Appendix 2

A DIGITAL PRESERVATION POLICY FOR WALES

TECHNICAL APPENDIX

This Appendix examines and expands upon the principles and technical challenges referred to in the main Digital Preservation Policy for Wales and outlines the necessary elements of a functional digital preservation model.

1 PRINCIPLES AND CHALLENGES OF DIGITAL PRESERVATION

- Digital objects are *encoded*, requiring technological mediation to render their content accessible. This depends on a complex set of interconnected technologies comprising all the elements required to correctly represent the object. These include the formats in which information is encoded, software required to interpret these formats, operating systems and hardware required to execute that software, and physical media on which that information is stored. The absence or failure of any part of this network may render the object inaccessible.

- Information technology continues to rapidly advance. As new products appear, older products cease to be supported. The currency of any given technology is typically very short, perhaps five to ten years. A principal challenge of digital preservation therefore lies in maintaining the means of access to digital objects in the face of rapid technological obsolescence. Digital storage media are susceptible to alteration, damage and decay over short timescales. The resultant potential for information loss must therefore also be mitigated.

- The authoritative nature of a record, from which its continued value derives, must also be maintained. An authoritative record may be understood in the context of four characteristics as defined by the international standard *Information and Documentation - Records Management* (ISO 15489)[1]:

- **Authenticity**: The assurance that the record is what it purports to be.

- **Reliability**: The record is a full and accurate representation of the business activity to which it attests. This requires the establishment of trust in the record keeping and archival processes used to manage the record throughout its lifecycle, and the continued ability to place the record within its operational context. This may be ensured through the operation of transparent and fully documented preservation strategies, and the provision of the metadata that is required to describe the content, context and provenance of the record.

---

[1] International Organization for Standardization, 2016. *ISO 15489-1:2001, Information and Documentation - Records Management - Part 1: General*, Available at: http://www.wgarm.net/ccarm/docs- repository/doc/doc402817.PDF.

- **Integrity**: The record is maintained to ensure it is complete, and protected against unauthorised or accidental alteration. This may be ensured through bitstream preservation; the provision of metadata to describe all authorised actions undertaken in the course of content and bitstream preservation; and robust access protocols.
- **Usability**: The record may be continuously accessed by users, across changing technical environments. It must be locatable and retrievable, capable of being rendered in a current technical environment, and supportive of interpretation by users. This may be ensured via content preservation methods and the provision of metadata sufficient to allow the record to be located, retrieved and interpreted.
- The authoritative nature of a record may be lost if, to eliminate software dependence, the structure and context within which the information resides is sacrificed. Transforming file formats and/or transferring information between storage media alone, as opposed to also preserving the structure of the actual records containing the information, results in unreliable end products. Documentation of actions taken, the reasons for taking these and validation that the substantive content has not been altered are required to preserve authority.
- Authority may also be lost if there is uncontrolled copying of the authentic original 'master' record with identification of the latter and its information content thus becoming blurred. Version control of digital records should be adopted and maintained in order to avoid this.

2 DIGITAL PRESERVATION FUNCTIONAL MODEL

- This section outlines the functions that a repository should utilise to successfully undertake digital preservation in terms of the *Open Archival Information System (OAIS) Model* (ISO 14721)[2].
- **The Open Archival Information System (OAIS) Model**

  Using standards can provide unambiguous benchmarks for defining digital preservation, requirements and measuring outcomes. They can support interoperability both between systems and across time. Of particular relevance is the Open Archival Information Systems (OAIS) Reference Model (ISO 14721), an international standard which defines a high-level functional model for a digital repository, which proposes common terms and conceptss wid ely used across the digital preservation community. OAIS is a conceptual framework, not a concrete implementation plan. This policy follows the broad guidance given in the functional model of the OAIS reference model.

2 Consultative Committee for Space Data Systems, 2012. *Reference model for an open archival information system (OAIS)*, Available at: http://public.ccsds.org/publications/archive/650x0m2.pdf.

The overall strategic priority is that content, in the form of digital records, be preserved, reliable and accessible over time for a pre-defined community of users: (*The Designated Community*). This is supported by the management of a number of functions that together make up the digital preservation process, all of which should be present in an archival institution in order to successfully undertake digital preservation activities:

- **Pre-Ingest Function**

Though not explicitly specified in the OAIS Reference Model, a pre-ingest function has been demonstrated as very beneficial to the remainder of the digital preservation process and is standardised as ISO 20652: *Producer Archive Interface – Methodology Abstract Standard (PAIMAS)*[3]. It should aim to ensure the following:

- Quality, comprehensibility and accessibility of information packages via quality assurance and enforcement of minimum standards at the point of the 'Producer-Archive Interface'.
- Issues that could affect preservation activities (consent, confidentiality, ethics, legal issues and data formats) are considered and addressed before deposit.
- Planning, rights and access are secured.
- The Archive institution involves the depositors in any decision-making process about which information properties of a digital object shall be retained.
- Records are submitted at a standard which requires a lower level of processing at the ingest stage.
- Metadata is created to enable identification and discovery.
- Checksums are generated so files are checked upon ingest.
- Potentially greater levels of usability are achieved via provision of adequate documentation.
- Financial costs of the actual ingest process are reduced where possible as a result of the above.

- **Ingest Function**

Ingest comprises the actual receipt of information in the form of records from a producer, and the validation that information supplied is uncorrupted and complete. It identifies the specific properties of the information to be preserved and authenticates that the information is what it purports to be.

The 'original' version of a record deposited, retained for preservation in its original format, stored in the appropriate directory on the preservation system and, together with accompanying files and metadata needed to access and reconstruct the information in an authentic manner, is referred to as the ***Submission Information Package (SIP)*** in OAIS terminology.

