



Archaeology
Data Service

ADS Repository Operations Manual

Kieron Niven, Olivia Foster, Teagan Zoldoske, Jenny O'Brien, Ray Moore, Tim Evans

November 2022

Document Control Grid

Title:	ADS Repository Operations Manual
File name:	ADS_repository_operations_v7
Location:	Internal Drive
Status:	LIVE
Version:	7
Last updated:	04/11/2022
Created date:	01/06/2004
Review due:	01/09/2023

Authors:	Kieron Nlven, Olivia Foster, Teagan Zoldoske, Jenny O'Brien, Ray Moore, Tim Evans
Maintained by:	Tim Evans
Required Action:	None

Contents

Purpose of this document	4
Introduction	4
Conventions within this document	5
Ingest and the Submission Information Package (SIP)	5
Archival Storage and the Archival Information Package (AIP)	6
5.1 AIP Directory Naming	7
5.2 AIP Directory Structure	8
5.3 Permitted subdirectories	8
5.4 Original	9
5.5 Admin	11
5.5.1 Project Metadata	12
5.6 Data directories (preservation and dissemination)	15
5.7 Previous	19
5.8 Migration	20
6. Data Management	21
7. Administration	22
8. Preservation	22
9. Access	22
10. Deaccession	23
11. Bibliography	23
12. Examples of ADS directory structure	23
12.1 Example 1: a 'simple' one-off deposit	24
12.2 Example 2: a run of journals with new data received on a regular basis	25
12.3 Example 3: an update to a database (showing use of /previous/ directory)	26

12.4 Example 4: what happens when we carry out a file migration	27
12.5 Example 5: a single deposit of an HER database	28
12.6 Example 6: removal of data following deaccession	29
Appendix 1: Example file type directories under data directories	30
Appendix 2: Where to store files	32
Appendix 3: Requirements for Scanned Hard Copy Material	33
Appendix 4: Reserved File Names	34
Appendix 5: File naming policy	36

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".

The six mandatory responsibilities for an OAIS compliant archive are:

- Negotiate for appropriate deposits in accordance with the [Collections Policy](#) and the [Instructions for Depositors](#).
- 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 follow the workflow outlined in the [ADS Ingest Manual](#). The preservation of data in accordance with the guidelines given in the Ingest Manual 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 which provides data consumers and users with access to the DIP.

For example, see 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>.

3 Conventions within this document

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 is 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 methods:

- [OASIS](#)
- [ADS-easy](#)

- External data exchange service: These include, but are not restricted to, the University of York DropOff (file sharing) service, alongside other commercial file sharing services DropBox, GoogleDrive, etc. The ADS also utilises SFTP transfers with depositors where necessary.
- Exchange of digital media: The exchange of DVD, USB or portable hard-drives in person, or through the postal service.

The ADS supplies guidance on the selection of material for deposition, with more specific instructions on the deposition of personal, confidential and sensitive data. 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.

The ADS will accept SIPs that fit the requirements outlined in the **Collections Policy** and the Instructions for Depositors. This process of 'acceptance' is governed by the **Assessment and Appraisal Policy**.

Detailed information on the accessioning and ingest process is outlined in the **ADS Ingest Manual**. 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 Instructions for Depositors and other documentation (see below), are added to the AIP.

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 Metadata System (OMS), holds both technical and contextual metadata pertinent to the individual files.

In circumstances where depositors provide 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 the OAIS model. The AIP can also hold those preservation files normalised or 'migrated' to newer, more sustainable, formats following migration.

5.1 AIP Directory Naming

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

arch-`{collections number}`-`{version number}`

Example: arch-335-1

The collections number will be a unique identifier for the collection generated programmatically by the CMS.

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.

In instances where files, or data, have been normalised 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. Within the directory, superseded normalised files are stored following the original structure of the AIP.

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.

