Ingest Manual

Version 5.0

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

1. Purpose of this document

This document outlines the process used to accession and ingest data submitted to the Archaeology Data Service. This includes information on the workflows for submissions from ADS-easy, OASIS Images, OASIS and other, external digital exchange services. Alongside these, the ADS continues to accept data through the exchange of physical media. Whilst ingest is broadly similar for each of these discreet workflows, there may be subtle differences which are documented below. Once accessioning into the repository is complete, the workflow for all submissions is broadly consistent.

2. Introduction

The ADS receives data from depositors through a series of workflows[1], via the

- ADS' dedicated submission portals – ADS-easy[2] and OASIS Images[3]
- OASIS[4] (grey literature reports)
- the digital exchange of data, outside of the dedicated ADS portals[5]
- the submission of data on physical media[6]

Once the *Submission Information Package* (SIP) is received the data enters the ADS workflow which is essentially a two-stage process, involving the

[1] https://archaeologydataservice.ac.uk/advice/DepositingData.xhtml#How%20to%20Deposit
[2] https://archaeologydataservice.ac.uk/easy/
[3] https://oasis.ac.uk/form/images.cfm, access only to registered users. See Hardman, C (2014) *OASIS: new image upload facility.* OASIS Blog: Archaeology Data Service. https://archaeologydataservice.ac.uk/blog/oasis/?p=41
[4] https://oasis.ac.uk/, see
[5] Where possible the ADS advocates the use of the *University of York Drop-off Service* (https://www.york.ac.uk/it-services/services/dropoff/), although GoogleDrive, DropBox, etc. are also utilised.
[6] For example, DVD, USB-drive, external hard drive, etc.

- accessioning of the data, and the
- preparation of the *Archival Information Package* (AIP) and *Dissemination Information Package* (DIP) for preservation and dissemination

The accessioning of data begins as soon as possible after data receipt in order to identify issues that might affect the preservation and dissemination of the dataset and bring these to the attention of the depositor. Typically, the depositor addresses any issues, with the assistance of repository staff, and returns any replacement files or updated metadata. The issuing of the deposit licence follows; once 'signed', the collection is kept in a 'holding pattern' until time can be allocated to the collection within the workflow. The second stage of the archiving process involves the preparation of data for inclusion in the AIP and DIP, the creation of the archive interface, and the documentation of all preservation processes and activities.

The ADS' *Guidance on the Selection of Material for Deposit and Archive* provides assistance for those unsure about what an archive should contain[7]. The *Guidelines for Depositors* outline the collection level documentation and file level metadata requirements for each deposition.[8] Specific information on *accepted and preferred formats* is available from the ADS website.[9] If additional information or assistance is required, depositors are encouraged to contact the ADS.[10]

The ingestion process for all SIPs received by the ADS should occur as a priority as soon as possible after receipt of data. Since 2014, the ADS has also accepted deposition of data through its ADS-easy[11] and OASIS Images[12] portals. These services give depositors the ability to create metadata and upload data directly to the ADS. Deposition through these portals follows the same basic ingest processes as that for standard depositions, however, online deposition of data is subject to its own discrete process and workflow (incorporated below). Checklists for both standard and ADS-easy accessions are available to guide digital archivists through the process,[13] while this *Ingest Manual* provides additional and specific guidance. Only when all parts of the checklists are completed can work begin on the creation of the AIP and DIP.

[7] https://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml
[8] https://archaeologydataservice.ac.uk/advice/DepositingData.xhtml#Depositing%20with%20the%20ADS
[9] https://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml
[10] https://archaeologydataservice.ac.uk/about/contact.xhtml
[11] http://archaeologydataservice.ac.uk/easy/
[12] https://oasis.ac.uk/form/images.cfm, access only to registered users.
[13] Detailed checklists are available within the ADS wiki - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=CmsManualHomePage, internal access only. Static versions of both the standard and ADS-easy accession checklists are available through the ADS website (https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists).

3.Data Accession

3.1 Data transfer session

As noted above, the ADS accepts submissions through a variety of methods.[14] This can include the digital transfer of data and metadata, or the exchange of physical media. The transfer of data deposited through the ADS' dedicated submission portals is largely programmatic. The exchange of data digitally, outside of the dedicated portals, or through the submission of physical media, requires the manual movement of data; first to a local drive, where it can be assessed and reviewed, before it is transferred to the preservation servers. All internal data transfers use the SFTP protocol.

3.2 Pre-accession checks

All accessions are subject to assessment during accession, but in some circumstances, a depositor may submit a dataset for review prior to formal accession. This allows the repository to appraise the dataset for any problematic formats, or incomplete metadata. This approach is particularly useful when the dataset is large, or where the depositor has not deposited a data set with the repository previously. To facilitate this assessment data is stored, temporarily, in the ADS file store.[15] Checks carried out during this assessment include, but are not limited to

- all formats are suitable for deposition
- the completion of appropriate collection and file level metadata
- the dataset is accurately documented
- an appropriate licence has been agreed with the depositor (although this may not be issued)
- that funding for the archive has been arranged and any charges documented in the CMS
- that details for the depositor, or other technical contact, are recorded
- any additional information, impacting the creation of the AIP and DIP, or the dissemination of the dataset, has been documented in the CMS

