



AHDS Digital Preservation Glossary

Final version

prepared by

**Raivo Ruusalepp
Estonian Business Archives, Ltd.**

January 2003

Table of Contents

1. INTRODUCTION.....	1
2. PROVENANCE AND FORMAT.....	1
3. SCOPE AND ENHANCEMENT.....	1
4. TERMINOLOGY.....	2
ACCESS.....	2
ACCESS FORMAT.....	2
ACCESSION.....	2
ACQUISITION/COLLECTION POLICY.....	2
ARCHIVE (NOUN).....	2
[ARCHIVING (VERB).....	2
AUDIT TRAIL.....	3
BACK-UP COPY.....	3
CATALOGUING.....	3
CHECK-SUM.....	3
COLLECTION.....	3
COPY (NOUN).....	3
COPY (VERB).....	4
COPYRIGHT.....	4
CONSUMER.....	4
DATA TYPE.....	4
DEPOSIT (NOUN).....	4
DEPOSITOR.....	4
DESCRIPTION.....	4
DESIGNATED COMMUNITY.....	5
DIGITAL PRESERVATION.....	5
DIGITAL RESOURCE.....	5
DIGITAL SIGNATURE.....	5
DISSEMINATION FORMAT.....	5
DISSEMINATION VERSION.....	6
DOCUMENTATION.....	6
ENCRYPTION.....	6
FILE.....	6
FILE COMPRESSION.....	6
FILE FORMAT.....	6
FINDING AIDS.....	7
FIXITY INFORMATION.....	7
HARDCOPY.....	7
HOLDINGS.....	7
INGEST.....	7
LEGACY SYSTEM.....	7
LIFE-CYCLE CONCEPT.....	8
LONG-TERM.....	8
METADATA.....	8
MIGRATION.....	8
OFF-LINE (STORAGE).....	8
OFF-SITE (STORAGE).....	9
ON-LINE (STORAGE).....	9
OPEN ARCHIVAL INFORMATION SYSTEM (OAIS).....	9
OPERATING ENVIRONMENT.....	9
ORIGINAL VERSION.....	9
PRESERVATION COPY.....	9
PRESERVATION FORMAT.....	9

PRESERVATION METADATA.....	9
PRESERVATION STRATEGY	10
PRESERVATION VERSION.....	10
PROPRIETARY	10
REFRESH.....	10
RENDER.....	10
RENDITION	10
RESOURCE DISCOVERY METADATA.....	10
RIGHTS OWNER	11
SUBMISSION SESSION.....	11
THESAURUS	11
USER.....	11
USER-FRIENDLY	11
VALIDATION	11
VARIANT	12
VERSION.....	12
VIRUS.....	12
APPENDIX	13
SOURCE GLOSSARIES USED.....	13

1. Introduction

This glossary contains definitions of key digital preservation terms for use throughout the AHDS and its Service Providers. It has been prepared as part of a preservation consultancy commissioned by the AHDS.

As the digital preservation discipline is still developing, there are no widely-agreed standards for the use of terminology. Consequently the AHDS has identified the need for this glossary, developed expressly for this community but based as much as possible on work published elsewhere.

The AHDS digital preservation glossary aims to:

- ? facilitate effective exchange of knowledge; and
- ? allow uniformity of practice among Service Providers.

2. Provenance and Format

This glossary is based on:

- [1] an earlier draft produced by the Cornwell Management Consultants plc. (July 2002);
- [2] interviews with the AHDS Service Providers;
- [3] AHDS Service Provider texts and documents;
- [4] other glossaries (a list is provided in the Appendix).

The body of the glossary contains a series of terms, each of which is given a formal definition. Where several different definitions have been found relevant, alternatives are provided with references to their sources. It is expected that the subsequent discussion will identify the best definition for the AHDS context.

These definitions are followed, where relevant, by an indication of the source of the definition; however, note that in some cases the wording of the source definition has been adjusted for AHDS use. Where no source is shown, the definition has been developed specifically for this glossary.

[Terms in *italics* are defined within this glossary.

Definitions are followed by explanatory notes, which are shown in *italics*.]

3. Scope and enhancement

The glossary covers digital preservation terminology that is relevant in, and sometimes specific to, the context of the AHDS. It does not define various technical computing terms and acronyms unless they have a specific use within the AHDS. It is envisaged that the glossary be updated and expanded periodically.

4. Terminology

Access

Terms and conditions of granting permission to use *data resources* and collections in an *archive*. Access may be restricted in some instances because of copyright, confidentiality or statutory requirements.