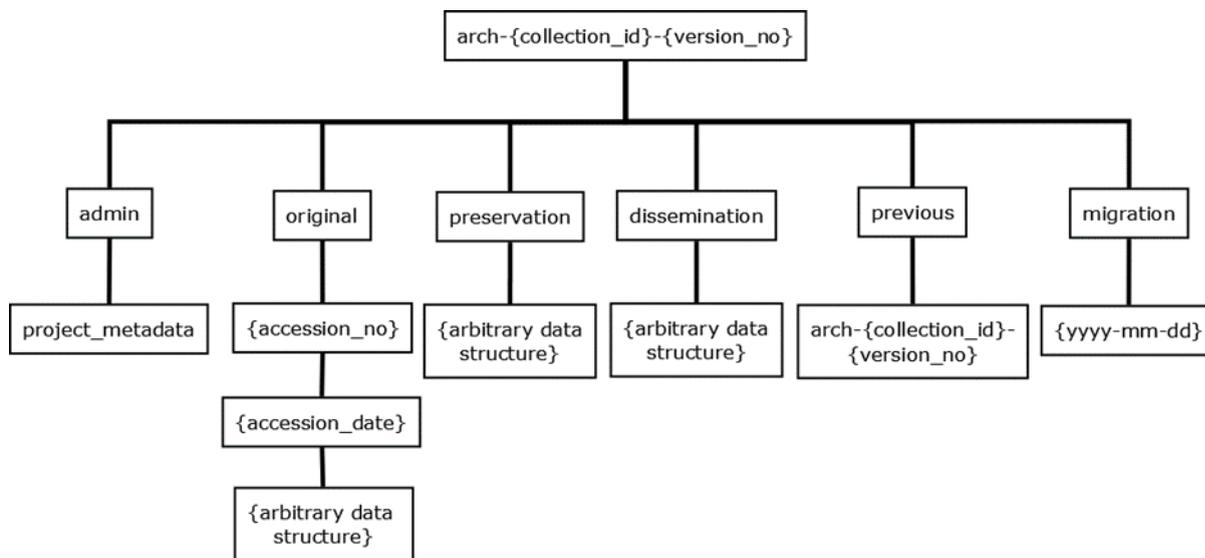


Figure 1: AIP directory structure

5.3 Permitted subdirectories

The AIP directory structure should only ever contain the following subdirectories which 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).
- **/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, into preservation formats.
- **/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 in 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.

5.4 Original

The /original/ directory should contain 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.
/ 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 form 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.
/ 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).

The repository **actively discourages changes to directory structure and file names 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 Instructions for Depositors. For those depositors submitting data

using OASIS Images or ADS-easy programmatic checks ensure that data follows the file naming policy (see Appendix 4 and Appendix 5).

However, in rare instances where the data structure is problematic or incongruous, ADS staff may rearrange a dataset into a more logical structure. All directories and folders within the SIP should be named in accordance with the file naming policy (see Appendix 4 and Appendix 5).

Where file naming does not follow the guidance provided, they will be renamed to follow ADS guidelines. The ADS also maintains a list of 'reserved file names' which are used for specific purposes within the AIP, a clash with a reserved file name will result in the deposited file name being changed in accordance with ADS guidelines. In all instances, the documentation of any changes made by the , within the CMS and OMS, is essential.

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}

Example: arch-335-1 / original / 770 / 2008-04-23 / {depositor's file structure}

5.5 Admin folder

The /admin/ directory should contain any documentation or files pertinent to the management and administration of the AIP. In October 2021 this was expanded to include the manifest

The /admin/ directory should contain the following files:

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 or when issued digitally stored in PDF form.	A signed (physical or electronic) deposit licence accompanies all depositions. All scanned licences and documents should follow the Requirements for Scanned Hard Copy Material in Appendix 3.

		<p>Previously the type of licence was documented in the file name alongside the CMS, but 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.</p> <p>Additional copies of this information are stored in the 'uploads' area of the CMS, and documented in the 'licences' section.</p>
manifest.txt	A list (with checksums) of all objects within the AIP.	Based on the standardised checksum file output, and can be read by other software.

