INGEST MANUAL (VERSION 6)

ARCHIVES MANAGER AND DIGITAL ARCHIVISTS
ARCHAEOLOGY DATA SERVICE
https://archaeologydataservice.ac.uk/

Created date:	2004
Last updated:	28 July 2020
Review Due:	31 July 2021
Authors:	Ray Moore, Kieron Niven, Olivia Foster, Tim Evans, Digital Archivists
Maintained by:	Archives Manager, Digital Archivists
Required Action:	
Status:	Live
Location:	https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml
Previous version	https://archaeologydataservice.ac.uk/resources/attach/ADS_Ingest_Manual_v5.pdf

1. Purpose of this document

1.0.1 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 discrete 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

2.0.1 The ADS receives data from depositors through a series of submission streams[1] that allow the exchange of data, alongside technical and contextual metadata. Depositors use one of the following repository deposition streams:

- *OASIS*[2] [3]
- *OASIS Images*[4] [3]
- *ADS-easy*[5] [3]
- external data exchange service[6]
- exchange of digital media[7]

2.0.2 On receipt of the *Submission Information Package* (SIP), data enters the ADS workflow via a two-stage process involving:

- accessioning of the data (see Section 3)
- preparation of the *Archival Information Package* (AIP) and *Dissemination Information Package* (DIP) for preservation and dissemination (see Section 4)

2.0.3 The ADS website provides detailed repository-specific information to depositors on the preparation and submission of the SIP through the *Guidelines for Depositors*.[8] They also

[1] https://archaeologydataservice.ac.uk/advice/DepositingData.xhtml#How%20to%20Deposit, accessed 15 June 2020.
[2] http://oasis.ac.uk/pages/wiki/Main (requires registration) accessed 15 June 2020.
[3] Deposition through these streams 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. These are discussed below.
[4] A raster image service available through the OASIS system - http://oasis.ac.uk/pages/wiki/Main (requires registration), accessed 15 June 2020.
[5] http://archaeologydataservice.ac.uk/easy/ (requires registration to use), accessed 15 June 2020.
[6] These include, but are not restricted to, the University of York DropOff (file sharing) service - https://www.york.ac.uk/it-services/services/dropoff/, alongside other commercial file sharing services DropBox, GoogleDrive, etc. The ADS also utilises SFTP transfers with depositors where necessary.
[7] The exchange of DVD, USB or portable hard-drives in person, or through the postal service.
[8] Guidelines for Depositors - https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml, accessed 15 June 2020.

provide specific information on accepted formats[9] alongside collection-level documentation[10] and file-level metadata requirements.[11] While the *Charging Policy* provides information on the associated charges for deposition.[12]

2.0.5 The ADS also provides general advice for depositors:

- *Data Management and Sharing Plans*[13]
- *Guidance on the Selection of Material for Deposit and Archive.*[14]
- *Digitising Journal Articles and Grey Literature Reports.*[15]
- *Policy and Guidance on the Deposition of Personal, Confidential and Sensitive Data*.[16]
- *Archaeology Data Service / Digital Antiquity Guides to Good Practice*.[17]

2.0.6 If additional information or assistance is required, depositors are encouraged to contact the ADS.[18]

2.0.7 *Checklists* for both standard and ADS-easy accessions are available to guide repository staff through the process[19], while this Ingest Manual provides additional and specific guidance.

3. Data Accession (Standard)

3.1 Data transfer

3.1.1 As noted above, the ADS accepts submissions through a variety of methods.[20] This can include the digital transfer of data and metadata, or the exchange of physical media.

[9] *Data Type Requirements* - https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, accessed 28 July 2020.
[10] *Collection-level Metadata Requirements* - https://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level%20Metadata%20Requirements, accessed 15 June 2020.
[11] *File-level Metadata Requirements* - https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#File-level%20Metadata%20Requirements, accessed 15 June 2020.
[12] These are outlined in the ADS *Charging Policy* - https://archaeologydataservice.ac.uk/advice/chargingPolicy.xhtml, accessed 15 June 2020.
[13] *Data Management and Sharing Plans* - https://archaeologydataservice.ac.uk/advice/DataManagementPlans.xhtml, accessed 15 June 2020.
[14] *Guidance on the Selection of Material for Deposit and Archive* - https://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml, accessed 15 June 2020.
[15] *Digitising Journal Articles and Grey Literature Reports* - https://archaeologydataservice.ac.uk/advice/scanningGuide.xhtml, accessed 15 June 2020.
[16] *Policy and Guidance on the Deposition of Personal, Confidential and Sensitive Data* - https://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml, accessed 15 June 2020.
[17] *Archaeology Data Service / Digital Antiquity Guides to Good Practice* - https://guides.archaeologydataservice.ac.uk/g2gpwiki/, accessed 15 June 2020.
[18] Contact - https://archaeologydataservice.ac.uk/about/contact.xhtml, accessed 15 June 2020.
[19] These checklists are available internally. Static versions of these checklists are available through the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml, accessed 15 June 2020.
[20] See the How to Deposit Data section of the ADS website - http://archaeologydataservice.ac.uk/deposit/How.xhtml.