Often referred to as Archival Access. In data processing, access is the process of entering data into and retrieving data from computer memory.

Access Format

A *format* used to present a preview of a *digital resource* during resource discovery.

This is typically an image with relatively low resolution, bit depth and pixel count, used to create a “thumbnail” rendition efficiently. Often referred to as a “thumbnail”.

See also: *Dissemination Format, Preservation Format, Dissemination Version.*

Based on: [14].

Accession

The processes of receiving, preparing, cataloguing and storing *digital resources* in a form suitable for *digital preservation*.

See also: *Deposit, Ingest.*

Based on: [12]. Similar to *Ingest* in [10].

Acquisition/Collection Policy

The official statement issued by an *archive* identifying types of *data resources* it will collect or acquire and the terms and conditions under which it will do so.

Archive (noun)

An organisation that intends to preserve information for access and use by a designated community.

Also used as Archives, Archival Repository.

Source: [10]. Compatible with [2].

[Archiving (verb)]

In the AHDS context, this verb could be used to denote the creation of the preservation version and submitting it into the AHDS archive.]

Audit trail

A means of tracking all the interactions with records within an electronic system so that any access to the system can be documented as it occurs for the purpose of preventing unauthorised actions in relation to the records.

Alternative:

A record showing who has accessed a computer system and what operations he or she has performed during a given period of time.

Source: [18], [16].

Back-up Copy

A copy of information maintained *off-site* from the original as a safeguard against disaster or computer failure.

Back-ups are usually copied to storage devices that can be removed from the computer and kept separately from the original. The essential attribute of a back-up copy is that the information it contains can be restored in the event that access to the master copy is lost.

Cataloguing

The process of arrangement and *description* of collections to produce a structured list or catalogue, which enables users to locate the data resources they need.

Check-sum

A number derived algorithmically from a document or other arbitrary-sized stream of information, intended to check the transmission or recording of information.

Source: [17]

Collection

In the AHDS, a data collection encompasses the deposited original data resource, the preservation and dissemination versions created from it and the documentation that accompanies the data resource.

Some of the Service Providers call a collection a 'study'.

See also: *Documentation, Data Resource.*

Copy (noun)

A bitwise reproduction.

A copy is identical to the original, bit for bit, except in some cases for the unique identifier (e.g., file name).

See also: *Variant, Refresh.*

Copy (verb)

The act of creating a bitwise reproduction.

See also: *Refresh, Migration*.

Copyright

A statutory right that grants creators (authors) certain exclusive rights in their creations for a legally established duration of time.

See also: *Rights Owner, Proprietary*.

Source: [16].

Consumer

See: *User*.

Data Type

- (1) Set of data with predefined characteristics (e.g., integer, floating point unit number, character, string, pointer).
- (2) On the application software level broader categories of data types are in use: text, database, tabular data, graphics image, digital audio, digital video, GIS, CAD, etc.

Deposit (noun)

One or several *digital resources* received by an *archive* for preservation.

Alternative:

The *data resource(s)* placed in the custody of an *archive* without transfer of legal title. [16]

Also referred to as Submission and Submission Information Package (particularly in the OAIS reference model).

Depositor

An individual, group, or organisation which offers a *deposit* to an *archive* for preservation and dissemination.

Based on: [6].

Description

The process of capturing, analysing, organising and recording information that serves to identify, manage, locate and explain data resources and the contexts that produced them.

See also: *Cataloguing, Metadata*.

Based on: [16].

Designated Community

An identified group of potential *users* who should be able to understand a particular set of information. The Designated Community may be composed of multiple *user* communities.

Based on: [10].

Digital Preservation

Storage, maintenance and access to *digital resources* over the long term.

Abbreviated from [2]. Compatible with [11].

Alternative:

Ensuring the usability of a *digital resource* through changing technological regimes with a minimum loss of the resource's intellectual content.

Source: M. Feeney (ed.), "Digital Culture: Maximising the Nation's Investment", NPO, 1999, p. 41

Digital Resource

Digitally encoded information which forms a logically meaningful unit that is *deposited* with an *archive* and what the users may *access* from the *archive*.

For example, an image, or a table from a database.

A digital resource may form a separate physical unit (a 'file' in IT terms), or it may be formed of several files.

A digital resource may consist of digital objects.

Based on: [6]. Compatible with [10] but with a different meaning.

Digital Signature

Information which, using cryptographic techniques, provides guarantees of the authenticity and/or reliability and/or authorship of a *data resource*.