5.5.1 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
email_{yyyy-mm-dd}.txt, or email_{yyyy-mm-dd}.pdf Additional terms, to notate the subject of the email, may also be included in filenames. This is useful in cases	Copies of any emails sent to, or received, from the depositor, or data producers.	Store only those mails which: <ul style="list-style-type: none"> - help document the SIP - give us permission to remove or edit files in the SIP - have been sent with original data to be accessioned into the archive

<p>where the list of emails is extensive, thus allowing repository staff to highlight the subject of the email, for example, email_deposit_{date_received}.txt (discussed below), email_copyright_{date_received}.pdf.</p> <p>'Notes' on email content should also be added to the record within the CMS, but this provides a similar function within the filestore.</p>		<p>- clear up copyright issues surrounding aspects of the data set.</p>
<p>deposit_receipt_{accession_id}.csv [deprecated, but reinstated]</p>	<p>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.</p>	<p>For a number of years this was 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 should also be attached to the CMS record ('Upload' section).</p>
<p>Project metadata files (e.g. ADS_collection_metadata.docx)</p>	<p>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</p>	<p>If the documentation/metadata relates to actual data, whether a group of files or a particular file, this should be stored in a /documentation/</p>

	<p>'file naming policy' with any file name changes documented in the CMS.</p> <p>Any metadata or documentation held in the /preservation/ directory should not be duplicated within the /admin/ directory.</p>	<p>directory alongside the data it refers to.</p>
<p>Introduction and Overview text (e.g. introduction.docx, or overview.txt)</p>	<p>Introduction and Overview text for the web interface should be stored here.</p>	<p>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.</p>
<p>email_deposit_{date_received}.txt, or email_deposit_{date_received}.pdf</p>	<p>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.</p>	<p>This file is also attached to the CMS record ('Upload' section).</p>
<p>dc_metadata.txt [deprecated]</p>	<p>A TXT file containing the Dublin Core metadata record loaded into ArchSearch and exported from the CMS.</p>	<p>No longer actively undertaken, may exist for older archives. The metadata continues to be collected, but is stored in the CMS.</p>
<p>original_{accession_id}.txt [deprecated]</p>	<p>A TXT file containing checksums for all files within the original deposition.</p>	<p>No longer actively undertaken for current collections, but may exist for older archives.</p>

		Partially replaced by a more extensive deposit receipt which includes checksum information for the deposition.
data_log_{date_of_deaccession}.csv	A digital copy of the data log created on the deaccessioning of part of, or an entire, dataset.	This file is also attached to the CMS record ('Upload' section).

It is important that all files within the /admin/ directory are stored in a suitable preservation file format.

The admin directory should have a structure much like this:

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

Example: arch-335-1/admin/project_metadata/

5.6 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.	On the implementation of the OMS, which provides file location and data type information, the storage in specific directories became unnecessary, and, consequently, deprecated.

<p>/ {file type} / {arbitrary} / [deprecated] / {arbitrary} / [current]</p>	<p>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</p>	<p>Any significant changes to directory structure are recorded in the 'processes' section of the CMS.</p> <p>Previously the {arbitrary} directory structure, inherited from the original deposition, was only permitted within the {file type} directory.</p>
<p>/ {documentation} / [current] (e.g. /preservation /documentation/)</p>	<p>Used to store documentation, or metadata, that relates to the entire collection, or groups of files within the collection.</p>	<p>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.</p>
<p>/ {file type} / {arbitrary} / documentation / [deprecated] / {arbitrary} / documentation / [current] (e.g. /preservation/{arbitrary} /documentation/)</p>	<p>Used to store documentation, or metadata, that relates to the entire collection, or groups of files within the collection.</p>	<p>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).</p>

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.	<p>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.</p> <p>These images are not part of the AIP/DIP, except in instances where the files are specifically identified as data</p>

		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}

Example: 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}

Example: 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}

Example: arch-335-1 / preservation / documentation / {depositor's file structure}

And/or:

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

Example: arch-335-1 / preservation / {depositor's file structure} / documentation

5.7 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:

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

Example: arch-335-2 / previous / arch-335-1/ {extant file structure}

5.8 Migration

The /migration/ directory should contain 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.

		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:

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

Example: 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 CMS, OMS and the AIP. The repository provides downloadable metadata templates and advice, through the [Instructions for Depositors](#), to ensure that both collection and file-level documentation are adequate.