Those depositions sent through the repository's dedicated submission streams (i.e. OASIS Images[4] and ADS-easy[5]) are part of a more automated workflow.

3.1.2 The ADS also supports the digital exchange of data outside of these dedicated streams,[6] or through the submission of physical media. Each of these require the 'manual' movement of data; first to a local drive, for data assessment and review, before transfer to ADS servers. All such data transfers use the SFTP protocol.

3.2 Pre-accession checks

3.2.1 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 dataset with the repository previously. To facilitate this assessment data is stored, temporarily, in the ADS file store.[21] Checks carried out during this assessment include, but are not limited to, ensuring that:

- all formats are suitable for deposition
- appropriate collection and file level metadata has been completed
- the dataset is accurately documented and complete
- an appropriate licence has been agreed with the depositor (although this may not be issued)
- funding for the archive has been arranged and any charges documented in the Collection Management System (CMS)
- details of 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

3.3.1 The ADS' *Guidelines for Depositors*,[8] and specifically the list of formats in the *Data Type Requirements*[9], provide depositors with guidance on data types and formats that are appropriate for submission. While repository staff also provide support and guidance to depositors to ensure that submissions follow these requirements. Often, the identification of problematic formats or data types typically occurs during negotiation for deposition, but can also be part of the checking prior to accession.

3.3.2 Data submitted through *ADS-easy*[5], *OASIS Images*[4] and *OASIS* is subject to programmatic controls, as outlined in the *Data Type Requirements*, to restrict submissions to the formats appropriate to each data type. Additional, manual checks ensure that they adhere to the repository's policies and guidelines.

3.3.3 Submissions outside of the above digital submissions streams require more manual checking of the SIP to ensure that they adhere to repository guidelines and policies.

[21] In the 'Unaccessioned holdings' directory.

3.3.4 Depositors submit new versions, or replacement files, in those case where problematic formats, data corruption, or improperly formed files are noted. In circumstances where such 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

3.4.1 In accordance with the *Security Overview*[22], the *Information Security Risk Assessment*[23], and the policies provided by the *University of York IT Services*, all PCs are protected using anti-virus software and a firewall.[24] The repository strictly follows the guidance provided by the *University of York IT Services* in order to mitigate against potential infection. The use of virus and malware prevention software ensures that the SIP is virus-free before any transfer of data to local hard drives or servers.

3.4.2 Those deposits made through *ADS-easy*, *OASIS Images* and *OASIS*, are submitted to programmatic virus checking during the upload process. At the same time, additional checking accompanies all data transfers between systems and servers during the ingestion of data.

3.4.3 Digital deposits outside the above submission streams, and those accessions involving the exchange of physical media, are subject to the same checks outlined in the *Security Overview*, albeit that these checks are initiated manually.[25]

3.4.4 On the detection of viruses or malware, the repository follows the policies and guidelines detailed in the *Security Overview,*[22] and provided by the *University of York IT Services*.[26] Any infected data are isolated and disinfected. In these instances a 'clean' copy of the file, or data, a replacement files is requested from the depositor and replaces the infected files.

3.5 Media and file readability check

3.5.1 All media and data submitted to the ADS are 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.[27] Checks will vary according to the data type and format submitted and the ADS' *Data Procedures* provide individual

[22] Security Overview - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security, accessed 15 June 2020.
[23] Information Security Risk Assessment - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk, accessed 15 June 2020.
[24] Virus & malware protection - https://www.york.ac.uk/it-services/security/virus/, accessed 15 June 2020.
[25] Anti-virus software carries out 'on-access' scanning, but manual checks are also carried out.
[26] https://www.york.ac.uk/it-services/security/virus/#tab-3
[27] *Appendix 1* provides information on the checks carried out by the ADS during accession.

guidance on this.[28] The preparation of both AIP and DIP allows further opportunity for assessment of data from collections.

3.5.2 In those instances where problems with the SIP are identified, these are reported to the depositor and, where necessary, replacement files sought. These replacement files then form a new accession to the collection, again subject to all the checks listed above, with the unreadable/problematic data removed from the dataset.

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

3.6 Documentation and integrity check

3.6.1 The ADS requires **all** submissions be accompanied by appropriate collection and file-level metadata. Specifics of these requirements are outlined in the *Guidelines for Depositors*[8] and through a series of metadata templates and exemplars.[29]

3.6.2 Where depositors use the ADS' digital submission streams (*OASIS*[2], *OASIS Images*[4] or *ADS-easy*[5]) a series of online forms or file uploads, allow the submission and validation of metadata. Programmatic and digital checks ensure that all necessary metadata fields are completed, while more 'manual' qualitative checks are carried out by repository staff as part of the ingest process. This ensures the accuracy of all supplied metadata.

3.6.3 Those datasets submitted digitally, outside of these submission streams, or through the exchange of physical media, are subject to the same checks and qualitative assessments, however, in these cases repository staff carry out these checks manually. Again, such checks ensure the currency of all metadata.

