Repository Operations Manual

Version 5.0

Created date:	2004
Last updated:	31st July 2019
Review Due:	August 2020
Authors:	Tim Evans, Ray Moore, Digital Archivists
Maintained by:	Archives Manager, Digital Archivists
Previous version:	Repository Operations Manual (Version 4.0)

1. Purpose of this document

This document summarises the internal requirements for preparing collections for preservation and for their onward management within the repository. It includes information on the ADS' implementation of the OAIS model and the form and structure of individual archives. The document provides guidance on the storage of datasets and the required documentation, alongside directory and file naming conventions used within the repository. The ADS acknowledges the need for consistency in data structure and form throughout the repository.

2. Introduction

This document makes considerable use of terminology and concepts from the OAIS reference model.

The definition of OAIS, as outlined by the Consultative Committee for Space Data Systems document, is given as "an archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community".[1]

The six mandatory responsibilities for an OAIS compliant archive are:

[1] Consultative Committee for Space Data Systems (2012) *Reference Model for an Open Archival Information System (OAIS). Magenta Book. Issue 2. June 2012.* pp1-2 https://public.ccsds.org/Pubs/650x0m2.pdf, see also Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS)'*. CCSDS 650.0-B-1 Blue Book,

- Negotiate for appropriate deposits in accordance with the Collections Policy[2] and the *Guidelines for depositors*[3]
- Obtain sufficient control of resources
- Determine scope of community
- Ensure independent utility of data
- Follow procedures for preservation
- Disseminate to the designated community

In the OAIS model, information packages move from producers, through the OAIS, and on to the data consumers. All SIPs follows the workflow outlined in the ADS' *Ingest Manual*.[4] The preservation of data in accordance with the guidelines given in the *Ingest Manual*[4] and the ADS' data procedures follows with the creation of the *Archival Information Package* (AIP) and *Dissemination Information Package* (DIP). A digital interface is produced for each submission with provides data consumers and users with access to the DIP.[5]

3. Conventions within this document

A Source Code block is used to highlight text that represents identifiers, directory names, file names and similar concepts.

Words or phrases in curly braces ('{' and '}') are placeholders that should be replaced with the appropriate content. For example, AIP-identifier indicates that a valid AIP identifier, of the form arch-{collection number}-{version number}, should be inserted, where, in turn, {collection number} and {version number} must also be replaced with actual values, while the dashes, which are not enclosed by braces, are literal text to include in the final AIP identifier.

Square brackets ('[' and ']') are used to indicate a set of choices, from which one choice should be selected. Each choice is separated from the next by a pipe character ('|'). For example {file name}.[pdf | .tif] indicates a file name that should finish with either the extension '.pdf' or the extension '.tif'.

4. Ingest and the *Submission Information Package* (SIP).

Within the OAIS model, information sent from data producers to the repository, through the exchange of the data and the creation of the *Submission Information Package* (SIP).The submission of data to the repository, through one of the submission streams, allows the exchange of data alongside technical and contextual metadata. Depositors use one of the repositories deposition streams:

---

[2] *Collections Policy* - https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml
[3] *Guidelines for Depositors* - http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml
[4] *Ingest Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest
[5] For example, Crossrail Ltd (2019) *Crossrail: Archaeological Investigations Conducted in Advance of Construction of the Elizabeth Line* [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1055125.

- OASIS[6]
- OASIS Images[7]
- ADS EASY[8]
- external data exchange service[9]
- exchange of digital media[10]

The ADS will accept SIPs that fit the requirements outlined in the *Collections Policy*[2] and the *Guidelines for depositors*[3], where questions arise about the suitability of a dataset for archive; the ADS refer to its Collections Evaluation Working Group, drawn from its Management Committee to facilitate the evaluation of datasets for submission.[11]

The ADS also supply guidance on the selection of material for deposition,[12] with more specific information on the deposition of personal, confidential and sensitive data[13] and digitisation.[14] The ADS also works closely with local and national agencies within the archaeological sector to provide help and support those creating and preserving digital data. The Collections Development Manager, and repository staff, provide specific guidance and assistance to depositors on the requirements outlined by the repository.

Detailed information on the ingest process is outlined in the *Ingest Manual*.[4] Once accession of the SIP is complete, the repository creates the *Archival Information Package* (AIP).

5. Archival Storage and the *Archival Information Package* (AIP)

AIP is the OAIS term for a coherent set of information preserved by the repository. Deposited data, or SIP's, that conform to the requirements provided in this document and outlined in repositories *Guidelines for Depositors*[3] and other documentation (see below), are added to the AIP.

---

[6] http://oasis.ac.uk/pages/wiki/Main
[7] An image only service available through the OASIS system - http://oasis.ac.uk/pages/wiki/Main
[8] http://archaeologydataservice.ac.uk/easy/
[9] These can include the ADS' parent organisation, University of York's, file sharing service (DropOff Service - https://www.york.ac.uk/it-services/services/dropoff/) or other commercial file sharing services DropBox, GoogleDrive, SFTP, etc.
[10] The exchange of DVD, USB or portal hard-drives in person, or through the postal service.
[11] In accordance with the *Collections Policy*, section *2.3 Criteria for evaluating electronic datasets* - https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml
[12] *Guidance on the selection of material for deposit and archive* - http://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml
[13] *Policy and Guidance on the Deposition of Personal, Confidential and Sensitive Data* - https://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml
[14] Digitising journal articles and grey literature reports - https://archaeologydataservice.ac.uk/advice/scanningGuide.xhtml

The AIP consists of files containing the data, documentation, metadata and administrative material (scanned or digital licence, correspondence etc.) for a collection.

The deposited data itself represents what the OAIS model defines as the content information, which is the actual material being preserved. In addition, a part of the AIP is the *Preservation Description Information* (PDI), which is the administrative metadata used to plan and manage the preservation of the content information. At the ADS, collection level metadata and administration information are stored in the *Collection Management System* (CMS), while an extension to this: the *Object Management System* (OMS), holds both technical and contextual metadata pertinent to the individual files.[15]

In circumstances where depositors provides a 'new edition' of a dataset, that is, one that replaces an earlier version, a new AIP is created and added to the existing AIP in line with OAIS model. The AIP can also hold those preservation files normalised or 'migrated' to newer, more sustainable, formats following migration.[16]

5.1 AIP Directory Naming

Each AIP must have a persistent identifier, taking the form:

arch-{collections number}-{version number}
(e.g. arch-335-1)

The collections number will be a unique identifier for the collection generated programmatically by the *Collections Management System* (CMS).[1515]

The first version of a collection should be given the edition number 1, and subsequent editions should be numbered sequentially using positive integers (e.g. 2, 3, 4 etc.). The second edition of a dataset will be given the number '2', the third '3', etc., so, for example, the identifier for collection '335' will become 'arch-335-2' on submission of the second edition, and then 'arch-335-3' for the third.