Alternative:

A digital mark that has the function of a signature in, is attached to, or is logically associated with a record, and is used by a signatory to take responsibility for, or to give consent to, the content of the record.

See also: *Encryption, Check-sum.*

Source: [17], [18].

Dissemination Format

A *format* used to present a *digital resource* to a *user* who has requested it.

See also Access Format, Preservation Format.

Dissemination Version

A version created from the *original version* of a *data resource* for dissemination to *users*.

In OAIS terms, the Dissemination Information Package that the user receives, may be derived from one or more Archival Information Packages.

See also: *Original Version, Preservation Version, User.*

Based on: [10]

Documentation

All the material that provides information and guidance on how to interpret the contents of a *digital resource*, and which describes its contents, provenance, structure and other attributes.

Information needed to permit access and retrieval of data resources.

See also: *Cataloguing, Metadata.*

Based on: [6].

Encryption

The process of encoding messages, including electronic data, for security purposes.

See also: *Digital Signature, Check-sum.*

Source: [17].

File

A logical assembly of data stored within a computer system. [16]

Alternative:

Set of data held on computer, given single identifier and treated as a unit. Files can be grouped into folders.

Source: [16], [17].

File Compression

Reduction of file size for storage, processing and transmission.

File Format

An attribute of a *file* which describes its encoding.

Alternatives:

The organisation of information for storage, printing, or displaying.

The specification of how the bits stored in a file should be interpreted. [6]

For example, .msg, .pdf, TIFF v6, JPEG 2000, Microsoft Word 7.0.

See also: *File, Access Format, Dissemination Format, Preservation Format.*

Finding Aids

Reference material such as catalogues, lists and indices providing information on the *data resources* held by an *archive*.

See also: *Catalogue, Resource Discovery Metadata.*

Fixity Information

The information which documents the authentication mechanisms and provides authentication keys to ensure that the *data resource* has not been altered in an undocumented manner.

See also: *Check-sum, Digital Signature.*

Based on: [10].

Hardcopy

Documentation and *data resources* in physical formats (e.g., paper) capable of being read without the assistance of a technical device.

Holdings

The whole of the archival material and collections in an *archives*.

See also: *Collection, Data Resource.*

Ingest

The OAIS entity that contains the services and functions that accept Submission Information Packages from depositors, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.

See also: *Accession, Deposit.*

Based on: [10].

Legacy System

An old application software that is continued to be used (e.g., because the cost of replacement or redesign is high).

Alternatives:

Previous generation or version of a system and its contents (legacy data) which needs special treatment to make it usable in a current IT environment.

Source: [16].

Life-cycle Concept

A concept that draws an analogy between the life of a biological organism, which is born, lives and dies, and that of a digital resource, which is created, used and then disposed of or transferred to an archives.

See also: *Digital Resource*.

Source: [16].

Long-Term

A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and file formats, and of a changing *user* community, on the information being held by an *archive*.

See also: *Digital Preservation*.

Based on: [10].

Metadata

Data about other data.

Metadata is collected with the purpose of explaining the technical and administrative processes used to create, manipulate, use and preserve the data resource. It is often subdivided into further categories.

See also: *Preservation Metadata, Resource Discovery Metadata, Documentation*.

Source: [10].

Migration

The conversion of data resource from one hardware/software configuration or generation to another.

This is generally done to minimise the risk of information loss caused by software or hardware obsolescence. It may involve new generations of hardware/software and/or moves to different hardware/software.

Consistent with [12], [16].

Off-line (storage)

Storage that is not linked to a live/on-line system and which, therefore, takes some time to access.

Off-site (storage)

Sufficiently distant from a location to be unlikely to be affected by a disaster which affects the location.

On-line (storage)

Storage that is linked to a live/on-line system and that allows immediate retrieval of data resources.

Open Archival Information System (OAIS)

An *archive*, consisting of an organisation of people and systems, that has accepted the responsibility to preserve information and make it available for a *Designated Community*. The term 'open' in OAIS is used to imply that its reference model and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

Based on: [10].

Operating Environment

All the hardware and software that is needed to run a *digital resource*.

Original Version

The original deposited *data resource* that is preserved by the AHDS Service Provider without any changes or alterations to the content.

See also: *Preservation Version*, *Dissemination Version*.

Preservation Copy

A *copy* made and used to preserve the intellectual content of a *digital resource*.

Preservation Format

A *format* chosen for preservation purposes.

See also: *Access format*, *Dissemination format*.