3.6.4 These assessments also ensure that there is parity between the data supplied and the metadata submitted. While the ADS does not require the inclusion of dedicated directory or file lists to ensure the 'completeness' of depositions, where they are included they are checked to ensure that the complete dataset has been received.[30] These checks ensure the identification of any missing files at the earliest opportunity.[31]

---

[28] These are available internally only, although static versions of these *Data Procedures* are made publically available from the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures, accessed 15 June 2020.

[29] https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#File-level%20Metadata%20Requirements, accessed 15 June 2020.

[30] The issuing of the 'deposit receipt', noted above, allows depositors to identify problems or gaps in the submitted dataset.

[31] 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 an MD5 checksum for each file, allowing depositors to identify any problems that may have occurred during the data transfer.

3.6.5 Repository staff carry out qualitative checks on all documentation to ensure accuracy. Where these checks identify inaccurate or incomplete metadata, staff will work with the depositor to address any gaps or concerns.

3.6.6 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. The repository also ask that the depositor supply image(s) to illustrate the archive.

3.6.7 It is the responsibility of depositors, as per the terms of the deposit licence, to ensure that they have the 'rights', as data owner or copyright holder, to submit the dataset to the repository. Similarly, it is the responsibility of the depositor to ensure that the submission adheres to current legal and ethical guidelines. The repository does provide additional guidance on the deposition of personal, confidential, and sensitive data allowing depositors to plan for the submission of data.[16] In those instances where repository staff identify data that they believe infringes the licence or policies, these concerns are directed towards the depositor and, where necessary, the Collections Development Manager. A clear record of any such discussions in maintained within both the preserved dataset and the ADS' *Collection Management System* (CMS) ensuring clear documentation of any issues.

3.7 Record details of SIP in the CMS[32]

3.7.1 Once assessment of the SIP is complete, repository staff document the accession using the CMS.[10] This record the date of accession and depositor; the creation of a digital checklist is included to track the collection as it moves through the ADS workflow.

3.8 Create a new 'version' of the dataset[33]

3.8.1 In those instances where the SIP includes a new 'version' of the entire dataset then repository staff should ensure that the directory structure is as outlined in the *Repository Operations Manual*.[34] This is programmatic process initiated from the CMS.[35]

3.8.2 Any creation of a new version of the dataset should be correctly documented, and checked, within the CMS.

3.9 Standardise file names and check directory structure

3.9.1 Generally, depositors should follow the file naming guidance provided in the *Guidelines for Depositors*.[8]

[32] The ADS' *Collection Management System (*CMS) is available for internal access only.
[33] This process is only necessary where a new version of the dataset has been submitted.
[34] *Repository Operations Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 15 June 2020.
[35] This process also updates technical metadata, specifically file paths, within the OMS.

3.9.2 In the case of depositions submitted via *OASIS*[2], *OASIS Images*[4] and *ADS-easy*[5] stringent programmatic controls are in place[36] to ensure that depositors adhere to the file naming policy. [37]

3.9.3 For other digital depositions outside of these submission streams, or involving the exchange of physical media, manual checks ensure adherence to the file naming policy.[37]

3.9.4 Repository staff ensure that all files have been checked and, where required, updated in accordance with the file naming policy. In accordance with *Repository Operations Manual,*[34] documentation of any changes or updates within the CMS ensures a clear record of any changes to the dataset.

3.9.5 As noted in the *Guidelines for Depositors*[8] and *Repository Operation Manual*[34] the ADS does not make any specific stipulations regarding the directory structure used within the SIP. However, as the *Guidelines for Depositors* note, we do ask that "a logical file structure" is used that "allows data to be easily retrievable".[38] Data that is poorly structured may make preservation activities difficult, and will certainly affect the dissemination of data and restrict reuse. Where SIPs use a poor, or overly complex, data structure, it may be necessary for repository staff to make changes to the data structure of the submitted dataset. Appropriate documentation, using the CMS, ensures a clear record of any such changes.

3.10 Copy to data server

3.10.1 Once the data assessment has been complete and all necessary file name and structural changes have been made to the SIP, then it should be copied from the local drive to the preservation server.

3.10.2 All file data transfers use a dedicated client and follow the SFTP protocol.

3.10.3 Where the dataset has been deposited digitally using *ADS-easy*[5] or *OASIS Images*[4] all data will be stored on a dedicated server until accession; at this point data will be transferred to the preservation server during a semi-automated ingest process initiated in the CMS. These data transfers use the SFTP protocol, with checksum validation to ensure the successful transfer of files.

[36] This is part of the upload process within these systems.
[37] *Repository Operations Manual, Appendix 5: File naming policy* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 15 June 2020.
[38] *Guidelines for Depositors: File Management* - https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#File%20Management, accessed 15 June 2020.

3.11 Create checksums and technical metadata

3.11.1 The ADS uses the National Archives' file characterisation application, DROID[39], to create technical metadata for each file within the SIP.[40] This includes a checksum, used to monitor the 'health' of the file within the archive, alongside file format, file format version and MIME-type that facilitate the ongoing management of the dataset. All metadata is stored within the OMS.[41]

3.12 Issue licence and create data receipt

3.12.1 All depositors are required to 'sign' a deposit licence, issued digitally from the CMS as part of the accession process.[42]

13.12.2 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.[43]

13.12.3 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 access and reuse for the dataset.[44]

13.12.4 It is important to note that a signed deposit licence is required before archival work can commence.

13.12.5 The signed deposit licence, once returned, should be stored alongside the data as outlined in the *Repository Operations Manual*.[34] A copy of the licence should also be stored in the CMS.