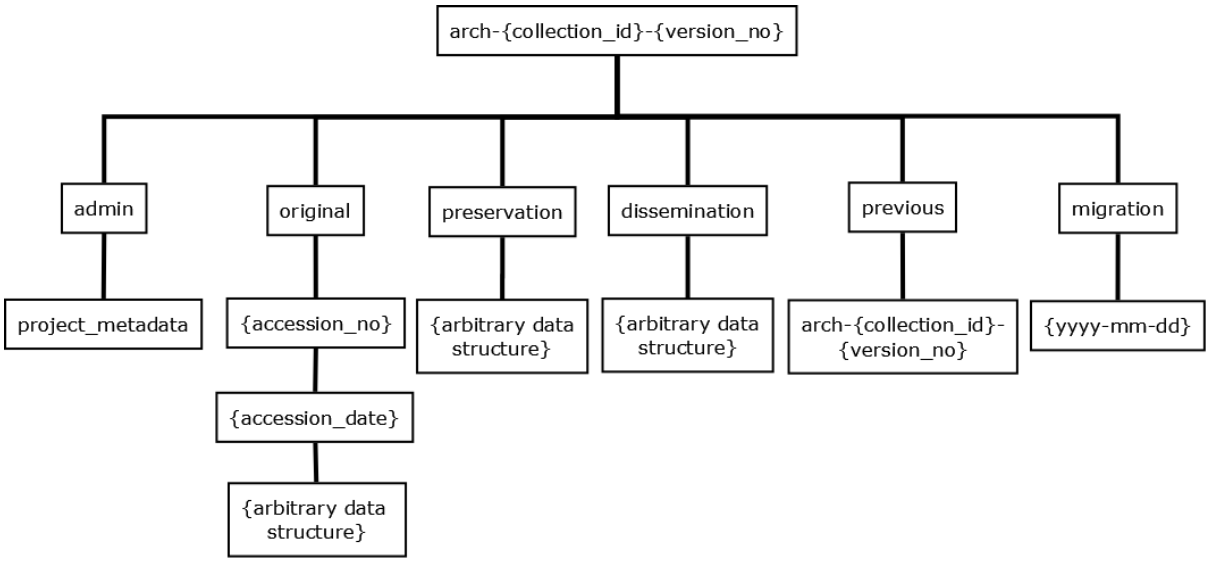
As noted above, an AIP may hold more than one edition of a dataset, with previous versions of the collection retained and stored, within a /previous/ directory, following the structure outlined below.[17]

In instances where files, or data, have been migrated to new, or more sustainable, formats subsequent to the initial phase of preservation, these are stored, in a /migration/ directory, following the structure outlined below.[18]

[15] Both CMS and OMS are only available for internal access only.
[16] Details of these aspects are available below.
[17] See section 9 *Examples of ADS directory structure*, Example 3 below.
[18] See section 9 *Examples of ADS directory structure*, Example 4 below.

5.2 AIP Directory Structure

The AIP directory tree is a logical rather than physical structure, with its contents spread across multiple storage devices (for example, for security purposes, the preservation data is stored in a separate location to the dissemination). Consequently, a consistent directory structure is required, although implementation of the OMS, which provides an accurate record of file locations and relationships, allows us to store information about files and other aspects of the AIP outside of this file structure. See Appendix 2 for further information on this.



5.2.1 The 6 main subdirectories

The AIP directory structure has seven main subdirectories (summarised below). These neatly organise the material in an AIP according to its purpose and origin.

- **/original/** – contains the SIP, i.e. data and metadata/documentation received from the depositor and accessioned in accordance with the *Ingest Manual*). [19]
- **/admin/** – contains administrative material, notably metadata about the contents of the dataset (as a whole, file/data specific documentation/metadata is stored with the data), the licence agreement (scanned or digital) alongside correspondence specific to the archive and its preservation (as a whole).
- **/preservation/** – contains preservation files and file-level metadata that have been normalised, in accordance with the *Ingest Manual*,[19] into preservation formats.

[19] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest.

- **/dissemination/** – contains dissemination files and file-level metadata, the DIP.
- **/previous/** – used to hold previous editions of the AIP.
- **/migration/** – used to hold old preservation normalisations of files that have been normalised into newer formats.

The naming of directories should be accordance with the guidance outlined above, with all names given in lower case. This level of the AIP directory must not contain any other files or directories. Details on the content and structures used within each of these directories are provided in the relevant section below.

5.2.2 Original

The /original/ directory should contain the following:

- All files received from the depositor and accepted as part of the final deposit.

The /original/ directory should contain:

Directory name	Description	Comments
/ original / {accession number} **[current]**	A subdirectory, named with the appropriate {accession number}.	Created under /original/ to hold data from each discrete accession/SIP. [20]
/ original / {accession number} / {accession date} **[current]**	A subdirectory, named with the appropriate {accession date}.	Used to distinguish data received from the depositor at different times in the accession process. Should be named with the {accession date}, in the form 'yyyy-mm-dd', (for example, /2016-04-30/). Subsequent depositions, which forms part of the same accession, can be added to the directory with correct {accession number}, and named, as above, with the {accession date} of the additional material.**Error! Bookmark not defined.**

[20] See diagram above, or Section 9 *Examples of ADS directory structure*, Example 1 below.

/ original / {accession number} / {accession date} / {arbitrary} /	Beneath this, the directory structure is somewhat arbitrary and dependent on the SIP provided by the depositor.	Generally, repository staff maintain the original data of the dataset received from the depositor (see note below).