Preservation Metadata

Preservation metadata – both technical and administrative – is kept to document the preservation processing in an *archive*, make it transparent and accountable and also in the preservation processing itself.

See also: *Metadata*, *Documentation*.

Preservation Strategy

Coherent set of objectives and methods for maintaining digital components and related information over time, and for reproducing the related authentic *data resources*.

See also: *Digital Preservation, Migration, Copy*.

Based on: [18].

Preservation Version

The preservation version is intended to guarantee the long-term survival of the *deposited digital resources*. When creating the preservation version, the AHDS Service Provider may convert the original version files to new file formats that are suitable for long-term preservation.

See also: *Original Version, Dissemination Version*.

Proprietary

A file format that is privately owned and controlled, the specifications are generally not open.

Based on: [16].

Refresh

The process of *copying digital resources* from one storage medium to the same storage medium.

This is generally done to minimise the risk of information loss caused by media degradation.

Based on: [11], [10].

Render

The process of producing a *rendition*.

Source: [5].

Rendition

The manifestation of a *digital object* presented (i.e., rendered) to which a *user* can refer.

This may include on-screen display, printed and audio and multimedia presentations.

Based on: [5].

Resource Discovery Metadata

The catalogue and index record component of a *collection*.

See also: *Cataloguing, Metadata, Documentation*.

Rights Owner

An individual, group, or organisation which holds intellectual property rights in some *digital resource(s)*.

See also: *Copyright*.

Submission Session

A delivered set of media or a single telecommunications session that provides *data resources* to an *archive*. The Submission Session format and contents are based on a model negotiated between the *archive* and the *depositor* in the Submission Agreement.

See also: *Accession, Deposit*.

Based on: [10].

Thesaurus

A controlled and structured vocabulary of keywords showing synonymous, hierarchical and other relationships and dependencies.

Source: [16].

User

Any member of the public who is allowed access to the *archive* and its *holdings*.

Alternatives:

Anyone who needs, uses or benefits from the *data resources* held by an *archive*.

The role played by those persons or client systems who interact with an archive to find preserved data resources of interest.

In the OAIS terminology User is the same as Consumer.

Based on: [10], [16].

User-friendly

Computer software or hardware that is simple to set up, run and use.

Source: [16].

Validation

Quality assurance performed by checking the contents of a *file* or a *digital resource* at the time of *deposit* or creation of *preservation and/or dissemination versions*.

Variant

A modified version of a *component* which includes all the information content of the original component but which differs by some attribute(s) other than unique identifier.

For example, a variant may have a different format, or it may feature the removal of a digital watermark.

See also: *Copy*.

Version

One of two or more forms in which a data resource is stored.

See also: *Original Version, Preservation Version, Dissemination Version*.

Based on: [18].

Virus

A computer program or code that is loaded (e.g., transferred) without user knowledge and runs without instruction.

Alternative:

A computer program that is transferred to one or more computers with the intention of corrupting or wiping out information in the recipient computer.

Based on: [16], [17].

APPENDIX

Source Glossaries used

- [1] ASTRA “Sesam” specification (1996)
- [2] Cedars Project (May 1999)
- [3] National Archives and Records Administration (USA), “Digital-Imaging and Optical Digital Data Disk Storage: Long-Term Access Strategies for Federal Agencies” (1994)
- [4] European Commission, “Digital Preservation Guidelines” (1998)
- [5] European Commission, “MoReq: Model Requirements for ERMS” (2001)
- [6] History Data Service Collections Manual (version 2.02) (August 2002)
- [7] ISO 15489:2001 “Information and Documentation – Records Management”
- [8] National Library of Australia, “PANDORA Data Dictionary” (1994)
- [9] “The NEDLIB Glossary” (2000)
- [10] CCSDS, “Reference Model for an Open Archival Information System (OAIS)” (version 650.0-B-1) (2002)
- [11] M. Jones, N. Beagrie, “Preservation Management of Digital Materials” (2001)
- [12] CPA & RLG, “Preserving Digital Information. Report of the Task Force on Archiving of Digital Information” (1996) [this report does not include an explicit glossary]
- [13] Public Record Office Digital Preservation Glossary (unpublished)
- [14] TASI, “Glossary of Technical Terms” (2001)
- [15] “Universal Preservation Format Glossary” (1999)
- [16] IRMT & ICA, “Managing Public Sector Records. Glossary” (1999)
- [17] UCL Summer School on Managing Electronic Records, “Glossary” (2001)
- [18] InterPARES, “The InterPARES Glossary: A controlled vocabulary of terms used in the InterPARES Project” (2002)