Digital preservation for local authorities

"The 100 year use case"

25 June 2018

1. Introduction

1.1 This is an evolving working document for Archives First: project two. The document will be developed as experience/insights are gained.

1.2 The purpose of the document is to record and communicate the 100 year use case where digital information needs to be preserved for 100 years. This is a requirement of the Statutory Guidance on Adoption 2013 in respect of a so-called "adoption record".

1.3 These records are currently subject to the GDPR as enacted by the Data Protection Act 2018 (DPA).

1.4 Not all local authority digital preservation use cases are affected by the DPA and not all information has a statutory retention as long as 100 years.

However, the 100 year use case is applicable in all instances where the requirement is to retain information of enduring value in perpetuity.

1.5 An OAIS approach is assumed. In particular the AIP container, by definition, includes all the information that is being preserved together with sufficient material for a user to access the information.

1.6 A feature of the use case is that information access is restricted*;* general access is not permitted for 100 years.

1.7 In order to further generalise the applicability of the use case AIP creation is an automated process. Manual processes would be both error prone and unable to scale to the volumes required.

1.8 Some additional technical comment is given in the Appendix in order to provide clarification.

2. Packaging (AIP creation)

2.1 Information in the AIP is born-digital together with digital attachments. The information will have been created and managed by a business transaction processing system over several years having been migrated from legacy systems.

2.2 The information asset owner is the appointed business manager.

2.3 The Data Controller responsible for compliance is the local authority.

2.4 A trigger event will cause information in the transaction processing system to be collated and exported as a structured collection of simple document and image format computer files.

2.5 Package metadata that is metadata describing the package rather than individual computer file metadata will be compiled and recorded as a computer file using an enduring format and schema (see the Appendix).

2.6 The AIP is created by including the structured collection of simple document and image files and the package metadata file in a single container file which has a UUID name. The AIP creation process includes calculating and recording package fixity values (see the Appendix).

2.7 Packages are created automatically.

3. Storage

3.1 Depositing the AIP in a trusted dark store.

3.1.1 The AIP and its package fixity values are created locally by the depositor.

3.1.2 A copy of the AIP is deposited in a reliable secure long-term digital storage system. It is assumed that this storage system is remote.

3.1.3 The AIP is encrypted whilst in transit and a suitable transmission security protocol is employed. The AIP is decrypted following receipt and is stored as plain text.

3.1.4 Several fixity values for the (plain text) AIP are calculated by the storage system and reported to the depositor for comparison with the locally created fixity values. The deposit is successful only if the fixity values correspond.

3.1.5 There is no local copy of the AIP.

3.2 Maintaining trust

3.2.1 Maintaining trust is an active management process. The trusted store regularly demonstrates the continued authenticity of the AIPs in its custody by recalculating and reporting fixity values.

3.2.2 No AIPs or any package content files are deleted (silently or otherwise).

3.2.3 The storage system conforms to all relevant reliability and security standards.

3.2.4 The storage system reports the results of DR AIP restore testing.

3.2.5 An AIP escrow arrangement is in place. The escrow copies are stored securely by a third-party. AIPs in transit between the storage system and the escrow store are encrypted; escrow copies of the AIPs are plain text. Escrow invocation is tested. Escrow invocation does not require any proprietary software.

3.3 Exit

3.3.1 The termination arrangement provides for an orderly transfer of AIPs to another trusted dark store.

4. Discovery

4.1 AIP discovery is facilitated by a locally maintained searchable catalogue that holds a copy of the AIP package metadata together with the AIP UUID.

5. Presentation (DIP creation)

5.1 A discovery system is maintained locally which provides the UUID name of a required AIP.

5.2 There is a secure user authentication procedure in place which the trusted dark store uses to verify the requester's credentials.

5.3 In response to a valid request from a verified user, the storage system will provide a copy of the AIP. The AIP is encrypted in transit.

5.4 Following decryption, the requester calculates several fixity values for the retrieved AIP which are compared with the locally stored values. The retrieval is successful only if the fixity values correspond.

5.5 Creating the DIP, that is managing the transformation AIP to DIP, is a mediated process (see the Appendix).

5.6 Document and image AIP content file formats are transformed automatically and without being executed.

5.7 The DIP is made available to a qualified end user, that is an end user to whom some or all of the information can be disclosed.