Note: The repository actively discourages changes to directory structure and filenames provided by the depositor within the SIP. In order to mitigate for poorly organised datasets, or problematic file names, all depositors are encouraged to follow the guidance within the *Guidelines for Depositors*.[21] For those depositors using *OASIS Images*[7] or *ADS-easy*[8] programmatic checks ensure that data follows the dedicated file naming policy. However, in rare instances where the data structure is problematic or incongruous repository staff may rearrange a dataset into more logical structure. Similarly, where file naming does not follow the guidance provide, they may be renamed to follow the guidelines. In both instances, the documentation of any changes, within the *Collections Management System* (CMS), is essential.[15]

SIP File name changes:

Character	Name	Replace with / Strip	Notes
	Space	_	
.	Period / full stop	**Strip**	
,	Comma	-	
|	Pipe	**Strip**	
/ \	Slash and Backslash	**Strip**	
?	Question mark	**Strip**	
*	Asterisk	**Strip**	
"	Quotation mark	**Strip**	
:	Colon	**Strip**	
;	Semicolon	**Strip**	
> <	Greater and less than	**Strip**	
+	Plus	_plus_	
[ ]	Square brackets	**Strip**	
%	Percent sign	**Strip**	

[21] *Guidelines for Depositors* – File Management - https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#File%20Management.

^	Caret	**Strip**	
( )	Round brackets	**Replace with '_' or '-'**	
!	Exclamation mark	**Strip**	
@	At	_at_	
#	Hash / pound	**Strip**	
$	Dollar	**Strip**	
`	Grave accent	**Strip**	
{ }	Curly brackets	**Strip**	
~	Tilde	**Strip**	May indicate a temporary or backup file
'	Apostrophe	**Strip**	
=	Equals	**Strip**	
&	Ampersand	_and_	

The original directory should therefore have a structure much like this:

arch-{collections number}-{version number} / original / {accession number} / {accession date} / {depositor's file structure}

(i.e. arch-335-1 / original / 770 / 2008-04-23 / {depositor's file structure})

5.2.3 Admin

The /admin/ directory should contain the following:

- Any documentation or files pertinent to the management and administration of the SIP are stored within the directory.

This /admin/ directory should contain the following files:[22]

File name	Description	Comments
licence.tif, or licence.pdf (licence_cc-by.pdf [**deprecated**]	A deposit licence scanned and saved as an uncompressed TIF or PDF file,[23] or when issued digitally stored in PDF form.	A signed (physical or electronic) deposit licence accompanies all depositions.[24] Previously the type of licence was documented in the file name alongside the CMS, but

[22] See also Appendix 4 regarding reserved file names.
[23] All scanned licences and documents should follow the 'Requirements for Scanned Hardcopy Material' (Appendix 3) below.
[24] The terms of access found within the Annex B of the deposit licence. Additional documentation of terms of access are documented within the *Collections Management System* (CMS), within the

		now is solely documented within the CMS. Any extensions to the licence, or explanations of the characteristics, should be stored in an ancillary file (suitable for preservation) with a name that adheres to the file naming policy, e.g. licence_addendum.txt. Additional copies of this information are stored in the 'uploads' area of the CMS, and documented in the 'licences' section.

5.2.3.1 Project Metadata

The /admin/ folder may additionally contain the following items that should be stored in a directory called /project_metadata/:

File name	Description	Comments
dc_metadata.txt **[deprecated]**	A TXT file containing the Dublin Core metadata record loaded into ArchSearch[25] and exported from the CMS.	No longer actively undertaken, may exist for older archives. The metadata continued to be collected, but is stored in the CMS.
original_{accession_id}.txt **[deprecated]**	A TXT file containing checksums for all files within the original deposition.	No longer actively undertaken for current collections, but may exist for older archives. Partially replaced by a more extensive deposit receipt with includes checksum information for the deposition.
email_{yyyy-mm-dd}.txt, or email_{yyyy-mm-dd}.pdf	Copies of any e-mails sent to, or received, from the depositor, or data producers.	Store only those mails which: - help document the SIP

dedicated 'licences' section. The CMS is only available for internal purposes, but the terms of use are clearly displayed (within the side bar) of each archive interface. For example, Kevin Camidge (2019) Wheel Wreck Investigation 2018 [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1055091.

[25] The repository integrated online catalogue of metadata records, which include ADS collections and metadata harvested from UK historic environment inventories - https://archaeologydataservice.ac.uk/archsearch/basic.xhtml.

		- give us permission to remove or edit files in the SIP - have been sent with original data to be accessioned into the archive - clear up copyright issues surrounding any of the data.[23]
Project metadata files (e.g. ADS_collection_metadata.docx)	Other files (normally supplied as part of the SIP) that help document or describe the project as a whole. All file names should follow the 'file naming policy'[26],[21] with any file name changes documented in the CMS.[27]	If the documentation/metadata relates to actual data, whether a group of files or a particular file, this should be stored in a /documentation/ directory alongside the data it refers too.
Introduction and Overview text (e.g. introduction.docx, or overview.txt)	Introduction and Overview text for the web interface should be stored here.	Generally included within the collection metadata template and not submitted separately. The repository may create introduction or overview text, often from other documentation within the collection, where it is absent. This is not stored separately.
deposit_receipt_{accession_id}.csv **[deprecated, but reinstated]**	A digital copy of the deposit receipt sent to the depositor after accessioning of data this should be stored, in a suitable preservation format (CSV), within the /project_metadata/ directory.	For a number of years this deprecated, but was later reinstated (in 2017) to make it clear to depositors what the SIP contained and to provide an audit of the original checksums. This may also be stored as an e-mail (see above), or since August 2017 as a separate export from the OMS (in CSV form). This file is also be attached to the CMS record ('Upload' section).

[26] See discussion on file naming above (5.2.2 Original).
[27] Any metadata or documentation held in the /preservation/ directory does not be require duplication within the /admin/ directory.

email_deposit_{date_received}.txt, or email_deposit_{date_received}.pdf	A digital copy of the deposit email sent to the depositor following the accession of the dataset. This should be stored in a suitable preservation format (CSV or PDF) within the /project_metadata/ directory.	This file is also be attached to the CMS record ('Upload' section).

Note: it is important that all files within the /admin/ directory are stored in a suitable preservation file format.[28]

The admin directory should therefore have a structure much like this:

arch-{collections number}-{version number} / admin / project_metadata /

(i.e. arch-335-1 / admin / project_metadata /)

5.2.4 Data directories (preservation and dissemination)

The AIP Directory folder contains the following directories for holding the data component of the archive:

- /preservation/ - All files allocated for preservation (AIP)
- /dissemination/ - All files allocated for dissemination (DIP)

These directories may contain the following subdirectories:

Directory name	Description	Comments
/ {file type} / or / {data type}/ **[deprecated]** (e.g. /pdf/, /xlsx/, /gis/, /geophysics/, etc.)	Previously data would be stored in discrete 'file type' or 'data type' directories within the preservation or dissemination directories.[29]	On the implementation of the OMS, which provides file location and data type information, the storage in specific directories became unnecessary, and, consequently, deprecated.