3.3 Check file formats are suitable for deposition

The ADS' *Guidelines for Depositors*[16], and specifically the list of *Preferred and Accepted File Formats*[17], provide depositors with guidance on data types and formats that are appropriate for submission. The identification of problematic formats or data types typically occurs during negotiation for deposition. The Collections Development Manager and Digital Archivists can provide support and guidance to depositors to ensure that submissions follow these requirements. Data submitted through ADS-easy, OASIS Images and OASIS is subject to

[14] See the How to Deposit Data section of the ADS website - http://archaeologydataservice.ac.uk/deposit/How.xhtml.
[15] In the 'Unaccessioned holdings' directory.
[16] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml.
[17] http://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml

programmatic controls that restrict submissions to the formats appropriate to the data type. Manual checks are made of all SIPs submitted digitally, or through the exchange of physical media, to ensure that they adhere to the repositories policies and guidelines. Replacement files are sought directly from the depositor where problems are noted. In those circumstances where resubmission is not possible, the ADS will accept the data in the extant format, but under the proviso that preservation activities will be carried out on a 'best efforts' basis.

3.4 Virus check

In accordance with the *Security Overview* and the policies provided by the University of York IT Services, all PCs are protected using anti-virus software and firewall.[18] The repository follows the guidance provided by the University of York IT Services in order to mitigate against potential infection. The use of anti-virus software and malware ensures that the SIP is virus-free before any transfer of data to local hard drives or servers. Those deposits made through ADS-easy, OASIS Images and OASIS, are submitted to programmatic virus-checking during the upload to the ADS. Additional checking accompanies all data transfers between systems and servers. The initiation of manual checks, as outlined in the *Security Overview*, are performed for data submitted digitally, outside of these systems, or through the exchange of physical media.[19]

The ADS follows the policies and guidelines outlined in the *Security Overview*[20] and the University of York IT Services in instances where viruses or malware are noted.[21] All infected data is isolated and disinfected. Should a problematic file be identified, a 'clean' copy of the file, or data, would be requested from the depositor.

3.5 Media and file readability check

All media and data submitted to the ADS is scrutinised to ensure that it has not become corrupted prior to or during transmission. In circumstances where the submitted dataset is small, it is often possible to open and examine all files within the collection to ensure that they are readable. Where the SIP is large, or particularly complex, this approach will not be possible. As a result, the appraisal of a representative sample of the dataset will take place, with care taken to ensure that the evaluation includes examples from all data types and formats. Checks also seek to identify password protection or encryption that adversely affect preservation and dissemination pathways.[22] Checks will vary according to the data type and format submitted, but the ADS' data procedures provide guidance on this.[23] The preparation of both AIP and DIP allows further opportunity for assessment of data from collections.

---

[18] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security, see also https://www.york.ac.uk/it-services/security/virus/.
[19] Anti-virus software carried out 'on-access' scanning, but manual checks are also carried out.
[20] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security.
[21] https://www.york.ac.uk/it-services/security/virus/#tab-3
[22] Appendix 1 provides information on the checks carried out by the ADS during accession.
[23] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of these are available on the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.

Replacement files are sought directly from the depositor where problems are noted. These replacement files then form a new accession to the collection, and subject to all the checks listed here, with the unreadable/problematic data removed from the dataset.

As an additional check, the ADS produces a 'data receipt' that lists all files received by the repository (see below). This data receipt includes a MD5 checksum for each file submitted, allowing depositors to identify any problems that may have occurred during data transfer.

3.6 Documentation and integrity check

The *Guidelines for Depositors*[24] provide depositors with assistance on the required documentation for all datasets, and should include appropriate collection and file-level metadata. The ADS provides metadata templates, in a variety of formats, to assist depositors with the creation of metadata[25], alongside guidance on these requirements. This documentation must accompany data submitted to the repository.

Where depositors use the ADS' submission portals (ADS-easy or OASIS Images) metadata is submitted through a series of online forms, or file uploads, which allow some validation of metadata, while additional checks are also carried out manually during accession. For those datasets submitted digitally, outside of these systems, or through the exchange of physical media, manual authentication of documentation and metadata is required. Where the deposition includes quantitative documentation, such as directory or file lists, additional checks ensure parity between this documentation and the data received. Such checks ensure the identification of any missing files at the earliest opportunity.[26] Digital archivists will also carry out qualitative checks on all documentation to ensure accuracy. Where these checks identify inaccurate or incomplete metadata, the digital archivist will notify the depositor and advise on addressing these concerns.

The collection level metadata that accompanies the SIP will include an introductory text used in the archive interface. Where necessary a depositor can include additional information in the form of an 'overview' that is also added to the archive interface. It is also useful to have some image(s) to illustrate the archive where appropriate. Digital archivists will contact the depositor where this information is missing.

It is the responsibility of depositors, as per the terms of the deposit licence, to ensure that all data submitted to the repository adheres to current legal and ethical guidelines. The ADS provides additional guidance on the deposition of personal, confidential and sensitive data allowing depositors to plan for the submission of data.[27] Where digital archivists identify data that they believe infringes this policy, concerns should be directed towards the depositor and, where necessary, the Collections Development Manager. Clarification should also be

[24] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml.
[25] https://archaeologydataservice.ac.uk/advice/Downloads.xhtml#Downloads
[26] As an additional check, the ADS produces a 'data receipt' that lists all files received by the repository (see below). This data receipt also includes a MD5 checksum for each file, allowing depositors to identify any problems that may have occurred during the data transfer.
[27] https://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml

sought from the depositor where queries over data ownership, or concerns over copyright infringement, are noted.