3.13 Attach correspondence to CMS

3.13.1 Deposition of data can sometimes be a protracted process requiring negotiation with the depositor. It is important that any correspondence is preserved and attached to the collection record in the CMS.

3.13.2 Any correspondence pertinent to the preservation or dissemination of the dataset should also be stored alongside data. This should be stored in accordance with the structure outlined in the *Repository Operations Manual*.[34]

---

[39] DROID - http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/, accessed 15 June 2020.
[40] Technical metadata includes physical location, filename, size, format, MIME type, PRONOM identifier and, most importantly, checksum/fixity value.
[41] Object Metadata Store (OMS) is used collect and store metadata associated with a file or object. This is for internal access only.
[42] A sample licence can be found within the *Guidelines for Depositors* - https://archaeologydataservice.ac.uk/resources/attach/ADS_Deposit_Licence_2018.pdf, accessed 15 June 2020.
[43] In instances where an individual is signing the deposit licence on behalf of an organisation, it is important that the appropriate 'organisation' is included in the licence.
[44] An overview of licence types and reuse is available - https://archaeologydataservice.ac.uk/advice/reuseLicence.xhtml, accessed 15 June 2020. While specific information on Identifying Copyright and Re-use Licences for data users is also available - https://archaeologydataservice.ac.uk/advice/IdentifyingCopyright.xhtml, accessed 15 June 2020.

3.13.3 All file names for correspondence should follow the guidance outlined in the *Repository Operation Manual.*[37]

3.14 Scan paper documentation

3.14.1 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 *Repository Operations Manual*.[45]

3.15 Acknowledge receipt of data and issue deposit receipt

3.15.1 Once the accession is complete, repository staff should acknowledge the receipt of all data by email. The email should highlight any issues, queries or concerns regarding the dataset. Repository staff should ensure the depositor is given an opportunity to address or respond to any problems, and any such responses should be clearly documented within the CMS, and emails retained within the file store in accordance with Repository Operations Manual.[34] [46]

3.15.2 A 'data receipt' should be attached to the acknowledgement email.[47]

3.15.3 All emails, and the deposit receipt, should be stored in accordance with *Repository Operations Manual*.[34]

3.16 Store original media

3.16.1 Once an accession is complete, physical media should be labelled with the correct collection number and accession id(s). Media should then be stored in the collections filing cabinets in the ADS offices.

3.16.2 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. A note is added to the CMS to document that media has be returned to the depositor.

3.16.3 In circumstances where data has been deposited electronically through ADS streams (*ADS-easy*[5] or *OASIS Images*[4]) or external file-sharing services, these should only be deleted once data has been copied to the preservation servers.

3.17 Document completion of accession process

3.17.1 Once the depositor has acknowledged that the submission is 'complete' and as intended, and a 'signed' deposit licence has been returned, the accession can be 'signed

[45] See *Repository Operations Manual, Appendix 3: Requirements for Scanned Hard Copy Material* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.
[46] In some circumstances, it may be necessary for the depositor to submit replacement data. In these circumstances, a new accession should be created for the new/replacement data and the old data removed from the original accession. All actions should be documented within the CMS.
[47] An outline on the contents and creation of the deposit receipt is found in Appendix 7.

off'. Repository staff should ensure that the 'accession' process is marked as complete on the internal workflow checklist.

4. Data Accession (OASIS Images[4] and ADS-easy[5])

4.0.1 Details on accessioning collections submitted digitally through the ADS dedicated submission streams (ADS-easy and OASIS Images) are provided through the ADS' internal wiki and should be carried out in conjunction with the ADS-easy accession checklist.[19]

4.0.2 The accessioning of data from ADS-easy follows much the same process associated with that of a standard collection, although elements of the process are undertaken programmatically prior to, or during, the import. There should be, for example, no need to check formats and undertake virus checks as the ADS-easy system carries these out automatically.

4.1 Importing collection metadata

4.1.1 The first stage in both is to import the *ADS-easy* project metadata via the 'Projects ready to import into the CMS' section of the CMS.[48]

4.1.2 Locate the project that you wish to accession and press the 'Import Data' button to begin the import process.

4.1.3 The initial process involves the transfer of collection metadata from *ADS-easy* system into the CMS. Repository staff should review all collection metadata during the import; quantitative checks should be undertaken during metadata creation ensure all fields are completed but additional checks ensure all metadata meets current, qualitative requirements.[49]

4.1.4 Repository staff will be required to link existing identities within the CMS to the appropriate 'people' and 'organisation' metadata within each deposition. This provides a mechanism to reduce the duplication of documentation within internal systems. Once complete staff should click the 'create new protect' button to import collection-metadata and create a CMS record.

4.1.4 A workflow checklist is created automatically during data ingest. Additionally, repository staff should manually check this to ensure that that the archive appears in the correct workflow.