[28] A list of suitable preservation formats is provided in the 'Data Procedures' within the ADS wiki. Internal access only, but static versions of these documents are available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.
[29] See Appendix 5: Data Types for a full list of data types used by the ADS below.

/ {file type} / {arbitrary} / **[deprecated]** **/** {arbitrary} / **[current]**	Currently there is no controlled file structure within either preservation or dissemination directories other than to preserve the inherited directory structure from the original deposit, as this can provide important contextual information about the dataset. Other logical structures may be created by the ADS in order to clarify complexities or to facilitate the display of files within the online interface	Any significant changes to directory structure are recorded in the 'processes' section of the CMS.[15] Previously the {arbitrary} directory structure, inherited from the original deposition, was only permitted within the {file type} directory.
/ {documentation} / **[current]** (e.g. /preservation /documentation/)	Used to store documentation, or metadata, that relates to the entire collection, or groups of files within the collection.	In circumstances where metadata refers to a group of files or data stored in a series of sub-directories, within the {arbitrary} file structure, the documentation or metadata should be stored at the highest level in a dedicated /documentation/ directory.
/ {file type} / {arbitrary} / documentation / **[deprecated]** **/** {arbitrary} / documentation / **[current]** (e.g. /preservation/{arbitrary} /documentation/)	Used to store documentation, or metadata, that relates to the entire collection, or groups of files within the collection.	The replication of the 'documentation' directory within the arbitrary file structure allows the storage of metadata, or documentation, specific to the data within that directory (*see above*).

They should not contain the following directories:

Directory name	Description	Comments
/ dissemination / thumbs / [**deprecated**]	For the storage of thumbnails used within the archive interface.	Previously these were stored alongside data within the DIP, but thumbs are now stored within the dedicated web directory. The

		repository carried out a process of 'migration' to ensure the removal of examples from the DIP.
/ dissemination / preview / [**deprecated**]	For the storage of preview images used within the archive interface.	Previously these were stored alongside data within the DIP, but preview images are now stored within the dedicated web directory. The repository carried out a process of 'migration' to ensure the removal of examples from the DIP.

They should not contain the following files:

File name	Description	Comments
{interface_image}.jpg or {interface_image}.png	Images used within the archive interface.	Those images included within the deposition for illustrative purposes within the archive interface are stored within the /admin/project_metadata/ directory and the 'upload' section of the CMS record.[15] These images are not part of the AIP/DIP, except in instances where the files are specifically identified as data and are reused within the archive interface.
Oracle loading files	Used for loading datasets into repository databases (e.g. ArchSearch or 'special collection' databases).	Preserved in the /admin/project_metadata/ directory, rather than being part of the AIP/DIP.

The /preservation/ directory should have a structure much like this:

arch-{collections number}-{version number} / preservation / {depositor's file structure}

(i.e. arch-335-1 / preservation / {depositor's file structure})

The /dissemination/ directory should have a structure much like this:

arch-{collections number}-{version number} / dissemination /
{depositor's file structure}

(i.e. arch-335-1 / dissemination / {depositor's file structure})

Where either directory contains documentation, or metadata, it should have the structure:

arch-{collections number}-{version number} / preservation /
documentation / {depositor's file structure}

(i.e. arch-335-1 / preservation / documentation / {depositor's file structure})

arch-{collections number}-{version number} / preservation /
{depositor's file structure} / documentation

(i.e. arch-335-1 / preservation / {depositor's file structure} / documentation)

5.2.5 Previous

The /previous/ directory should contain:

- Files that were part of the AIP replaced by a new edition.

Directory name	Description	Comments
/ previous / {previous edition} **[current]**	Used to store the previous edition of the AIP.	Preserves the file structure of the previous version.

The /previous/ directory should have a structure much like this:[30]

arch-{collections number}-{current version number} / previous /
arch-{collections number}-{previous version number} / {extant file structure}

(i.e. arch-335-2 / previous / arch-335-1/ {extant file structure}

5.2.6 Migration

The /migration/ directory should contain:

[30] See section 9 Examples of ADS directory structure, Example 3 below.

- Files created by the ADS that were once part of the AIP but are now considered obsolete, having been replaced by newly normalised files in more up-to-date file formats.

Directory name	Description	Comments
/ migration / {date files migrated} / {normalised data} **[current]**	Used to store normalised versions of files/dataset surpassed by more sustainable and current normalised versions.	The repository ensures the maintenance of the 'arbitrary' data structure, used by the depositor during the submission of the SIP, within the 'migration' directory. This ensures parity with the /original/, /preservation/, etc. directories. The repository ensures the documentation of all normalisation processes and actions within the CMS.[15] These are maintained/updated during migration to ensure the retention of a complete life cycle of the dataset.

The /migration/ directory should have a structure much like this:[31]

arch-{collections number}-{current version number} / migration / {date files migrated} / {extant file structure}

(i.e. arch-335-1 / migration / 2017-12-01 / {extant file structure}

6. Data Management

Documentation of descriptive information for data included in the AIP and SIP, and which supports the archive, is stored within the *Collection Management System* (CMS), *Object Metadata System* (OMS) and the AIP.[1515] The repository provides downloadable metadata templates and advice, through the *Guidelines for Depositors*,[3] to ensure that collection and file-level documentation is adequate. These original templates are stored, alongside the data, within the AIP and disseminated as part of the DIP.[32] These files normalised to stable preservation formats to ensure their preservation.

[31] See section 9 Examples of ADS directory structure, Example 4 below.
[32] See section 5 above. Documentation pertinent to the collection is stored within the admin/project_metadata directory, while data specific information is stored in the /documentation/ directories of the AIP.