3 Consultative Committee for Space Data Systems, 2004. *Producer-archive interface methodology*, Available at: http://public.ccsds.org/publications/archive/651x0m1.pdf.

- The Ingest function receives information from producers and packages it for storage. It accepts a **SIP**, verifies it, creates an **AIP** (an ***Archival Information Package****)* from the **SIP**, and transfers the newly created **AIP** to archival storage.
- The ingest function also may include the creation of metadata for a variety of purposes including to demonstrate an unbroken audit trail of actions to ensure the authenticity and integrity of records ingested.
- The ingest process should also include an element of depositor accountability whereby the latter are informed of actions undertaken within an archival institution before records are released to a wider user community.
- Depositor-submitted media or non-digital documentation in their original format may be returned or destroyed securely after completion of ingest, rather than their being retained.

- **Storage Function**

  This is the second functional component of OAIS. It manages the digital objects which are entrusted to the Archive, ensuring that what is passed to it from the ingest process remains accessible.

  - Storage should ensure confidentiality, integrity and availability of digital objects and if possible certified against the relevant parts of the ISO 27000 family of standards (*Information Security Management*)
  - The storage function creates **AIPs** or receives them from the ingest function and assigns them to long term storage in the appropriate permanent storage facility.
  - **AIPs** are similar in concept to **SIPs** (see above) but with appropriate alterations made so that the package is fit for permanent preservation and storage; for example, by conversion of elements of the package to formats more suited to long term preservation.
  - The storage function oversees all aspects of storage management, including maintenance of **AIPs**, media refreshment, monitoring and error checking to ensure bit-rate loss and degradation do not occur, including migration where necessary.
  - Requested AIPs are retrieved as needed by providing them to the Access function.
  - Archive institutions may follow a policy of multiple copy resilience as part of the storage function. Different versions of the complete system may be held on servers distributed across a number of locations for security via multiple redundancies.
  - Where original storage formats such as magnetic and optical media are retained in storage rather than being disposed of, best practice should be adhered to in terms of environmental conditions for storage media (BS ISO 18925:2013) and archival materials (BS 4971).

- **Data Management Function**

  This is the third major function of the OAIS model; it operates in conjunction with the Storage function. The Data Management Function coordinates the descriptive information of the AIPs and the system information that supports the archive; maintains the database that contains the archive's information by executing query requests and generating results;

generates reports in support of other functions; manages administrative metadata (which support internal operations including change control); and supports external finding aids.

- Any alterations to the preserved version of any part of a collection should be accurately documented; this is crucial in retaining the authenticity of any digital records.
- Where records and data are to be withdrawn for any reason, a distinction (recognised by The National Archives and the UK Data Archive) may be made between *'soft deletion'* (where references to withdrawn content are deleted, but not the content itself) and *'hard deletion'* (where the content **and** all references to it are deleted).
- Soft deletion avoids costs associated with wholesale removal of data collections, and avoids any risks which their physical removal might present to other parts of the collection.
- Hard deletion might be considered in cases where collections are archived, preserved and disseminated elsewhere.
- Where a collection is withdrawn, administrative metadata and any external view of the catalogue record should be updated to reflect the change of status of the collection, including where appropriate information about why the collection was withdrawn and dates of its availability.

- **Administration Function**

  This function manages the daily operations of the repository by:

  - Obtaining submission agreements from depositors.
  - Performing system engineering.
  - Auditing **SIPs** to ensure compliance with submission agreements.
  - Developing and ensuring adherence to policies and standards.
  - Dealing with customer service needs.
  - Managing legal requirements and rights management relating to Digital Records including Freedom of Information, Data Protection and other imposed access and copyright restrictions.
  - Acting as interface between Management and the Designated Community in the OAIS environment.

- **Preservation Function**

  This function supports all tasks in order to keep digital records permanently accessible and understandable even if the original computing system becomes obsolete, via:

  - The development of detailed preservation/migration plans.
  - Maintaining a 'technology watch' to monitor software, hardware, operating systems, determining which formats are at risk of obsolescence and which have a longer projected life, etc.
  - Evaluation and risk analysis of Digital Records.
  - Recommendations regarding updates and migration, based on the above points.

- Persistent maintenance of metadata and its relationship with the digital objects to which it relates including (as part of bitstream preservation) any physical or logical change to a digital object being logged and recorded in the associated metadata to provide an audit trail. All changes to metadata themselves should also be audited.

- **Access Function**

  The sixth and final function of OAIS relates to access, whereby users interact with the archival institution to find, request and receive digital records. The access function must also implement security relating to access, monitor access management failures and review access processes.

- The access function utilises **DIPs** (***Dissemination Information Packages***) which are again related to the **SIPs** and **AIPs** that are described above. **DIPs** have appropriate alterations made so the package is fit for dissemination to a defined audience; for example, by conversion of elements of the package to formats more suited to this such as image files that require less memory storage and/or are watermarked to preserve copyright; or provision metadata with technical or administrative fields removed.

- **Current System Selection**

- The ARCW group has (as of 2017), following a rigorous assessment procedure, selected **Archivematica**, a systems infrastructure which supports the policy and ensures the technical elements of the digital preservation function are undertaken.

- Archivematica is an integrated suite of open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO- OAIS functional model. It uses METS, PREMIS, Dublin Core, Droid, JHOVE and other recognized standards to generate trustworthy, authentic, reliable records, associated metadata and enable access to these records and their metadata.

- The linkage of Archivematica with another open source solution, Fedora, which manages the preservation function, provides the attributes necessary for the management and preservation of the digital content within systems architecture.