4.2 Accessioning data

4.2.1 Following creation of the CMS record, transferral of data from the *ADS-easy* system into the repository can begin. The 'Database Import' function, found within the CMS, initiates the process of copying data; however, before beginning the process checks of all file-level metadata (in the *ADS-easy* system) ensure that there are no problems with the associated

[48] Internal access only.
[49] Halting the import at this stage and re-opening the submission within ADS-easy allows the depositor an opportunity to make any necessary changes.

documentation. A view of the metadata is available from within the 'Database Import' window. Where issues are extensive, it may be worthwhile re-opening the project in *ADS-easy* so that the depositor can correct, or enhance, any metadata. Where the problems are minor, contact with the depositor should follow the completion of the import process.

4.2.2 Once initiated, the 'Database Import' copies data from the *ADS-easy* system into the repository file store. This process ensures that all formats are valid,[50] carries out virus and integrity checks on all data,[51] establishes that file names conform to the *file naming policy*,[52] and creates the appropriate archive structure, as outlined in the *Repository Operations Manual*.[34] When 'finished', checking the report (below the table) ensures that all data has been successfully copied and facilitates the identification of any problems.[53]

4.2.3 Any programmatic file name changes carried out during the import process will be included in this report.[54] The 'Processes' section of the CMS should be used to document any name changes.

4.2.4 Repository staff should carry out manual media and file readability checks on a representative sample of the dataset to ensure that all files open and function as intended.[55]

4.3 Import ADS-easy[5] metadata into the CMS

4.3.1 Once the data has been successfully imported, repository staff should initiate the second phase of the process and import the file-level metadata from *ADS-easy*; initiated from the 'Database Import' section of the CMS.

4.3.2 A report, at the bottom of the import screen identifies any 'technical' problems or issues experienced during the transfer of the metadata. Repository staff are encouraged to carry out qualitative checks on all file-level metadata to ensure that it meets current standards and requirements.[56] The provision of a simple viewer, accessed from the CMS, facilitates these checks.

---

[50] As outlined for a 'manual' import in *Section 3.3: Check file formats are suitable for deposition*, above.
[51] As outlined in Section *3.4: Virus check*, above. Additional checks are undertaken of checksums to ensure that the transfer of data from one server to another has not introduced error or corruption.
[52] As documented in Section 3.9: Standardise file names and check directory structure, above. Details of the file naming policy and reserved file names are also documented in the *Repository Operations Manual*, *Appendix 4: Reserved File Names* and *Appendix 5: File naming policy* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 15 June 2020.
[53] Although rare, the report includes information on discrepancies between checksums of the data stored within the ADS-easy system and the files once transferred into the repository. Repository staff should investigate any reported issues. Where necessary staff can reimport the entire data, retrieve data 'manually' from ADS-easy, or seek replacement files directly from the depositor.
[54] For example, 'WARNING - These filenames were changed because they contained special characters. Keep a note of these changes, as these will need to be documented as a 'Process'.
[55] As outlined in *Section 3.5 Media and file readability checks* above.
[56] As outlined in the *Guidelines for Depositors* - https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml, accessed 02 July 2020. Detailed information on these requirements is also available in the *Data Procedures*, internal access only. Static versions of these documents are available from the ADS website -

4.4 Technical metadata creation and metadata checks

4.4.1 The creation of technical metadata is a semi-automated process initiated by repository staff following the import of data and metadata from *ADS-easy*. This follows the same process outlined above in *Section 3.11 Create checksums and technical metadata*.

4.4.2 Repository staff are encouraged to check and update the collection metadata where necessary. The list of 'issues to look out for' provides some assistance in identifying common problems.[57] Again, it may be necessary for repository staff to contact the depositor directly to address some issues with the collection metadata.

4.5 Notification of receipt and issuing of deposit licence

4.5.1 This follows the same process outlined in section *3.15 Acknowledge receipt of data and issue deposit receipt* above.

4.6 Document completion of accession process

4.6.1 The follows the same process outlined in section *3.17 Document completion of accession process* above.

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

5.0.1 Once accession is complete, repository staff 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.[19] Alongside these checklists, the ADS' *Data Procedures* provide information on the processes and procedures for the preservation and dissemination of discrete data types.[28]

5.0.2 Only once preservation and dissemination activities are complete, and the creation of both AIP and DIP finished, will the process of AIP checking begin.

5.1 Check and assess the significant properties of files to be preserved and establish conversion plan.

5.1.1 Repository staff should check and assess the significant properties of the dataset alongside any associated documentation and metadata.[58]

5.1.2 In some circumstances previously undetected problems may arise once we start working with the data and preparing the AIP and DIP. Where issues are noted, the digital archivist should contact the depositor and, where necessary, encourage them to re-submit

https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures, accessed 02 July 2020.

[57] Internal access only.