Original templates, submitted as part of SIP, are stored alongside the data within the AIP and disseminated as part of the DIP (see Section 5). Documentation pertinent to the collection is stored within the admin/project_metadata directory, while data/file specific metadata is stored in the /documentation/ directories of the /preservation/ and /dissemination/ directories of the AIP.

These files are normalised to stable preservation and dissemination formats in line with current guidelines. Typically collection level metadata is disseminated through a dedicated metadata page for each collection.

In instances where additional or enhanced documentation is submitted within the SIP, this can be 'attached' to the collection record within the CMS. This metadata is also preserved and disseminated in a suitable format as part of the AIP.

Digital submissions of data using OASIS or ADS-easy allow Depositor to record collection metadata within an online form. This information is then imported, and stored, within the CMS and OMS. File specific documentation is uploaded, and metadata is stored and disseminated in the same way as standard collections.

The CMS and OMS are also used 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 that supports the continued management of datasets. 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 files and data sets throughout the preservation life cycle, and as defined in the [ADS Preservation Policy](#).

7. Administration

Administrative information about the AIP is stored within a combination of the CMS, 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`) and included within the collection record within the CMS. Where this documentation does not fit within the database structure it can also be 'attached' to the record.

Administrative documentation and metadata does not form part of Dissemination Information Package (DIP), however in certain circumstances, and with the agreement of the data provider, it may be necessary to share this information through the archive interface.

8. Preservation

The [ADS Preservation Policy](#) and Ingest Manual discuss the repository function with regard to preservation. The [ADS Data Procedures](#) provide specific detail on the processes and practices employed by the repository. The CMS and OMS document all actions performed on the AIP.

9. Access

The [ADS Preservation Policy](#) provides detailed information on the repository function with regard to the dissemination and access to datasets. The [ADS Data Procedures](#) provide specific detail on the processes and practices employed by the repository. The CMS and OMS should be utilised to document all actions performed on the DIP.

10. Deaccession

The [ADS Deaccession Policy](#) provides clear and concise instruction on the processes and procedures in place for the partial or the complete removal of a dataset from the repository. In such instances, the CMS and OMS records all actions and processes performed on the AIP and DIP during the removal of the dataset.

11. 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>, accessed 05 October 2021.

