific properties of the information to be preserved; it verifies that the information is what it purports to be. The version supplied by the depositor is known within the Archive as the 'original' version. This original version is retained for preservation in its original format and stored in the appropriate directory within the preservation system. This original version has a close correspondence to the OAIS Submission Information Package (SIP).

The ingest function also transforms all elements of the deposited files into a valid preservation format for the specified data type. Files for preservation are copied to a different machine, as the ingest and preservation directory structures are created.

The Archive currently believes that the construction of a DIP during the ingest process (rather than automatically from an AIP on demand) has considerable benefits for the preservation process. This allows the Archive to deprecate the use of an existing DIP to create a new DIP to avoid the loss of integrity in the underlying data by transformation through different software packages.

The ingest function also includes the creation of descriptive metadata for a variety of purposes and the production of DIPs in multiple formats to maximise usability.

As well as an unbroken audit trail of actions to ensure the authenticity and integrity of any data collection, the ingest process includes an element of depositor accountability whereby depositors are informed of actions undertaken within the Archive before the data collection is released to a wider user community.

The UK Data Archive will not preserve depositor-submitted physical media or non-digital documentation in their original format. These will either be returned or destroyed securely.

6.3 Archival storage function

The second functional component of OAIS is archival storage. Archival storage manages the digital objects which are entrusted to the Archive. In essence, the purpose of archival storage is to ensure that what is passed to it from the ingest process remains identical and accessible. In the Archive this function receives AIPs and DIPs from the ingest function and adds them to the permanent storage facility, oversees the management of this storage, including media refreshment and monitoring. This function is also responsible for ensuring that AIPs can be retrieved.

The Archive does not follow the OAIS model fully. In the model there is a 'Provide Data Function' whereby the Access function of an Archive can request AIP transfer from the storage area. In the reference model this process ensures that end users receive an authentic version of the data collection. However, due to a combination of factors relating to information security, access conditions and usability, the Archive has elected to prepare DIPs as part of the ingest process. These DIPs are accessible to users via a 'Provide Data Function' and identical copies are also stored alongside the AIPs in the preservation system. Thus for any data collection there are always at least three information packages residing on each of the different preservation systems: the original SIP, the ingested AIP and multiple DIPs. When a new version of a DIP is created it must be created from the AIP (or possibly the SIP if it is understood to be a de facto subset of the AIP) but not from an earlier recension of a DIP.

6.3.1 Physical data preservation and storage

In order to best safeguard long-term preservation, the Archive follows a policy of multiple copy resilience. Five versions of the complete preservation system are held. A main near-line copy and a shadow copy on two separate preservation servers, only accessible using a dedicated preservation account, the access online copy (on the mirror preservation server), and copies are generated for user access and dissemination. There is also a near-site online copy on a server located in another location within the University of Essex, and an off-site online copy. An encrypted hard disc-based offline copy is also maintained.

The Archive follows best practice in the storage and housing of magnetic and optical media. In particular, for environmental conditions for storage media (BS ISO 18925:2002) and for the storage of archival materials (BS 5454).

6.3.2 Media monitoring and refreshing strategy

The UK Data Archive operates a media monitoring procedure. This allows it to check for potential future problems of wear and tear on media and act before the problems become severe. Bad blocks on discs are automatically identified, administrators are notified and the disc replaced.

If any media have either recoverable or non-recoverable errors then they are regenerated from the on-site mirror preservation server.

6.3.3 Compression

In order to reduce the risk of damage to data, the UK Data Archive uses industry standard lossless compression tools, which are outlined in the Archive's Information Security Policy. Compression software is used under tightly controlled conditions on files stored on offline hard drives and as a concatenation tool for coherent collections of multiple TIFF files. Sufficient redundancy is used for long-term storage to warrant use of these tools.

Compression software technology is monitored continuously to ensure timely updates to enable successful migration.

6.4 Data management function

Data Management is the third major function of the OAIS reference model. It works in conjunction with the Archival Storage function. It maintains databases of descriptive metadata; supports external finding aids; and manages administrative metadata which support internal operations, including change control.

6.4.1 Version control/change procedures

Ensuring that any alteration to the preserved version of any part of a data collection is accurately documented is integral to the authenticity of any data collection. The Archive distinguishes between two forms of alteration post ingest:

New *version*: when there is a change to the *preserved* metadata

New *edition*: when there is change to data or documentation.

The word *preserved* is highlighted because not all metadata are proposed to be preserved under this policy. All 'finding-aids' (which the Archive defines explicitly as: abstract, subject category and keywords) are understood in this policy to be metadata that are not necessarily preserved at present.

When there is a new version of a data collection, the relevant descriptive and structural metadata must be reviewed and revised as necessary and the old file retained.

When there is a new edition of a data collection, all descriptive and structural metadata must be recreated, and the old file and the previous AIP and DIPs retained within the preservation system and identified as 'not for issue'.

6.4.2 Data collection withdrawal

The UK Data Archive operates a multifaceted policy towards data collection withdrawal. The Archive distinguishes between 'soft deletion' whereby certain references to the withdrawn content are deleted, but not the content itself, and 'hard deletion' whereby the content and all references to it are deleted. In the case of soft deletion the data collection is only accessible to Digital Preservation Systems and Security and Ingest team members.

The Archive chooses soft deletion as the default method of withdrawal since it is too expensive to remove data collections, and their physical removal would present unacceptable risks to other parts of the collection. The Archive has undertaken hard deletion of collections because they are archived, preserved and disseminated elsewhere, or where instructed by data owners to meet legal requirements

In cases of the withdrawal of a data collection, the administrative metadata are updated, and the external view of the metadata record is updated to reflect the change of status of the collection (with information about why the collection had been withdrawn, the dates of its availability, and, where appropriate, the reasons for withdrawal).