3.7 Record details of SIP in Collection Management System

Once assessment of the SIP is complete, digital archivists should document the accession using the ADS' *Collection Management System*, or CMS.[28] This should record the date of accession and depositor; the creation of a digital checklist should be included to track the collection as it moves through the ADS workflow.

3.8 Standardise file names and check directory structure

Depositors should follow the guidance provided in the *Guidelines for Depositors* with regard to file naming.[29] For depositions submitted via ADS-easy or OASIS Images, stringent controls have been put in place to ensure that depositors adhere to the file naming policy while uploading data to the repository. For other depositions, manual checks ensure adherence to the naming policy. Digital archivists ensure that all files have been checked and, where required, altered in accordance with the file naming policy. Where changes have been necessary these require documentation within the CMS.

The repository does not make any specific stipulations with regard to directory structure used within the SIP. However, the ADS does ask that "a logical file structure" is used that "allows data to be easily retrievable".[30] Data that is poorly structured may make preservation activities difficult, and will certainly affect the dissemination of data and restrict reuse. It may be necessary, therefore, for digital archivists to make changes to the data structure. If this proves necessary, it is important to record all changes within the CMS.

3.9 Copy to data server

Once data assessment is complete and any necessary file name and structural changes have been carried out, the digital archivist should copy the SIP to the ADS data preservation server. All file data transfers use a dedicated client, and follow the SFTP protocol. Where data is submitted through either ADS-easy, or OASIS Images, they will be stored on a dedicated server until accession; at which point data will be transferred to the preservation server during the automated ingest process initiated in the CMS. Data transfers use the SFTP protocol, with checksum validation to ensure the successful transfer of files.

[28] http://adsmanticore0.york.ac.uk:8080/cms3/pages/home.do, internal access only.
[29] https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#FileNaming and the *Repository Operations* (https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp). Current and more detailed information on file naming is also available in the ADS wiki - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=FileNamingPolicy, internal access only.
[30] https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#File%20Management

3.10 Create checksums and technical metadata

The ADS uses the National Archives' file characterisation application, DROID, to create technical metadata.[31] This metadata is used to populate the *Object Management System* (OMS) - physical location, filename, size, format, MIME type, PRONOM identifier and, most importantly, checksum/fixity value.

3.11 Issue licence and create data receipt

All depositors are required to 'sign' a deposit licence, issued from the CMS as part of the accession process.[32] It is important to ensure that the details in the deposit licence are correct, and completed according to the information listed in the collection metadata. Where appropriate the 'organisation' for whom the individual is signing the deposit licence is completed. The, non-exclusive, licence gives the repository permission to disseminate data on behalf of the depositor; copyright for the data is not transferred by this agreement. The deposit licence also outlines the terms of re-use for the dataset.[33] The ADS provides a summary of the deposit licence on the website.[34]

A signed deposit licence is required before archival work can commence.

The signed deposit licence, once returned, should be stored, alongside the data, as outlined in the *Repository Operations*.[35] A copy of the licence should also be stored in the CMS.

3.12 Attach correspondence to CMS

Deposition of data can sometimes be a protracted processes requiring negotiation with the depositor. It is important that any correspondence is preserved and attached to the collection record in the CMS. Any correspondence pertinent to the preservation or dissemination of the dataset should also be stored alongside data following the structure outlined in the *Repository Operations*.[36] All correspondence should adhere to the *File Naming/Renaming Policy*[37] (for data storage) and conventions used for adding documentation to the CMS.[38]

3.13 Update the changes log

Any new SIPs, alongside any changes or updates, require documentation in the 'changes log'.[39] This provides an additional check during the syncing process with external storage.

---

[31] http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/
[32] https://archaeologydataservice.ac.uk/resources/attach/ADS_Deposit_Licence_2018.pdf
[33] https://archaeologydataservice.ac.uk/advice/reuseLicence.xhtml
[34] https://archaeologydataservice.ac.uk/advice/depositLicence.xhtml
[35] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp
[36] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp
[37] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=FileNamingPolicy, internal access only.
[38] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=CmsManualNegotiations, internal access only.
[39] https://docs.google.com/spreadsheets/d/1mTr1IzGZvqDqrUQn4t_gMO5hMWIw7NyYPto4_WZQYq8/edit#gid=1085367917, internal access only.

3.11 Scan paper documentation

While the ADS operates a paperless office, on occasion depositions may include paper or physical copies of licences, documentation, or other metadata. In these circumstances, the scanning and storage of all paper-based documentation should follow guidance outlined in the ADS wiki[40] and *Repository Operations*.[41]

3.12 Acknowledge receipt of data and issue deposit receipt

Once the accession is complete, digital archivists should acknowledge the receipt of all data by email. A data receipt, generated in accordance with the guidance in the ADS wiki and *Repository Operations*, should be included.[42] The email should include any issues or queries noted during accession, with the depositor given opportunity to respond and address them. In accordance with the *Repository Operations*, a copy of the data receipt and associated email should be stored alongside the dataset, with another copy added to the CMS. All relevant details pertinent to the accession should be stored in the CMS.