Digital submissions of data, using OASIS,[6] OASIS Images[7] or ADS-easy[8], allow the input of the ADS' required metadata and documentation through a series of online forms. This information is then imported, and stored, within the CMS and OMS.[33]

Both CMS and OMS are also utilised to manage datasets. The OMS records technical metadata, including file location, filename, fixity value and size, alongside format, MIME type, PRONOM identifier and data type[29] that supports the continued management of datasets.[34] The creation of fixity values (checksums) allows the repository to monitor the 'health' of data during the preservation life cycle. Checksums are calculated when data is uploaded, through one of the repositories digital deposition streams, or during accession, in instances where data is submitted using other means. These fixity values (checksums) allow the repository to monitor the 'health' of files and data sets throughout the preservation life cycle.[35]

7. Administration

Administrative information about the AIP is stored within a combination of the *Collection Management System* (CMS), *Object Metadata System* (OMS) and as files within the AIP. Where administrative information is supplied as files within the SIP, these are stored within the '/admin/' and '/admin/project_metadata/' directories of the AIP (*e.g.* licence.pdf). Where appropriate

8. Preservation

The *Preservation Policy*[35] and *Ingest Manual*[36] discuss the repository function with regard to preservation. The ADS' *Data Procedures*[37] provide specific detail on the processes and practices employed by the repository.

9. Access

The *Preservation Policy*[35] and *Ingest Manual*[36] provide detailed information on the repository function with regard to the dissemination and access to datasets. The ADS' *Data Procedures*[37] provide specific detail on the processes and practices employed by the repository.

---

[33] Data specific and technical metadata and documentation are 'exported' from the CMS and OMS and stored within the DIP, alongside the data, and disseminated accordingly. Direct access to this metadata is, in some instances, provided within the archive interface.
[34] The latter is created using the *National Archives* (UK) *DROID* file characterisation software - http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/.
[35] For fuller discussion of this process, see the *Preservation Policy* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol.
[36] *Ingest Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest.
[37] Internal access only. Static versions of these Data Procedures are available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.

10. Bibliography

Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS)'*. CCSDS 650.0-B-1 Blue Book.

Consultative Committee for Space Data Systems (2012) *Reference Model for an Open Archival Information System (OAIS). Magenta Book. Issue 2. June 2012.* https://public.ccsds.org/Pubs/650x0m2.pdf

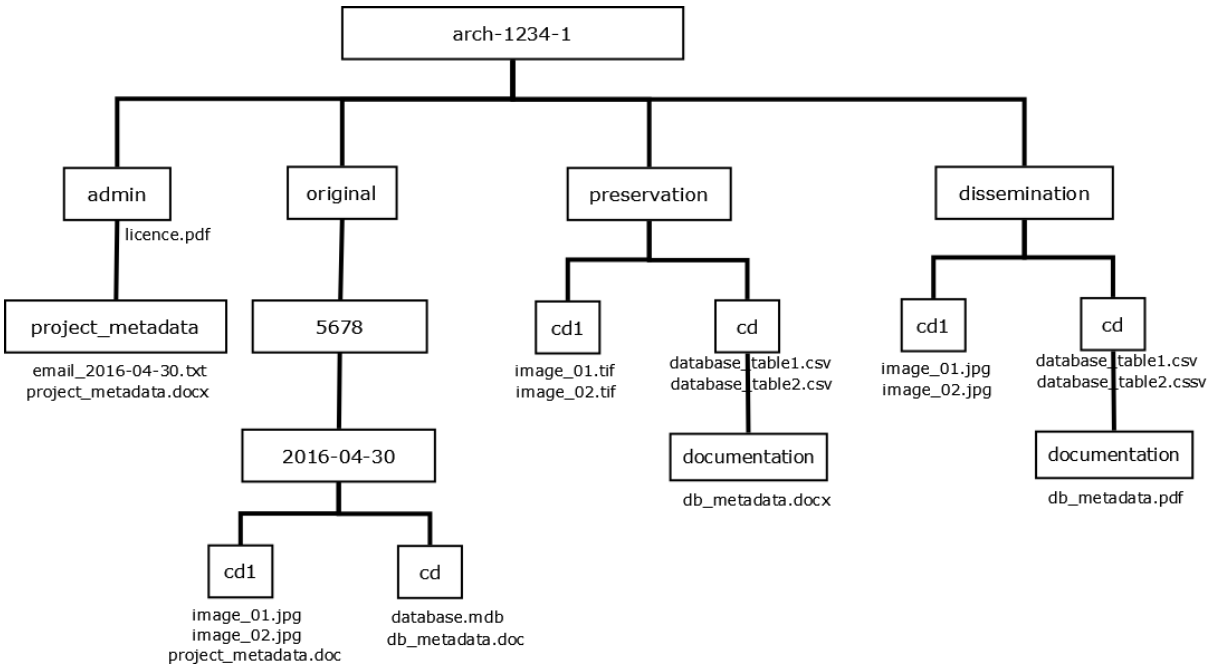
11. Examples of ADS directory structure

The following examples are intended to illustrate a number of different archiving scenarios showing how we should manage and organise the data within our directory structure:

Example 1: a 'simple' one-off deposit

In this example, a database with documentation, associated images and project metadata were deposited on 2 CDs and accessioned on the 30th April 2016.