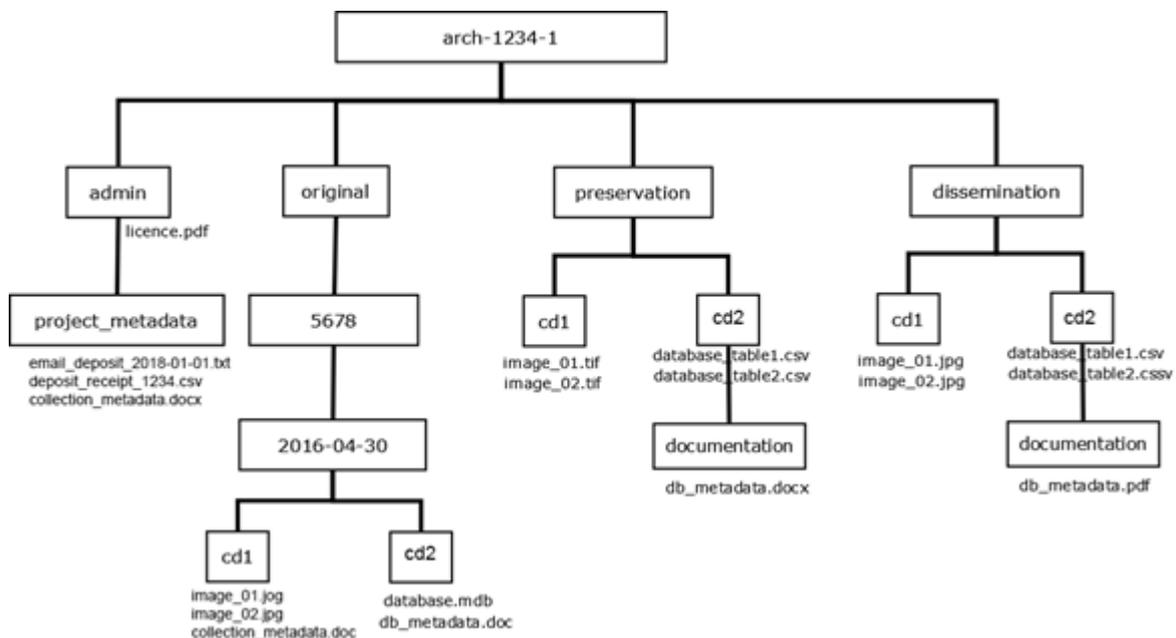
12. Examples of ADS directory structure

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:

12.1 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.

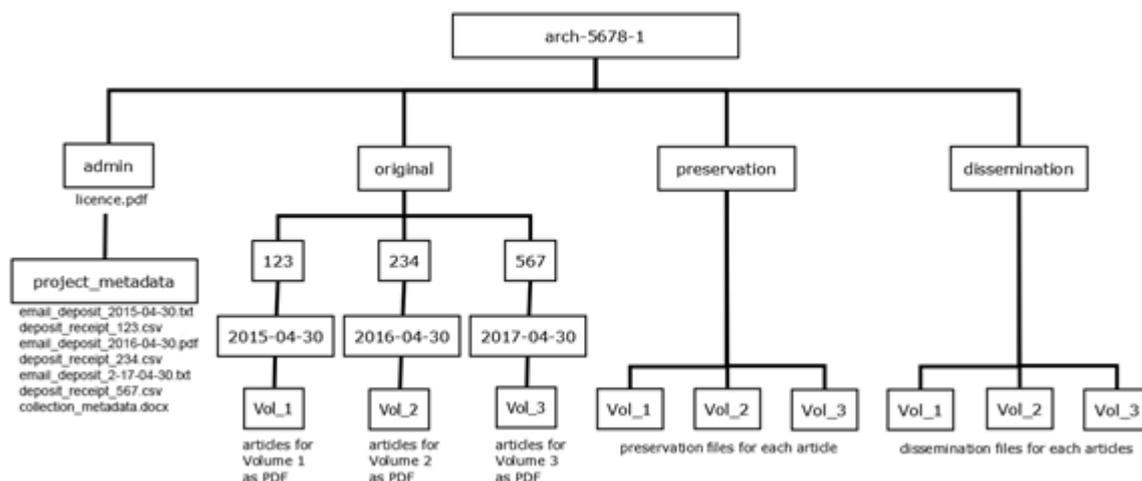
- In an email sent on 30th April 2016 the depositor clarified some queries about the database and gave permission to ignore one of the data tables that contained only test data so this email has been stored in /admin/project_metadata/. Documentation for the collection as a whole is stored in the /admin/project_metadata/ 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}/)