[58] As outlined in *Data Accession, Section 3.5 Media and file readability check* above. *Appendix 1* provides information on the general checks undertaken, while the ADS' Data Procedures include detailed documentation of specific checks carried out for discrete data types.

replacement data in order to maintain data integrity and authenticity. It may be possible for repository staff to undertake edits on behalf of the depositor, but any such changes require documentation within the CMS.[32]

5.1.3 Repository staff should develop a conversion plan for all archives. The ADS' *Data Procedures* provide detailed guidance on appropriate formats for preservation and dissemination.[28] In some circumstances, data may already be in formats suitable for preservation or dissemination and may not require normalisation, but in others, conversion may be more complex and require multiple steps. Therefore, it may be worthwhile mapping out each step to ensure the appropriate software and expertise is available. While the *Data Procedures* provide guidance on typical preservation and dissemination pathways, it may be necessary to consult with colleagues over the most appropriate course of action for your dataset. 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*.[28]

5.2 Preserving and disseminating data

5.2.1 The creation of a working or 'local' version of the dataset, on which all normalisation and updates can be carried out, will ensure the integrity of the original data (AIP). All data should be normalised to suitable preservation and dissemination formats in line with the preservation plan and *Data Procedures*.[28]

5.2.2 Repository staff 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.

5.2.3 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 transfers use a dedicated client, and follow the SFTP protocol. Fixity checks 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 Manual*.[34]

5.3 Create technical metadata

5.3.1 Once copied to the ADS servers the use of DROID[39] allows the creation of technical metadata for all files and normalised data within both the AIP and DIP. The transferal, and storage, of all technical metadata to the OMS is a semi-automated process initiated through the CMS.

5.4 Creating and matching objects

5.4.1 Repository staff should use the 'match objects' functionality, initiated from through the CMS, to link all related files into notional 'objects'.[59] This is largely programmatic, although it may be necessary to group files and objects manually. Checks should be made of all objects to ensure that they are accurate, including the creation of 'parent-child' objects where relationships between ingested and preserved files are complex (see 5.7).

5.5 Ensuring the correct data types have been assigned

5.5.1 Digital archivists should ensure that all files within the collection have been assigned a 'data type'[60] as part of the accession process (outlined in Section 3.10), or, in the case of depositions through the *ADS-easy* and *OASIS Images* submission streams, created by the depositor during data upload.

5.5.2 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.

5.5.3 Repository staff should ensure the accuracy of all assigned data types.[61] Where necessary, functionality within the CMS allows 'updates' to be made to the data type.

5.5.4 Other files, such as the deposit licence or deposit receipt, may have an internal and administrative function; repository staff should ensure that these objects have been assigned the correct data type.[62] Again, 'updates' to the assigned data type can be made through the CMS.

5.6 Ensuring the correct resource types have been assigned

5.6.1 To facilitate the internal management of data, repository staff should also ensure that the correct 'resource type' is allocated.[63] This allows greater granularity in characterising the content of files and objects. The CMS allows the correct resource type to be assigned.

5.7 Establishing relationships and relationship types

5.7.1 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.[64]

---

[59] See *Appendix 8: Match Objects, for a fuller explanation of this data management functionality*.
[60] See *Appendix 2: Data types*, below.
[61] 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. See *Appendix 2: Data types*.
[62] The data type for files holding other documentation relating to the preservation of the collection is 'Admin'. See *Appendix 2: Data types*.
[63] See *Appendix 4: Resource type* for a more detailed explanation
[64] The ADS uses an agreed list of 'relationship types' documented in *Appendix 3: Relationship types*.

5.8 Add and/or update collection and file level metadata

5.8.1 **ADS-easy**[5] **and OASIS Images**[4]**:** In instances where data has been submitted through *ADS-easy* or *OASIS Images*, metadata is completed through a series of online forms. These services provide some quantitative checking of metadata during upload, with additional 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[65], however repository staff should always carry out checks to ensure accuracy and completeness.

5.8.2 The identification and resolution of any gaps or problems with the dataset form part of the accession process. However, in instances where the identification of new/other issues with collection or file level metadata are noted, or where extant problems have not been addressed to the satisfaction of the repository staff, it may be necessary to contact the depositor (again) and updates to the data or metadata sought. In some circumstances, repository staff may carry out updates or enhancements to metadata as part of the archiving process; any such changes require documentation within the CMS.

5.8.3 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 within the DIP as a separate file in a suitable format.[66] The inclusion of this downloadable version of the file-level metadata within the archive interface provides data users with the necessary documentation to facilitate data reuse.[67] All metadata extracted from the OMS should be stored in accordance with the *Repository Operations Manual*.[34] Repository staff should ensure that the correct data type has been assigned[68] and any relationships between metadata and data documented.[69]

5.8.4 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 submitted using the ADS' dedicated metadata templates.[70] Collection and file level metadata submitted using these templates should be transferred to the CMS and OMS respectively.[71] As noted above (*Section 5.8.2*), care should be taken to follow the guidelines concerning