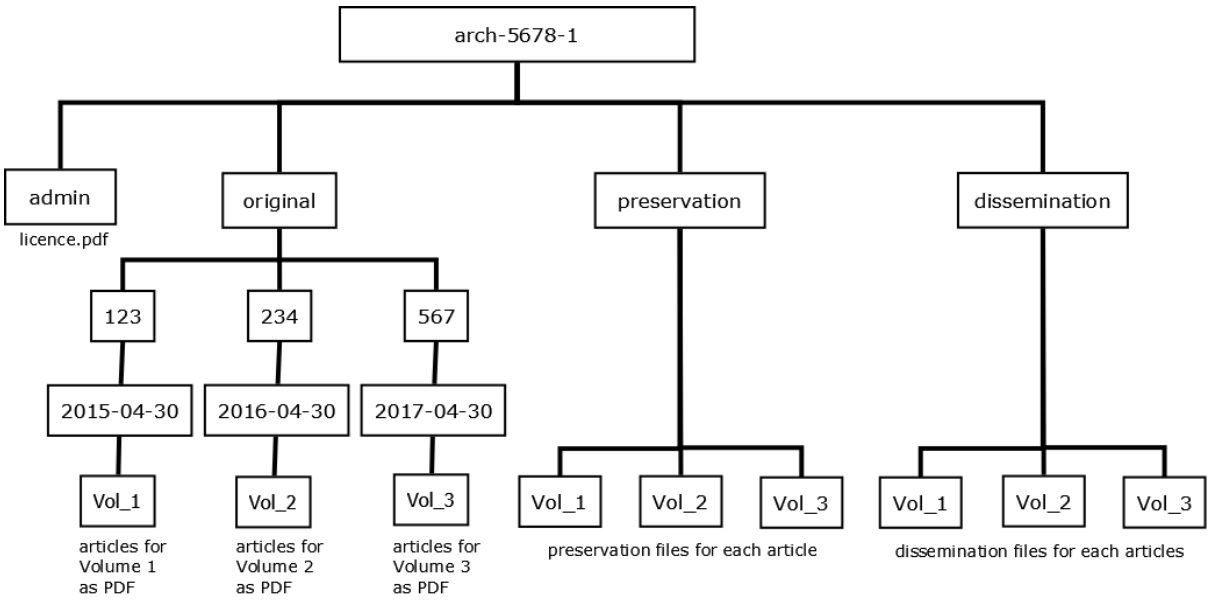
- In an email sent on 30th April 2016 the depositor clarified some queries about the database and gave permission to ignore one of the data tables that contained only test data so this email has been stored in /admin/project_metadata/. Documentation for the collection as a whole is stored in the /admin/project_metadata/ directory as well and may include a completed ADS project metadata template
- Database tables have been saved as CSV files under /preservation/{sub-folder}/ and /dissemination/{sub-folder}/. As CSV is both a preservation and dissemination format the files are stored in the /preservation/ folder and duplicated in /dissemination/.
- Likewise, the data dictionary and entity relationship diagram for the database are stored in the /documentation/ directory under /preservation/{sub-folder}/. These documents, where relevant, should be in suitable formats for the preservation.
- TIF images have been stored for preservation purposes under /preservation/{sub-folder}/ and batch converted to jpg images for dissemination (stored under /dissemination/{sub-folder}/)



Example 2: a run of journals (we receive new data on a yearly basis)

In this example, pdf files have been deposited for 3 volumes of a journal over the course of 3 years.

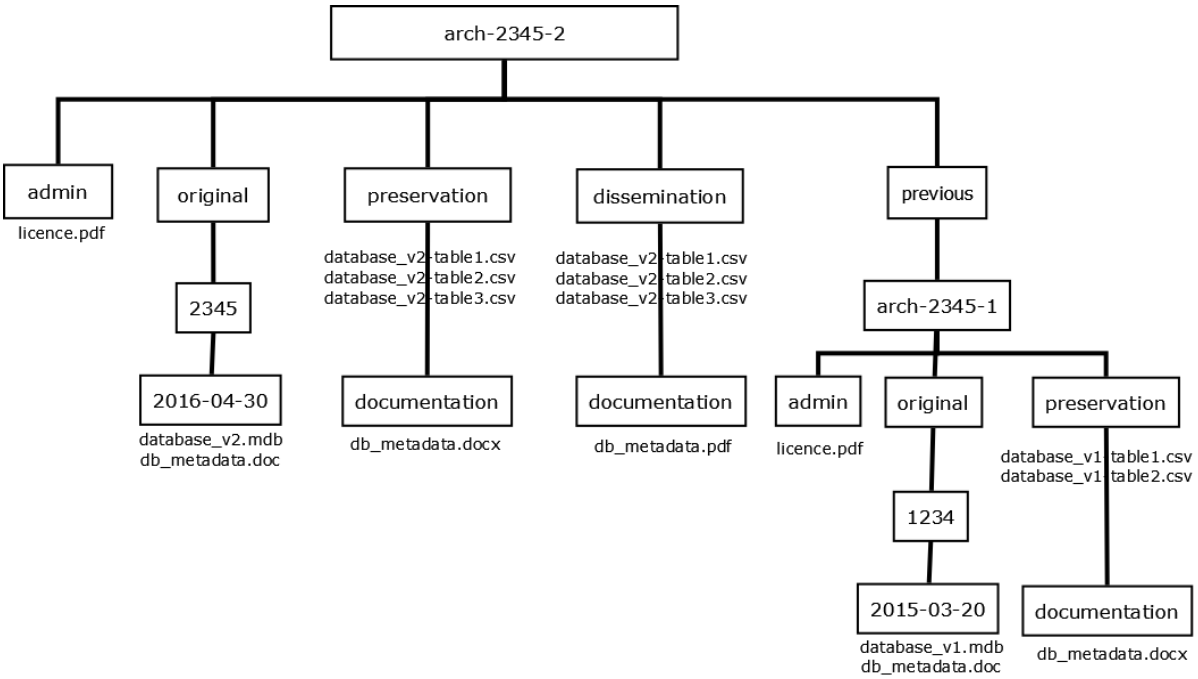
- 3 directories in /original/ store the 3 deposits of data (i.e. 123, 234, 567)
- Files are converted to preservation format and stored in /preservation/ under a volume number directory. Splitting files into separate volumes helps maintain order and structure
- PDF files are stored in /dissemination/vol_{no}/ for dissemination online
- The usual files are stored in /admin/.



Example 3: an update to a database (showing use of /previous/ directory)

In this example, a database (and documentation) has been deposited in 2015 and has been made available as a series of csv downloads. Subsequently in 2016, the final database containing updates and amendments was accessioned – this constituted a new edition of the archive