3.13 Store original media

Once an accession is complete, physical media will require labelling with the correct collection number and any accession ids. The media should then be stored in the collections filing cabinets. Where the depositor has requested that original media be returned to them (this may happen if data has been delivered on a memory stick, or portable hard drive) this should be returned to them only after the backup to deep storage has been carried out. In circumstances where data has been deposited electronically through ADS systems (ADS-easy or OASIS Images) or external file-sharing services, these should not be deleted until the backup to deep storage has been actioned. A note is added to the CMS to document that media has be returned to the depositor.

4. Preparation of the AIP and DIP for preservation and dissemination

Once accession is complete, the digital archivist will begin the process of normalising data, and the creation of both AIP and DIP. A dedicated checklist, available through the ADS wiki, guides digital archivists through the archiving process.[43] Additional, more detailed, documentation is available in the ADS wiki and this *Ingest Manual*. Only once preservation and dissemination activities are complete, and the creation of both AIP and DIP finished, will the process of AIP checking begin.

---

[40] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=ScanningGuidelines, internal access only.
[41] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp
[42] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DepositReceipt, internal access only. See also *Repository Operations* https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[43] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=Archiving%20procedures, internal access only. A static version of the checklist is available through ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists.

4.1 Consistency checks

Though a certain amount of checking and validation of the delivered data will have occurred at the accessioning stage, it is sometimes the case that previously undetected issues may arise once we start working with the data and preparing the AIP and DIP. The nature and complexity of consistency checks depends on the data type under scrutiny. Appendix 1 provides information on the checks carried out, while the ADS' data procedures[44] includes detailed documentation of all checks. Where the digital archivist identifies issues with either the AIP or DIP, the depositor is encouraged to re-submit replacement data in order to maintain data integrity and authenticity. Where necessary, digital archivists can also undertake edits on behalf of the depositor, although any updates should follow the guidance in the ADS' data procedures.[45] All such changes require documentation in the CMS as a 'process'.

4.2 Develop a conversion plan

Digital archivists are encouraged to develop a conversion plan for all archives. If normalisation into the preservation or dissemination format is complex, it may be necessary to map out each step, ensuring that the appropriate software and expertise is available. The data procedures provide guidance on standard preservation and dissemination pathways for the vast majority of cases; however, it may be necessary to consult with colleagues, or the Archives Manager, over the most appropriate course of action for your dataset. While many normalisations are relatively straightforward, involving single-step conversions, others might involve multi-stage processes using a variety of different software packages. In instances where the dataset is large, or the number of files significant, batch processing may provide a reliable and consistent approach to normalisation. Digital archivists should ensure that all normalisations are consistent with the guidance outlined in the data procedures.

4.3 Selecting preservation and dissemination file formats

The ADS' data procedures, available internally, provide guidance on appropriate formats for preservation and dissemination.[46] With reference to these procedures, and the content of the archive as documented in the CMS and OMS, a suitable preservation and dissemination pathway is established. Data submitted in formats suitable for preservation or dissemination may not require normalisation; in these circumstances, it is important to consider the preferred version of the format for inclusion in the AIP and DIP. It is should be noted that in other instances data may be preserved and disseminated in different formats. All data normalisations should ensure that the significant properties of the data are preserved.

[44] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only.
[45] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only.
[46] Static versions of these procedures are available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.

4.4 Conversion of files

Once the conversion plan is established, the digital archivist should follow the plan and enact all normalisation activities. The creation of a local version of the SIP will allow digital archivists to carry out normalisations, and others changes to the dataset, without affecting the integrity of the original deposition. All changes to the dataset should be in accordance with the conversion plan (see 4.3) and the data procedures.[47]

4.5 Validate file conversion

Digital archivists should ensure that the normalisation of data, alongside any other changes, are successful and that the significant properties of each file remain unchanged. As noted above, it may not always be practical to check each individual file within a dataset, but it is important to check a representative sample of the dataset.

4.6 Copying files to the preservation/dissemination directories

Once all normalisation processes are complete and the resultant files validated to ensure the preservation of all significant properties, the transferal of data from local directories to the preservation and dissemination servers can begin. All file data transfers use a dedicated client, and follow the SFTP protocol. Fixity checks should ensure that transfers of data have been successful. Digital archivists should ensure that data is stored in accordance with original directory structure and as outlined in the *Repository Operations*.[48]

4.7 Update the changes log

Any transferal of data to the preservation servers requires documentation in the 'changes log'.[49] This provides an additional check during the syncing process with external storage.

4.8 Create technical metadata

See Section 3.10. Once copied to the ADS servers the use of DROID, the file characterisation application, allows the creation of technical metadata for all files and normalised data within both the AIP and DIP. [50] The transferal, and storage, of all metadata to the OMS is a semi-automated process initiated through the CMS.

---

[47] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of the data procedures are available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.

[48] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

[49] https://docs.google.com/spreadsheets/d/1mTr1IzGZvqDqrUQn4t_gMO5hMWIw7NyYPto4_WZQYq8/edit#gid=1085367917, internal access only.