---

[65] See *Section 3.1: Data transfer session*.
[66] Note all collection level metadata is directly available through each archive interface; therefore, any extraction from the CMS is not necessary. See, for example, Chiz Harward (2020) *St Nicholas House, 47 London Road, Gloucester, Gloucestershire. Archaeological Watching Brief (OASIS ID: urbanarc1-342180)* [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1078328, specifically https://archaeologydataservice.ac.uk/archives/view/urbanarc1-342180/metadata.cfm, accessed 15 June 2020.
[67] The extraction of the metadata from the OMS and the creation of the downloadable file are documented within the 'processes' section of the CMS.
[68] See *Appendix 2: Data types*.
[69] See *Appendix 3: Relationship types*.
[70] See *Guidelines for Depositors: Downloads* - https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, accessed 15 June 2020.
[71] A dedicated metadata loader, provided through the CMS, allows staff to add raster file-level metadata directly from the spreadsheet into the OMS. For other data types, the transfer of file-level metadata is a largely manual process.

problems or issues with data or metadata. 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*[34] and disseminated in appropriate formats outlined in the *Data Procedures*.[28] Repository staff should ensure that the correct data type has been assigned[68] and any relationships between metadata and data documented.[69]

5.8.5 For all types of submission, depositors may upload additional or supplemental documentation, beyond the required standard metadata, as separate, discreet files. The accession of additional metadata should follow the workflow outlined above for all depositions.[72] Supplemental metadata is subject to the same qualitative checks carried out for all data and metadata. All supplemental metadata should form part of both the AIP and DIP in an appropriate formats outlined in the *Data Procedures*.[28] It should also be stored in accordance with the *Repository Operations Manual.*[34] Repository staff should ensure that the correct data type has been assigned[68] and any relationships between metadata and data documented.[69]

5.9 Record conversion and editing processes undertaken

5.9.1 All conversions, or processes, carried out on the dataset should be documented in the 'Processes' section of the CMS. A semi-automated system allows repository staff to generate most of these programmatically following file matching (see *Section 5.4 Creating and matching objects*). This process makes assumptions about the nature of the conversion, so all processes require checking to ensure accuracy. Repository staff can make edits or updates to processes through the CMS interface. In some cases, it may be necessary to add processes manually.

5.9.2 The CMS documents the type of process carried out[73], and includes a detailed record of each process.[74] A record of any problems, or other information, should be included in the comments section of each process.

5.10 Interface creation

5.10.1 Following the creation and documentation of both AIP and DIP, the creation of a dedicated interface for each collection should follow. Guidance on this is available from the ADS wiki,[75] and in consultation with the Collections Development Manager.

5.10.2 For 'standard' depositions the ADS uses a series of templates and exemplars to facilitate this process. These templates can be adapted and amended where required.

5.10.3 For 'special collections' a bespoke interface must be created. Agreement on the form and content of the interface is generally during negotiations for deposition, and documented within the CMS – whether it includes a database search, a map interface, 3D viewer, etc.

[72] See *Section 3 Data Accession (Standard).*
[73] See *Appendix 5: Process type*.
[74] See *Appendix 6: Process documentation*.
[75] See Creating an archive interface, internal access only.

5.10.4 Once completed, whether the interface is a 'standard' deposition or 'special collection', repository staff should ensure that the interface meets the required standards in terms of accessibility, validation, and compatibility.[75]

5.10.5 The CMS allows repository staff to keep a clear record of any interface functionality, queries, directories, database tables, and repository templates.[76]

5.11 Updating associated documentation

5.11.1 The CMS can be used to document any 'notes' concerning the dataset and the preservation process, while it also allows the storage of any associated documentation and correspondence relating to the dataset.[32] Any documentation relating, specifically, to the preservation or dissemination of the dataset should also be included in the AIP. All documentation should be stored in accordance with the *Repository Operations Manual*[34] and named in accordance with the *File Naming Policy*.[37]

5.11.2 Repository staff should ensure the completion of all required collection-level metadata within the CMS.

5.11.3 Any correspondence or documentation pertinent to the deposition, maintenance or management of the dataset should preserved within the CMS. These can be added as 'notes' or attached to the CMS record.[77]

5.12 Updating associated collection metadata

5.12.1 Repository staff should check that the transfer of all collection-level metadata from the requisite template submitted as part of the SIP, or, where data has been deposited digitally via *ADS-easy* or *OASIS Images*, has been transferred to the CMS.

5.12.2 As noted above, typically, the identification of any problems or gaps in the metadata occurs during accession with any updated and additional metadata received from the depositors. In rare circumstances, however, repository staff may notice issues whilst working with the dataset, in such circumstance they be required to contact the depositor directly for additional information. As observed above, it may also be necessary for repository staff to augment and enhance metadata; the CMS allows the creation of a clear record of any such processes and activities.

5.12.3 The dissemination of the collection metadata through the 'metadata' page within the archive interface ensures the data users have clear access to the documentation for the archive.[78]

[76] In the 'web admin' section.
[77] Ensure that all attached files are submitted in a suitable preservation format.
[78] See, for example, Nigel Nayling, University of Wales Trinity St David, Toby Jones, Newport Museums and Heritage Service (2017) *Newport Medieval Ship* [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1044659, particularly https://archaeologydataservice.ac.uk/archives/view/newportship_2013/metadata.cfm, accessed 29 June 2020.