12.2 Example 2: a run of journals with new data received on a regular basis

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

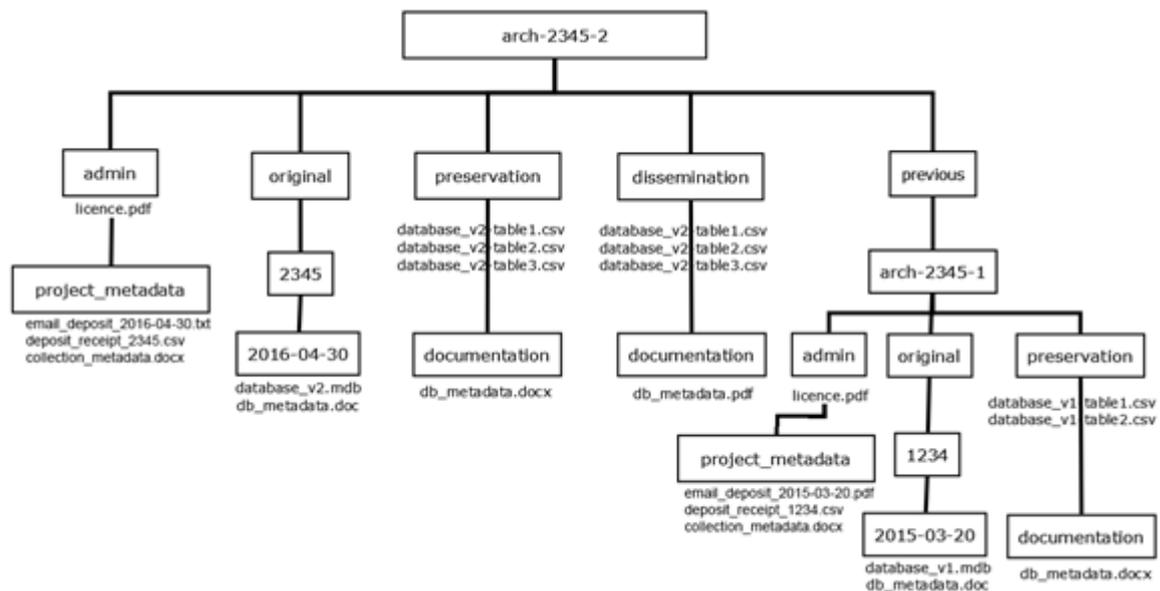
- 3 directories in /original/ store the 3 deposits of data (i.e. 123, 234, 567)
- Files are converted to preservation format and stored in /preservation/ under a volume number directory. Splitting files into separate volumes helps maintain order and structure
- PDF files are stored in /dissemination/vol_{no}/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



12.3 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).

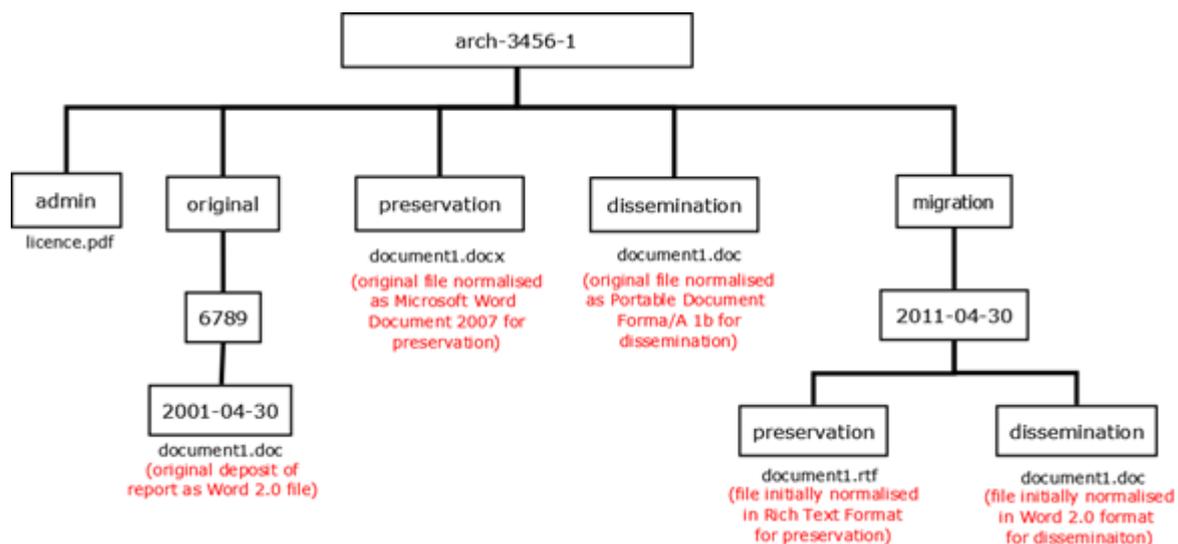


12.4 Example 4: what happens when we carry out a file migration

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.