6.5 Access function

The sixth and final function of OAIS relates to access. In this function end users interact with the Archive to find, request and receive data collections. These three separate, but interrelated operations are:

- finding data collections
- requesting data collections
- delivering data collections, either directly or remotely.

The access function must also implement the security relating to access; all access management failures are logged and processes reviewed periodically.

6.6 Administration function

In the OAIS model the 'administration function' manages the day-to-day operations of the repository. In the UK Data Archive the roles of this function are spread across different internal sections.

7. Preservation planning strategy

Under this section the fourth and fifth OAIS functions are discussed separately.

7.1 Preservation strategy overview

The UK Data Archive has chosen to implement a preservation strategy based upon open and available file formats, 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 alongside the requirements of levels of authenticity and legal admissibility. Hence, different ingest processes may be required for material with different levels of quality and significance. Data collections are assigned an ingest activity level as outlined in the Archive's Collections *Development Selection* and *Appraisal Criteria* document. Specifications for ingest standards including validation and integrity checks are fully documented.

The Archive's preservation strategy is predicated on two basic principles: first, that digital storage media are inherently untrustworthy unless stored appropriately; second, that all file formats and physical storage media will ultimately become obsolete.

Therefore, the environmental parameters which control the storage media are tightly controlled to reduce the vulnerability of these media. Additionally, the strategy to reduce the risk of obsolescence is based on storing multiple copies on different storage media. These are reviewed regularly, and data are copied onto new media when appropriate.

A similar strategy is employed to deal with the obsolescence of file formats. Appropriate information-rich preservation formats have been identified and are used in conjunction with formal documentation procedures. These formats are chosen with specific reference to the 'data types' under consideration. The Archive follows international best practice in its choice of preservation formats and data migration procedures.

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 strategies of the Archive aim to maintain a flexible preservation system that evolves to meet the demands of changing technology and new and increasing user expectations.

Every *data collection* within the preservation system follows a consistent directory structure for storage, and this is enforced by automated checks. This has many benefits, such as the ability to find 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.

7.2 Integrity measures

The UK Data Archive takes its role as custodian of data collections seriously. To this end the complete chain of custody of all data collections is documented through metadata. All actions are explicit, complete, correct and current. However, only the 'original' version can be said to be an integral copy of the version deposited with the Archive. The preservation and dissemination versions are considered to be authentic and there is an audit trail of all alterations in the preservation and dissemination versions which relates back to the original deposited version.

7.3 Monitoring, review and feedback

The preservation policy of the Archive is monitored and reviewed in the light of changing technologies on an annual basis to ensure timely updates. The Repository & Preservation Manager starts the review process in line with the Archive's controlled document management procedures.

This document is communicated to all Archive staff and will be published on the Archive's website, and shared with other archives and bodies that are involved in digital curation activities.

Queries concerning the preservation of the digital collections in the Archive, or digital preservation in general should be directed to the Archive's Digital Preservation, Systems and Security section: (preservation@essex.ac.uk).

8. IT infrastructure

The preservation of the Archive's collections relies on an IT infrastructure that is fit for purpose and is continually monitored and periodically reviewed to ensure timely upgrades in both hardware and software.

In order to ensure resilience and provide an adequate level of redundancy, the preservation system consists of on-site, near-site and off-site storage. For the same reasons, mirror versions of on-site systems are provided.

Adequate storage capacity for all holdings is maintained.

The Archive provides necessary secure networking and communications equipment, providing adequate connectivity, the ability to restrict access to valid Mac addresses and a facility to segment the network for switched separated firewall connectivity.

All servers in the Archive are protected by power surge protection systems.

Disaster recovery procedures are in place.

9. Security

The UK Data Archive is committed to taking all necessary precautions to ensure the physical safety and security of all data collections that it preserves:

- fire prevention and protection system
- physical intruder prevention and detection systems
- environmental control systems.

The server rooms are equipped with multiple key entries and a security-protected swipe-card system linked to an on-site alarm system and to the University Security Office. The swipe-card system is maintained by archive staff, and access is restricted to two key members of staff.

The server rooms are located outside of the secure working area of the Archive.

The Social Sciences Research Centre (SSRC) building in which the Archive is housed, is locked between 19.30–7.30 and all weekend, and is regularly patrolled by University of Essex security staff.

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.

The Archive's suite of information and premises security are documented in our Statement of Applicability for ISO/IEC 27001. The Archive was first recommended for certification to this standard in June 2010 and maintains that certification. These are detailed in the Archive's *Information Security Policy, Information Security Management Policy* and *Premises Security Procedures*.

10. Co-operation

The Archive 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.

11. Funding and resource planning

The UK Data Service is partially dependent on funding from the Economic and Social Research Council (and others) to ensure the longevity of the resources which it holds. If there were an indication that this funding was to cease, succession planning activities (as required by the Succession Plan described in *Business Continuity Incident Management Procedures*) would be completed and made operational.

The UK Data Service is committed to supporting continued funding for all 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 recruitment, induction, and ongoing staff training.

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

The preservation of data and documentation to ensure they remain usable over time is a core activity of the Archive. The Archive, therefore, makes every effort to remain up to date with any relevant technological advances to ensure continued access to its collections.

www.data-archive.ac.uk

dainfo@essex.ac.uk

+44 (0)1206 872001