[50] http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

Digital archivists use the 'match objects' functionality, initiated from within the CMS, to link all related files, and data, into notional 'objects'.[51] Typically matches are made programmatically using the automated 'computer matching' function within the CMS. Additional relationships between files and objects can be added/updated manually. In instances where the deposition is particularly large, then the CMS will be unable to support the 'printing' of all files/objects to interface, consequently, 'match objects' can also be run from outside of the CMS.

All files within the collection should have been assigned a 'data type'[52] as part of the accession process (outlined in Section 3.10), or, in the case of depositions through the ADS-easy and OASIS Images portals, created at upload. At the same time, the creation of AIP and DIP may result in the creation of new files and 'objects', it may, therefore, be necessary to assign new data types to these outputs. Digital archivists should assess all data types to ensure their accuracy. Where necessary functionality provided within the CMS allows digital archivists to 'update' the data type where it has been incorrectly identified.[53] Digital archivists should ensure that those files containing collection, or file level, metadata have been ascribed the correct data type.[54] Other files, such as the deposit licence or deposit receipt, may have an internal and administrative function and digital archivists should ensure that the OMS documents the appropriate data type.[55]

To facilitate the internal management of data, digital archivists should also ensure that the correct 'resource type' is allocated.[56] This allows greater granularity in characterising the content of files and objects. The CMS allows the assignment of the correct resource type.

Digital archivists should ensure that all relationships between files, or 'objects', are documented within the OMS where the nature of that relationship, or 'relationship type', can also be expressed.[57]

[51] Documentation of the 'match objects' process can be found within the ADS wiki - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=Match%20Objects%20Overview, internal access only.
[52] See Appendix 2.
[53] The system, for example, may have problems differentiating between files containing data and documentation, or identifying the correct data type when the format is not unique. A 'csv', for example, may contain metadata or documentation, but can equally contain geophysical, spreadsheet, or database data.
[54] The data type for files containing metadata or documentation relating to the dataset is 'documentation'. See Appendix 2.
[55] The data type for files holding other documentation relating to the preservation of the collection is 'Admin'. See Appendix 2.
[56] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=ResourceTypes, internal access only and Appendix 4.
[57] The ADS uses an agreed list of 'relationship types' documented in the ADS wiki - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=Object%20Relationships%20type, internal access only, see also Appendix 3, all have been mapped to the equivalent 'PREMIS Relationship Type' (http://id.loc.gov/vocabulary/preservation/relationshipType/collection_PREMIS.html).

4.9 Add and/or update collection and file level metadata

Collection, and file level metadata is required for **all** depositions to the ADS. The *Guidelines for Depositor* provided guidance on the appropriate metadata required for each data type[58], while checks during the accession process ensure that data meets the required standards of documentation (see Section 3.5).

In instances where data is submitted through ADS-easy or OASIS Images, metadata is completed through a series of online forms. The applications provide some quantitative checking of metadata during upload, with qualitative checks made as part of the accession process. The transferal of this metadata to both the CMS and OMS respectively forms part of the accession process for these submissions[59], however Digital Archivists should always carry out checks to ensure accuracy and completeness. The identification of gaps, or problems, form part of the accession process, and where necessary updates should have already been sought directly from the depositor. In instances where the identification of new/other issues with collection or file level metadata are noted, or where extant problems have not be addressed to the satisfaction of the Digital Archivists, contact should be made with the depositor and updates to the metadata sought. In some circumstances, Digital Archivists may carry out updates or enhancements to metadata as part of the archiving process; any such changes require documentation, as a discrete process, within the CMS.[60] In circumstances where documentation has been submitted through the online forms within ADS-easy or OASIS Images metadata will be stored within the CMS and OMS. While it is not necessary to extract any metadata from the OMS for inclusion in the AIP for preservation reasons, as standard practice all file level metadata is exported from the OMS and stored (in a separate file) within the DIP. The inclusion of this downloadable version of the file level

[58] https://archaeologydataservice.ac.uk/advice/Downloads.xhtml#Downloads.
[59] See Section 3.1.
[60] See Appendices 5 and 6.

metadata within the archive interface provides data users with the necessary documentation to facilitate reuse.[61] All extracted metadata should be stored, in accordance with the *Repository Operations*,[62] alongside any data, and disseminated in appropriate formats.[63] Digital Archivists should ensure the correct classification of all extracted file level metadata with an appropriate data type (i.e. 'documentation'), while at the same time documenting all relationships between the metadata and data.[64] All collection level metadata is directly available through each archive interface;[65] therefore, any extraction from the CMS is not necessary.

The submission of the SIP digitally, outside of ADS-easy or OASIS Images, or through the exchange of physical media, requires collection level and file level metadata to be deposited as a separate file. Typically, using the dedicated metadata templates.[66] In these instances, and as standard practice, Digital Archivists should transfer this metadata into the CMS and OMS respectively. A dedicated spreadsheet metadata loader, provided through the CMS, allows staff to add raster file-level metadata directly from the spreadsheet into the OMS, but for other data types, the transferal of file-level metadata is a largely manual process. While the assessment of all metadata for completeness and accuracy forms part of the accession process, the transferal of metadata can lead to the identification of additional problems. In these circumstances, the notification of the depositor should follow and, where possible, updates to the metadata provided. In some circumstances, it may be necessary for Digital Archivists to address problems with the metadata, but any such updates require documentation in the CMS. In other cases, Digital Archivists may wish to enhance metadata; any such enrichment requires documentation in the CMS. Following transferal of all documentation from the templates to the OMS, checks ensure parity between the metadata stored in the two locations. All completed metadata templates should be stored, in accordance with the *Repository Operations*,[67] alongside any data, and disseminated in appropriate formats.[68] Digital archivists should make sure a record of the appropriate data type (i.e. 'documentation') is part of the CMS,[69] with all relationships between the metadata and data recorded.