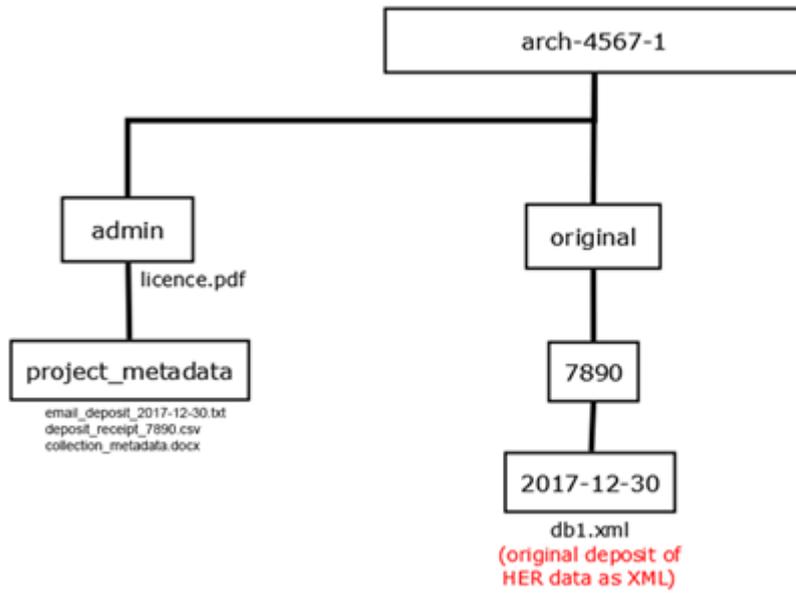


12.5 Example 5: a single deposit of an HER database

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 has 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.

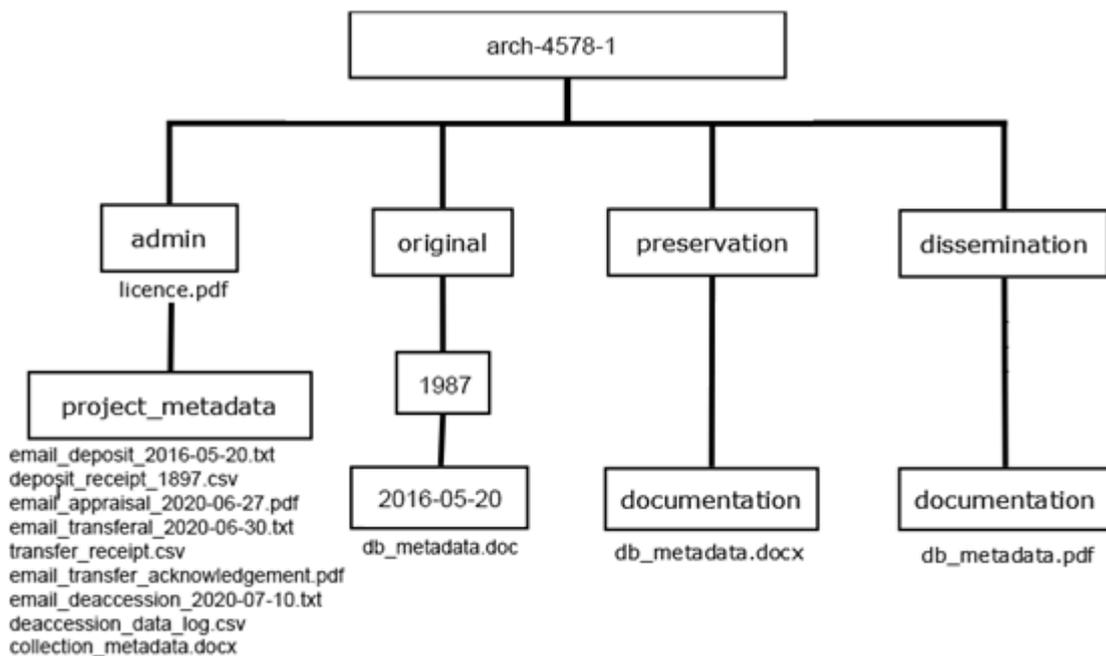
- /preservation/ and /dissemination/ data directories are not necessary.



12.6 Example 6: removal of data following deaccession

In this example following a request from a depositor to transfer the digital archive to another repository, an appraisal by ADS staff and agreement with the destination archive reached. The Deaccession and Removal Policy provides complete documentation of this process.