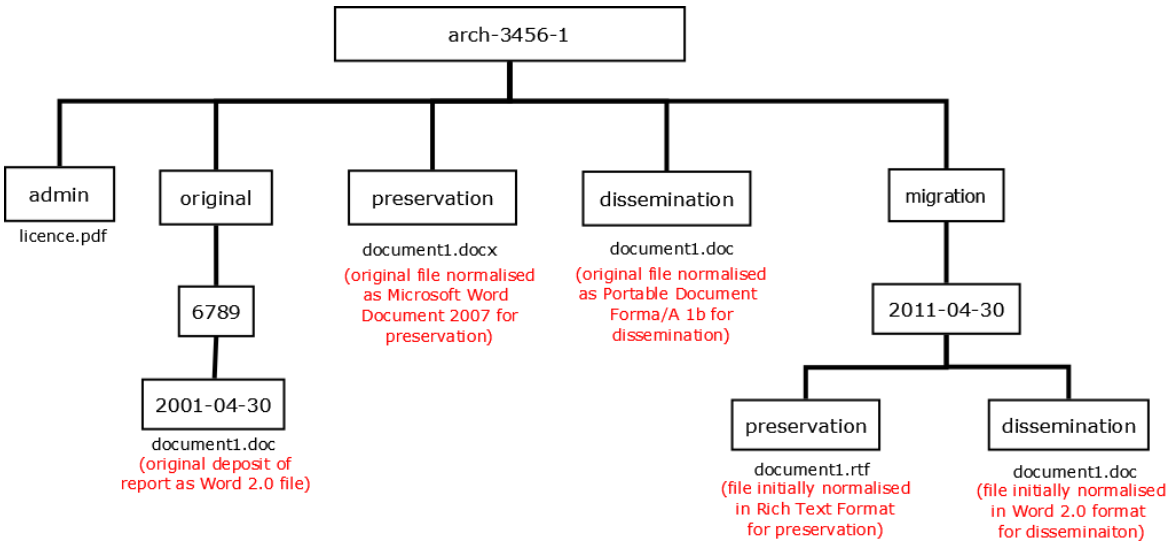
- 1st deposit: database is saved as csv files for preservation and dissemination and database documentation is stored in a /documentation/ directory
- 2nd deposit: the new and complete version of the database is converted to csv and stored under /preservation/ and /dissemination/ with the original documentation (still applicable as the data structure has not changed).
- The whole AIP is renamed to reflect the fact that it is now the second version.
- The old original, admin and preservation csv files (and the directory structure that they sit within) from the 2015 version are moved into the /previous/ directory, stored under the original AIP name (version 1)



Example 4: what happens when we carry out a file migration

In this example a selection of archaeological reports were accessioned in 2001. They were submitted as Microsoft Word 2.0 files and we disseminated them in their original format and preserved them as Rich Text Format.

In 2011 we decided that we needed to migrate both the preservation and the dissemination files into newer formats. Here is how the resulting directory structure would look.

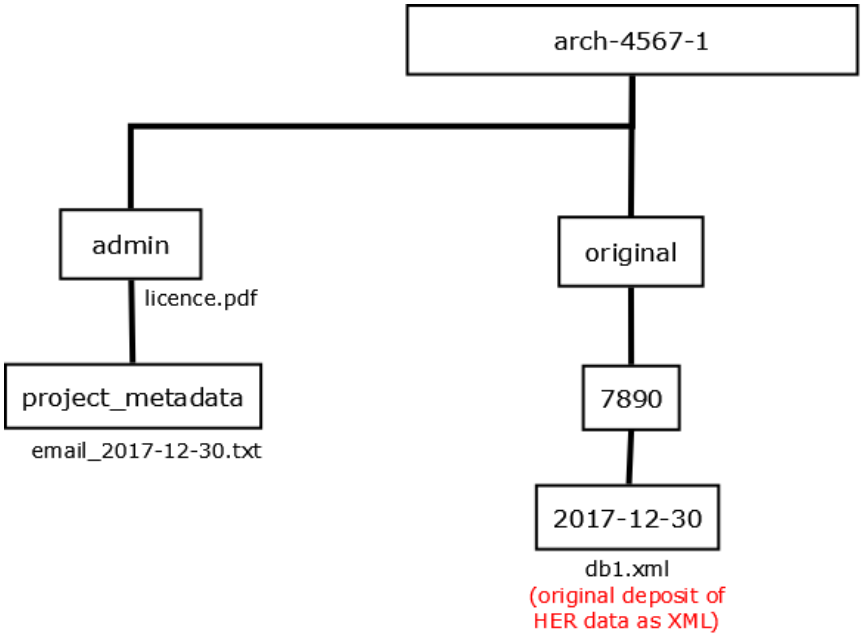


- Old DOC and RTF versions of the files are moved into /migration/ (stored under the date that the files were moved, and with the directory structure that shows the purpose and original data structure)
- New DOCX and PDF versions of the files are created for preservation and dissemination purpose and preserve the original data structure
- Metadata in the CMS/OMS is updated to reflect the changes

Example 5: a single deposit of an HER database

In this example we have received a database from a Historic Environment Record (HER). They have signed a licence which is different to our standard deposit licence and have agreed that the data is for dissemination through ArchSearch alone. The ADS therefore have no commitment to preserve this data.

- Beyond preparing the data and loading into ArchSearch no action is needed
- Data files used to load the database tables into Oracle need not be stored here
- /preservation/ and /dissemination/ data directories are not necessary



Appendix 1: Example file type directories under data directories

As of August 2017 the ADS no longer adheres to a prescribed directory structure, other than to store data underneath the appropriate /preservation/ and /dissemination/ directories. Beneath this the ADS preserves the meaningful structure given to the data by the depositor. Where the dataset lacks a logical structure or the extant structure proves problematic, the digital archivists can create a more appropriate structure. Changes to the directory structure require documentation in the *Collections Management System* (CMS).[15]

Historically, however, the ADS stored data underneath the /preservation/ and /dissemination/ directories in subdirectories relating to the file extension, and in some circumstances the data type.[29] A TIF file, for example, would be stored in a directory called /tif/, following the directory structure outlined below:

arch-{collections number}-{current version number} / preservation /
tif / {extant file structure} / {tif files}

(i.e. arch-335-1 / preservation / tif / {extant file structure} / {tif files}

The notable exception to this structure was where data was organised according to data type. In the case of an ESRI Shapefile (SHP), and its associated files, files were stored in a dedicated /gis/ subdirectory.

arch-{collections number}-{current version number} / preservation /
gis / {extant file structure} / {file name}.shp, {file name}.shx, {file name}.dbf

(i.e. arch-335-1 / preservation / gis / {extant file structure} /
mygis.shp, mygis.shx, mygis.dbf

Similarly, in the case of a geo-rectified TIF files, these were stored in a /geotif/ directory in order to differentiate them from 'standard' TIF files.

arch-{collections number}-{current version number} / dissemination /
geotif / {extant file structure} / {file name}.tif, {file name}.tfw

(i.e. arch-335-1 / dissemination / geotif / {extant file structure} / mygeotif.tif,
mygeotif.tfw

Historically this structure was adhered to afford easy management of datasets, but with the advent of the CMS and particularly the OMS, where more detailed information about file location and data type can be more effectively stored the need for this structure was negated.

The movement to the current file structure allows for increased consistency across the SIP, DIP and AIP, removing duplication of the structure within extension/data type subdirectories and allowing for greater simplicity in the data structure.

Appendix 2: Where to store files

The ADS currently has a split AIP with preservation files held in /ADS_preservation/ and dissemination files in /adsdata/. For security reasons these are spread across two virtual servers with different levels of access.