For all types of submission, depositors may upload additional or supplemental documentation, beyond the required standard metadata, as discreet files. The accession of

[61] The extraction of the metadata from the OMS and the creation of the downloadable file are documented within the 'processes' section of the CMS.
[62] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[63] See the data procedures - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of these procedures are available through the website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.
[64] See Section 4.6.
[65] For example, https://archaeologydataservice.ac.uk/archives/view/redhill_he_2019/metadata.cfm.
[66] https://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level%20Metadata%20Requirements.
[67] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[68] See the data procedures - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of these procedures are available through the website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.
[69] See Appendix 2.

additional metadata should follow the workflow outlined above for all depositions.[70] All supplemental files are subject to the same qualitative checks carried out for all metadata during accessions. All supplementary documentation will require normalisation in accordance with the data procedures. Attribution with the correct data type is required for the supplemental metadata (i.e. 'documentation'), while all relationships between the documentation and data should be defined in the OMS. Any supplemental metadata should form part of the AIP and DIP, and disseminated alongside the associated data through the archive interface.

4.10 Record conversion and editing processes undertaken

All conversions, or processes, carried out on the dataset should be documented in the 'Processes' section of the CMS. A semi-automated system allows digital archivists to generate most of these programmatically following file matching (see Section 4.8). This process makes assumptions about the nature of the conversion, so all processes require checking to ensure accuracy. Digital archivists should carry out any edits, updates, or add additional information through the CMS interface. In some cases, it may be necessary to add processes manually. The CMS documents the type of process carried out[71], and includes a detailed record of each process.[72] Digital archivists should document any problems or additional information within the comments section of each process.

4.11 Interface creation

Following the creation, and documentation, of both AIP and DIP, a dedicated interface for each collection is required. Guidance on this is available from the ADS wiki, [73] and in consultation with the Collections Development Manager. For standard deposition the ADS uses a series of templates and exemplars to facilitate this process. These templates can be adapted and amended where required. Once completed the digital archivists should check that the interface meets the required standard in terms of accessibility, validation and compatibility.[74] All pages created and templates used require documentation in the CMS.[75]

All interfaces, once complete, are shared with the depositor to ensure that they meet the required standard that has been agreed. The circulation of these web pages to repository staff ensures that they adhere to the guidelines (noted above) and that they function as intended. All collection interfaces are subject to more rigorous checking as part of the AIP check.

---

[70] See Section 3.1.
[71] See Appendix 5.
[72] See Appendix 6.
[73] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=CreateArchiveInterface, internal access only.
[74] This documentation is available through ADS wiki - http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=CreateArchiveInterface, internal access only.
[75] In the Web Admin section, http://adsmanticore0.york.ac.uk:8080/cms3/authenticate.do, internal access only.

4.12 Submit AIP for checking[76]

On completion of both the AIP and DIP, AIP checks ensure that the outcomes meet the requirements outlined in the *Ingest Manual*, *Repository Operation*[77] and data procedures.[78] ADS staff carry out the visual inspections of all archives prior to release, with more in-depth assessments carried out by an assigned digital archivist following the completion of preservation and dissemination activities. Historically, all archives had AIP checks carried out, but as workload has increased, it has become impossible carry out the assessments for all archives. In cases where the collection is relatively simple, i.e. it contains a small number of raster images and some metadata, formal AIP checks may not be necessary and the visual inspection of the archive, carried out by all ADS staff, should highlight any issues. This is often the case for archives submitted through the ADS-easy or OASIS Images portals. The repository still regards AIP checking as an important part of the preservation process, particularly as it ensures that the correct adherence to the recommended archival procedures and policies. Instances where AIP checking may be necessary, include, but are not restricted to, the following:

- a collection is large, or particularly complex
- the digital archivist is unsure whether preservation and dissemination activities have been completed successfully
- the archive includes 'new' data types or formats
- digital archivists are undergoing training, or to ensure digital archivists are consistently applying current policies and procedures
- the Archive's Manager, or other staff, feel checks may be necessary

Collections requiring AIP checks are marked within the CMS, and the Archives Manager allocates the task to a Digital Archivist. A dedicated checklist provides guidance on the assessment carried out as part of the AIP checking processes,[79] with additional information provided in the *Ingest Manual*, *Repository Operations*[80] and data procedures.[81] Once complete, any problems or issues with the AIP and DIP are highlighted and appropriate action taken to address them. As noted above, checks of the archive interface also form part of the AIP process.[82]

At the same time, the sharing of the archive interface with the depositor allows their input into the archive interface. Any comments or requests by the depositor should be addressed

[76] Although the term AIP check implies assessment of the AIP only, these checks also consider the DIP, the archive interface and correct documentation of the entire archive.
[77] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[78] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only.
[79] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=AIPChecking, internal access only, although a simplified version is available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists.
[80] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[81] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of the data procedures are available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.
[82] See Section 4.9.

where appropriate and within reason. The submission of archives for sign-off by the Archive Manager, or Collections Development Manager, follows.