- Thorough documentation of the appraisal and deaccession process are required, both within the CMS and within the data store.
- Documentation of all data transferred to the destination is required.
- The retention of a complete data log of the entire dataset.
- Retention of ALL metadata to enable clear documentation of files removed (particularly for DOIs).



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 CMS.

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. 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}**

example: 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**

**example: arch-335-1 / preservation / gis / {extant file structure} /
mygis.shp, mygis.shx, mygis.dbf**

Similarly, in the case of 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**

**Example: 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 not permitted**. If there is nothing in a directory then it does not need to exist.

Appendix 3: Requirements for Scanned Hard Copy Material

A3.0.1 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.

A3.0.2 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.

A3.0.3 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

A4.0.1 The ADS maintains a small list of reserved file names used only in specific circumstances.¹

Term	Description
licence.pdf / licence.tif	A signed (physical or electronic) deposit licence accompanies all depositions. Note this uses UK spelling for the noun.
email_{yyyy-mm-dd}.txt	Copies of any emails sent to, or received, from the depositor, or data producers.
email_deposit_{yyyy-mm-dd}.txt	Copy of email issued to acknowledge the receipt of the dataset. This email accompanies the deposit receipt. Where possible this should include acknowledgement and agreement from the depositor that the dataset is complete and as described in the deposit receipt.
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
dc_metadata.txt	[deprecated] A TXT file containing the Dublin Core metadata record loaded into ArchSearch and exported from the CMS.
original_{accession_id}.txt	[deprecated] A TXT file containing checksums for all files within the original deposition.
arch-{collections number}-{version number}.xml	[deprecated] An export of collection level metadata from the CMS.
arch-{collections number}-{version number}_gpmd.xml	[deprecated] An export of Generic Preservation Metadata Database (GPMD) from the CMS.

¹ Names such as licence.tif or licence.pdf.

checksum.txt	[deprecated] Checksums for the entire AIP.
--------------	--

Table A4.0.1: A list of reserved file names.

A4.0.2 These reserved file names can be modified, but any modifications should be placed after the main name of the file, and be carried out in accordance with the guidelines below:

- Use of sequential numbers.
 - e.g. Multiple emails are received by the depositor on the same day
email_{yyyy-mm-dd}_1.txt, email_{yyyy-mm-dd}_2.txt
 - e.g. A three page hardcopy deposit licence scanned and saved as a series of TIF images
licence_1.tif, licence_2.tif, licence_3.tif
- Use of date (in ISO 8601 format)
 - e.g. Multiple emails are received from a depositor on different days
 - email_2020-06-14.txt, email_2020-06-15.txt, etc.
 - e.g. Different 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: File naming policy

A5.0.1 File names of deposited data should follow the following guidelines:

- File names should use only alpha-numeric characters (a-z, 0-9), the hyphen (-) and the underscore (_). No other punctuation or special characters should be included within the filename.
- Use the underscore character to imply a space within your file name.
- A full stop (.) should only be used as a separator between the file name and the file extension and should not be used elsewhere within the file name.
- File extensions are normally 3 characters long and should be lower case.
- Both upper and lower case characters can be used in a file name but keep files within your project consistent and ensure that supplied documentation accurately reflects the case of your filenames. Also remember CAPITALS ARE HARD TO READ and affect ordering.
- Individual file names, regardless of file structure, should be unique within a dataset.
- Keep file names consistent. Descriptive or non-descriptive file names can be used. A descriptive file name helps explain the contents of the file. For example: TSA04_final_report_v3.pdf (version 3 of final site report for site TSA04 as a pdf file), or 12102004_trench_1.tif (digital photograph of trench 1 taken on 12/10/2004). A non-descriptive file name might be a unique ID number allocated to an image within an accompanying image catalogue database. Non-descriptive file names are acceptable as long as content is adequately described in accompanying metadata.

A5.0.2 SIPs that contain files that do not follow these naming conventions should be updated as during accession, as outlined in the Ingest Manual,⁵ following the guidelines outlined in the table below:

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	
^	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_	

Table A5.0.2: Guidance on renaming files submitted as part of the SIP.