5.8 The end user is advised to employ anti-virus software.

6. Managing risks

This section is much influenced by the Planning Tool for Trusted Electronic Repositories (DigitalPreservationEurope, 2008), (PLATTER).

6.1 Financial

6.1.1 Both the depositor and the trusted store are exposed to existential financial (including organisational) risk.

6.1.2 Suitable escrow arrangements mitigate the effect of failure by the trusted store.

6.1.3 Failure by the depositor is not managed. (Management options could include insuring the credit risk, risk pooling, or last resort arrangements.)

6.2 Key personnel

6.2.1 Both the depositor and the trusted store are vulnerable to the loss of key personnel.

6.2.2 The risk is mitigated by

i. avoiding there being a single key individual
ii. relying only on industry standard practice
iii. maintaining relevant skill-sets
iv. maintaining full documentation

6.2.3 The depositor’s authorised users are key personnel.

6.2.4 All staff in the depositor's chain of authority are key personnel.

6.3 Preservation plan

6.3.1 The depositor is vulnerable to the future obsolescence of file formats used in the AIP content which prevent access.

6.3.2 The risk is mitigated by managing the range of file formats used. If a format is deemed to be at risk because, for example, it is proprietary and no open reader exists, then a non-proprietary version is included within the AIP.

6.3.3 Demonstrating authenticity requires local access to a register of AIP fixity values. This is supported by a local operational system which is maintained day to day in the usual way. Fixity value data is retained using a non-proprietary format.

6.3.4 Both the depositor and the trusted store are vulnerable to the adverse effects of technological developments (both hardware and software).

6.3.5 This risk is mitigated by the identification of critical technology and an appropriate "technology watch".

6.4 Succession plan

6.4.1 The depositor and the trusted store are exposed to succession failure, both technological and human.

6.4.2 The risk is mitigated by relevant transition and handover procedures including testing.

6.4.3 All key personnel and technology are included in the succession plan.

6.5 Discovery system

6.5.1 The discovery system is provided by a locally maintained catalog and is exposed to multiple failure modes (i.e financial, organisational, technological etc.).

6.5.2 Risk is partially mitigated by appropriate discovery metadata (package metadata) being included in each AIP which could be used to re-populate a catalog.

6.6 Disaster plan

6.6.1 A disaster is an unexpected and rapid change event that adversely affects the ability of either the depositor or the trusted store to provide the expected level of service.

6.6.2 The risk of a disaster is mitigated by there being an agreed disaster plan which includes invocation, communication and response.

DigitalPreservationEurope, 2008. DPE Repository Planning Checklist and Guidance DPE-D3.2. [Available from: https://digital.library.unt.edu/ark:/67531/metadc799759/m2/1/high_res_d/platter.pdf]

Appendix

AIP container file

A popular information package container specification is BagIt created by the Library of Congress. A reference implementation is available.

Historically both tar and zip serialization were supported by the reference implementation but latterly BagIt ignores serialization leaving this to the user.

Zip is now commonly used due to the widespread availability of open source cross-platform tools.

Either tar or zip serialized container files can be compressed. However this should be avoided.

AIP container file names should be unique. A popular way to achieve this is to use a UUID. A file name extension should be optional.

Several fixity values for the AIP are calculated and recorded. A fixity value is a cryptographic hash or message digest obtained by encoding the container file bit string. Message digests are often described ambiguously as checksums.

AIP content is not “virus checked”.

AIP content is not encrypted. The need to maintain decryption keys for a 100 years external to the AIP and for this to be a pre-requisite to accessing preserved information breaks OAIS.

DIP container file

The DIP container file is similar in outline to the AIP container file.

The differences are,

- the DIP is essentially ephemeral and the container file name need not be unique since it will be used by only a single end user,
- there is no requirement to manage DIP fixity,
- in addition to the transformed content files, the DIP also contains relevant intellectual property and terms of use statements.

Package metadata

METS (Metadata Encoding for Transmission Standard) maintained by the Library of Congress is an example of a relevant enduring XML schema.

Any ancillary schema used will also be enduring either because they are open or because schema documentation is included in the AIP.