4.13 Archive release

Following sign-off by the Archive Manager/Collections Development Manager, the archive is ready for release. An agreement for the date of release is sought from the depositor. The repository does allow depositors to request an embargo date for those archives that contain sensitive information, or to allow full publication of any outputs.[83] Following appraisal by the depositor, and completion of the proper AIP checks, the removal of all pages and files of embargoed content from the production server ensures the protection of any associated data. In instances where an advanced DOI is required, for publication say, the creation and continued support of a landing page may be necessary. Once an embargo has passed, the return of all pages and files to the production server, and the necessary updates to DOI, ensures the dataset is publically accessible.

Detailed documentation for the release process is available in the ADS wiki[84], although a shorthand version is available within the *Procedure Checklist*.[85] A final check is made of the archive to ensure that all processes have been documented, metadata completed and all 'objects' created. Part of the release process involves the minting of the persistent *Digital Object Identifier* (DOI) for the collection.[86] All releases are publicised through the ADS' 'Collections History' page[87] and resource discovery metadata transferred into the *ArchSearch* catalogue.[88] OAI-PMH targets allow the ADS to share resource discovery metadata with other catalogues and organisations.[89] The ADS Library allows the publication of citations for reports, documents and articles included in the collection.[90] Digital archivists also take steps to ensure the documentation of any relationships between digital resources within the CMS, and outside the repository, are added and up-to-date. Once released the repository publicises the publication of the collection through its website and social media.

---

[83] See the *Collections Policy* - https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml.
[84] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=Cats, internal access only.
[85] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=Archiving%20procedures, internal access only. A static version is available from the ADS website https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists.
[86] DOI's are minted through the British Library (https://www.bl.uk/), part of the DataCite (https://datacite.org/) consortium. The creation of DOIs for individual files/objects and groups of objects may be agreed with depositors in some circumstances. In either instance, the creation of DOIs forms one of the last parts of work on an archive. See the Preservation Policy for a fuller discussion - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol.
[87] https://archaeologydataservice.ac.uk/about/collectionsHistory.xhtml.
[88] https://archaeologydataservice.ac.uk/archsearch/basic.xhtml. A fuller discussion of the ADS' qualified Dublin Core (DC) metadata is available in the Preservation Policy https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol.
[89] https://archaeologydataservice.ac.uk/advice/OAIPMH.xhtml. For a fuller discussion of these repositories and organisations, see the Preservation Policy - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol.
[90] https://archaeologydataservice.ac.uk/library/.

5. Updating content

In some instances, collections may require the submission of additional or replacement data. Such submissions can be part of a regular/planned program of updates, or a series of piecemeal/irregular additions to an existing collection. The ADS supports these activities in an effort to provide up to date and accurate datasets.

5.1 Adding data to a collection

When a depositor wishes to add further data to an existing collection, the digital archivist should create a new accession within work on AIP and DIP, progressing as they would with a standard 'new' deposition.[91] Where necessary depositors should submit additional collection and file level metadata so that any documentation is current and accurate. Digital archivists should make sure that the 'last modified date' and, particularly, the 'updated year' fields are updated within the CMS so that it that the collection history is clear to both staff and data consumers. The addition of data to an existing collection should not require the minting of a new DOI for the collection.

5.2 Replacing part/all of a dataset

Updates involving the replacing of data are little more complex as depositors/users may require access to previous versions of part/all the dataset. Following the OAIS model[92] the ADS retains all previous versions of a file and editions of a dataset. The creation of a new edition of the entire dataset follows deposition of updated material, with all previous versions of files and editions of the dataset stored in accordance with the *Repository Operations*.[93] This is a semi-automated process initiated through the CMS. Digital archivists should ensure that the 'last modified date' and, particularly, the 'updated year' fields are updated within the CMS so that it that the collection history is clear to both staff and users. The treatment of all replacement or updated data should follow standard practice for all new accessions, and are subject to the same workflow identified in the *Ingest Manual* (see above), *Repository Operations*[94] and the ADS' data procedures.[95]

The retention of all editions of the dataset, and versions of individual files, as part of the AIP is standard; the DIP, however, should consist of the most current edition of the dataset only. Each edition of the dataset should have its own dedicated DOI, but only the most recent version of the DIP should be active. The maintenance of previous DOIs should continue,

[91] Depositions of new, or replacement, data for an existing collection cannot be made through the ADS' dedicated deposition portals and must be received through an external digital transfer, or through the exchange of physical media.
[92] https://www.dpconline.org/knowledge-base/preservation-lifecycle/oais
[93] The *Repository Operations* provides specific, technical guidance on edition/version handling in terms of the structuring of the dataset. See https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, specifically Example 3.
[94] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[95] http://adslocalwiki0.york.ac.uk:8080/wiki/Wiki.jsp?page=DataProcedures, internal access only. Static versions of some of these are available on the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures.

with dedicated landing pages created to ensure that they resolve and provide notification to users about the succession of editions. All previous editions of the dataset are available to users on receipt of a formal request.