In order to avoid duplication, follow these guidelines when constructing directories in these two separate areas:

/ADS_preservation/ should contain:

- The /admin/ folder and all of its contents
- The /original/ folder and all of its contents
- The /preservation/ folder and all of its contents
- The /previous/ folder (where it exists) and all of its contents
- The /migration/ folder (where it exists) and all of its contents
- The only situation where the /dissemination/ directory may appear here are where there are dissemination files which are not available for download on-line (because of delayed release or large file size for example)

/adsdata/ should contain:

- The /dissemination/ folder and all of its contents
- The /migration/ folder (where it exists) and all of its contents

The structure should allow reconciliation of the complete AIP should it be necessary.

The creation of empty directories is discouraged. If there is nothing in a directory then it does not need to exist.

Appendix 3: Requirements for Scanned Hard Copy Material

The licence form, hard copy documentation and other supporting information must be scanned at, or above these minimum standards:

- Black and white documents, 200dpi 1 bit.
- Greyscale documents, 200dpi 8 bit.
- Colour documents, 200dpi 24bit RGB.

Scanned images must be saved as TIF v6.0 (pref) or PDF. Image dimensions should be adjusted according to the size of the scanned paper (A4, legal etc.). Illegible images should be rescanned at a higher resolution and/or colour depth.

If scanned images are saved as TIF, a separate file must be created for each page scanned rather than a single multi-page TIF file.

Appendix 4: Reserved File Names

A number of file names (such as licence.tif or licence.pdf) are reserved, and only used in specific circumstances.

Reserved file names can be modified. Any modifications should be placed after the main name of the file, and separated from it by a dash.

A reserved file name can be modified to include sequential numbers if there is a need to save the information in multiple files. For example, under /admin/, a three page hardcopy deposit licence scanned and saved as a series of TIF images could be named as follows: licence-1.tif, licence-2.tif, licence-3.tif.

A reserved file name may be further modified to include a date, in ISO 8601 format, if there is a need to distinguish between two versions of a file created on different dates that are both still current in some way (for example licence-2004-03-01.pdf) If both a date and a sequential number modifier are used, then files MUST be named as follows (using the example from above): licence-2004-03-01-1.tif, licence-2004-03-01-2.tif, licence-2004-03-01-3.tif.

Appendix 5: Data Types

At ingest each file is assigned a discreet data type appropriate to the type of data it contains. These are assigned programmatically using the file profiling software, DROID,[34] and the file extension. In circumstances where the file type and extension are not unique to a specific data type, then manual intervention may be necessary to ensure the correct identification.[38] Depositors using the OASIS Images[7] and ADS-easy[8] services identify the appropriate data type as part of the upload process. As discreet data types requires specific documentation and metadata, it is imperative that this identification is accurate.[39] Current data types include:

- 3D Model
- Audio
- Database
- Geophysics
- GIS
- Harris Matrices
- Image
- Laser Scanning
- LIDAR
- Mass Spectrometry
- Photogrammetry
- RTI
- Spreadsheet
- Statistics
- Text
- Vector
- Video
- Websites
- Admin[40]
- Documentation[41]

[38] A CSV file, for example, can contain Statistics, Database, Spreadsheet and Geophysical data and be classified as any of these data types. Similarly, a TIF file can simply be a raster 'Image', but it may also contain geographical information intended to rectify the image. The geo-rectified TIF image would be classified as GIS data.
[39] Details of these metadata requirements are outlined in the *Guidelines for Depositors* - https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml, with more extensive discussion within the ADS' Data Procedures (static versions of this documentation are available through the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures).
[40] The 'Admin' data type classifies administrative documentation, such as the deposit licence or deposit receipt, relevant to the collection.
[41] The 'Documentation' data type is use to identify both collection and file level metadata.

Appendix 6: PREMIS and the expression of relationships between objects.

The OMS structure allows the expression of non-hierarchical relationships between objects. Digital archivists will identify two files, or objects, and characterise the relationship between them using the PREMIS relationship type.[42] , created by the Library of Congress (USA). The current list of relationship terms is as follows:

- Is Documented In
- Is Source Of
- Has Source
- Includes
- Has Part
- Has Sibling
- Is Represented By
- Supersedes
- Has Version

Full definitions of these relationships are available directly from the Library of Congress, but those listed above are summarised below:

PREMIS_REL_TYPE	PREMIS Definition
Reference: "A relationship in which one object provides documentation for another."	
Is Documented In	A relationship between an environment object and the information that documents it. (The ADS uses this generally for digital objects, not environment objects).
Derivation: "A relationship in which one object is the result of a replication or transformation performed on the related object. The intellectual content of the resulting object is the same, but the object's instantiation, and possibly its format, are different."	
Is Source Of	… the related object is a version of this object created by a transformation, this is a derivation relationship, not a structural one.
Has Source	… the related object as a result of a transformation, this is a derivation relationship, not a structural one.
Structural: "A relationship between parts of an object. This is essential preservation metadata, because if a preservation repository can't put the pieces of a digital object back together, it hasn't preserved the object."	

[42] Published by the *Library of Congress* (USA) - http://id.loc.gov/vocabulary/preservation/relationshipType/collection_PREMIS.html.

Includes	… for the relationship of a representation to a file.
Has Part	A relationship in which the object is contained in the related object when these are the same object category. For instance, a Web page intellectual entity is part of a larger Web site intellectual entity.
Has Sibling	… the object shares a common parent with the related object (the ADS notes this when a parent object is not deposited).
Is Represented By	A relationship in which an abstract intellectual entity is represented as a file or representation.
Supersedes	A relationship between an environment object and another where the described object replaces another. This allows for an audit trail of environments to be maintained. (ADS can use this for digital objects not just environment objects.

Some examples of how each of these relationships is understood are given below:

Example 'parent' object	Relationship Type	Example 'child' object
Image1.tif	Is Documented In	image_metadata.csv
	Is Source Of	
GIS_files.zip	Includes	gis.shp (replicated for other files within the zip)
Spreadsheet.xlsx	Has Part	sheet1.csv (replicated for other tables within the same spreadsheet)
database_table1.csv (originally part of a database which has been deposited as separate csv files)	Has Sibling	database_table2.csv (replicated for other files within the original database)
	Is Represented By	
Report_2018.pdf	Supersedes	Report_2017.docx
RTI model	Has Source	Images used to create model