6. Appendix 1: Data validation and consistency checks

Here are some examples of the types of checks which may be carried out as a part of the ingest process. This list has been derived from the *AHDS Archive Ingest Procedure Framework: HS Preservation Procedures Manual*, working draft 1.3 prepared by Raivo Ruusalepp, Estonian Business Archives Ltd, December 2002/January 2003.[96]

- Check that digital resources and their items adhere to the relevant formal definitions of their structure (e.g., an XML document conforms to its XML schema, a relational database conforms to its SQL schema, an image conforms to its stated image format – dpi, colour depth, compression, etc.).
- Image compression algorithm, dimensions, orientation, resolution, colour space, etc. correspond to the values stated in documentation.
- Digital audio compression algorithm, length of the recording, sampling frequency, bit rate, etc. correspond to the values stated in documentation.
- Digital video compression algorithm, length/duration of the recording, codec structure, frame rate, sound format, etc. correspond to the values stated in documentation.
- Linkages and dependencies between items within a particular type of digital resource should be checked for correctness (e.g., in a database, foreign keys having a matching primary key; in a spreadsheet, formulas refer to correct cells, etc.).
- Linkages and dependencies to other digital resources are correct (e.g., hyperlinks point to a currently valid URL, details of published works in a bibliography are correct, etc.).
- Items within a digital resource adhere to the relevant definition (e.g., a numeric field in a database contains a number, text strings do not exceed a stated maximum length, etc.).
- Items within a digital resource contain 'sensible' values that do not contradict relevant logical assumptions (e.g., age of a person should not be less than 0) and subject/resource type specific concerns
- Documents (word processor files) should be checked for changes or errors in footnotes, tables of contents, links, auto-fields and formatting that may hinder the later use of the data resource.
- GIS, CAD and virtual reality data resources may require domain- or research area specific consistency checks to be applied (e.g., scale of different layers in a GIS, level of precision and sufficiency of coordinates in a CAD and VR data, etc.).
- Simple data types (numbers, text strings, dates, etc.) are not truncated, restricted in range, formatted or otherwise defined in a potentially confusing or ambiguous way (e.g., dates contain four digits for the century, date format, memo fields in a database do not contain embedded end-of-lines, etc.).
- Coded data must be checked that the data have been consistently assigned the documented code.
- Any codes that are used in data must be used consistently and according to the specified coding rules.
- Standardised data has been standardised consistently and according to specified rules or a recognised schema for the standardisation.

[96] This document in now no longer available.

- Exceptions to particular standards, coding schemes, formats, etc. are documented and justified in the documentation for the data collection

7. Appendix 2: Data types

The ADS uses dedicated data types to classify digital objects (files) within its *Object Management System* (OMS)[97]. These help identify the type of data that is stored and is particularly useful when the type of data contained within the file is unclear from the file extension. A CSV file, for example, stores Spreadsheet, Database or Geophysical data. These data types provide a shorthand to ensure that digital archivists receive the correct metadata for each file.

- 3D Model
- Database
- GIS
- Geophysics
- Harris Matrices
- Image
- LIDAR
- Laser Scanning
- Mass Spectrometry
- Photogrammetry
- RTI
- Spreadsheet
- Text
- Vector
- Video
- Websites

Added to this classification are ‘non-data’, data types. These classify internal documentation generated to document a collection and metadata created by the depositor.

- Admin
- Documentation

[97] Internal access only.

8. Appendix 3: Relationship types

The ADS adheres to the PREMIS concept of a relationship, i.e. "a relationship in which one object provides documentation for another".[98] Agreed ADS Relationship Type List:

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

PREMIS_REL_TYPE	PREMIS Definition
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)
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
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

[98] http://id.loc.gov/vocabulary/preservation/relationshipType/collection_PREMIS.html

9. Appendix 4: Resource type

To facilitate the internal management of data the repository uses a discreet 'resource type', based on the *FISH Resource Description Thesaurus*[99], which characterises the content of data. These add granularity to the 'data type' classification.[100]

- Correspondence
- Diary
- Notebook
- Oral History Transcript
- Publication
- Thesis
- Site Record
- Report
- Provisional/Working Report
- Ephemera

[99] http://heritage-standards.org.uk/wp-content/uploads/2016/05/Thred_class_v20.pdf
[100] See Appendix 2.

10. Appendix 5: Process type

Digital archivists should ensure **all** processes and actions carried out on a dataset have been correctly documented within the CMS. All preservation and dissemination activities are classified with the appropriate 'process type'.

- Capture
- Compression
- Creation (Documentation)
- Creation (Metadata)
- Deletion
- Editing (Aesthetic)
- Editing (Corrective)
- Migration (Dissemination)
- Migration (Preservation)
- OCR
- Rename
- Restructure
- Other Event

11. Appendix 6: Process documentation

Digital archivists ensure that all processes carried out on the dataset, during the creation of both the AIP and DIP, are recorded correctly within the CMS. The form within the application uses the following criteria to document each action.

Type	See Appendix 9 above
Source Format	The original format of the data.
Destination Format	The resultant format of the data
Start Date	When the process was begun
Completion Date	When the process was completed
Description	The nature of the process being carried out
Result	Success/Partial Success/Failure
Input	The files on which the process is carried out on
Output	The files resulting from the process
Hardware	PC/SUN/Mac
Software	Any software used to carry out the process
Operating System	Operating system of the hardware (e.g. Windows 10)
Comments	Any comments from the agent carrying out the process
Agent	The individual carrying out the process
Accession ID	The accession id for the data.