5.13 Submit AIP for checking[79]

5.13.1 On completion of both the AIP and DIP, AIP checks ensure that the outcomes meet the requirements outlined in this Ingest Manual, the *Repository Operations Manual*[34] and *Data Procedures*.[28] An assigned digital archivist initiates a semi-automated process of checks, alongside visual inspections of the AIP (and DIP), following completion of the preservation and dissemination procedures. 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 repository staff, should highlight any issues. This is often the case for archives submitted through the *ADS-easy* or *OASIS Images* submission streams. The repository still regards AIP checking as an important part of the preservation process, particularly as it ensures the correct adherence to recommended archival procedures and policies. Instances where AIP checking may be necessary include, but are not restricted to, the following:

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

5.13.2 Collections requiring AIP checks are marked within the CMS, and the Collections Development 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,[19] with additional information provided in this *Ingest Manual*, the *Repository Operations Manual*[34] and the *Data Procedures*.[28] 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.

5.13.3 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 where appropriate and within reason. The formal submission of archives for "sign-off" by the Collections Development Manager follows.

5.14 Archive release

5.14.1 Following sign-off by the Collections Development Manager, the archive is ready for release. An agreement for the date of release is sought from the depositor.

[79] 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.

5.14.2 **Embargo:** 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.[80] Following appraisal by the depositor, and completion of the AIP checks (where carried out), 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 advance 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.

5.14.3 Detailed documentation for the release process is available in the ADS wiki,[81] although a shorthand version is available within the *Procedure Checklist*.[19] A final check is made of the archive (AIP and DIP) which ensures that all procedures have been followed, all processes documented, and all metadata completed.

5.14.4 Repository staff run DROID[39] to ensure that all files within the AIP and DIP are recorded within the OMS[41] and all files have been correctly linked together into notional 'objects' using the 'match objects' functionality.[82] Checks should be carried out to ensure that all 'data types' have been correctly recorded. These actions ensure that an accurate record of the archive contents, and up-to-date technical metadata, form part of the documented dataset, facilitating the ongoing management of the dataset.[83]

5.14.5 The 'date of release' (and associated 'ready for release') should be added to the CMS. Part of the release process involves the minting of the persistent Digital Object Identifier (DOI) for the collection.[84] All releases are publicised through the ADS' 'Collections History' page[85], resource discovery metadata[86] transferred to the *ArchSearch* catalogue[87], and the archive is included in the archive index.[88] In instances where the archive includes formal reports or other 'library-worthy' documents, citations are also added to the *ADS*

[80] See the Collections Policy - https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml, accessed 29 June 2020.
[81] Internal access only.
[82] See Appendix 8: Match Objects, for a fuller explanation of this data management functionality.
[83] For example, changes to the 'Admin' directory may result from the AIP-checking process or in instances where detailed documentation of the release process is necessary.
[84] DOIs 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 , accessed 30 June 2020.
[85] Collections History - https://archaeologydataservice.ac.uk/about/collectionsHistory.xhtml, accessed 30 June 2020.
[86] 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, accessed 30 June 2020.
[87] *ArchSearch* - https://archaeologydataservice.ac.uk/archsearch/basic.xhtml, accessed 30 June 2020.
[88] Archive index - https://archaeologydataservice.ac.uk/archive/, accessed 30 June 2020.

Library.[89] Where the archive is 'marine based', metadata is added to the *Marine Environmental Data and Information Network* (MEDIN) data portal.[90]

5.14.6 The release process also involves the transfer of the complete AIP to off-site, 'deep storage' where the dataset is backed up to ensure that the long-term preservation of the dataset.[91]

5.14.7 Once released, the repository publicises the publication of the collection through its website and social media.

6. Updating content

6.0.1 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.

6.1 Adding data to an extant collection[92]

6.1.1 When a depositor wishes to add further data to an existing collection, repository staff should create a new accession within the AIP and data from the new SIP added to the 'original' directory as outlined in the *Repository Operations Manual*.[34] The archive then enters the workflow as a new accession with data added to the SIP and DIP accordingly and repository staff carrying out the same procedures and processes as with a 'new' collection.

6.1.2 Where necessary, depositors should submit additional collection- and file-level metadata so that all documentation is current and accurate. All updated and replacement metadata should be incorporated into the CMS[32] and OMS[41] respectively, with clear documentation of all actions undertaken for the preservation and dissemination of the dataset. Where necessary consultation with the Collections Development Manager ensures any changes to the archive interface are as agreed. Submission for an AIP check ensures that all preservation actions and processes are in accordance with current practice.[93]

6.1.3 On release, repository staff should ensure that the 'last modified date' and, particularly, the 'updated year' fields are updated within the CMS. This ensures a clear collection history for both staff and data consumers.

---

[89] *ADS Library* - https://archaeologydataservice.ac.uk/library/, accessed 30 June 2020.
[90] https://portal.medin.org.uk/portal/start.php, accessed 30 June 2020.
[91] The repository uses Amazon Web Services S3 Glacier cloud storage for off-site, 'deep storage'. See *Systems Overview* for wider discussion - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems, accessed 30 June 2020. A full account of the process is available internally.
[92] 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.
[93] In some circumstances it may be many years since the initial deposition so repository staff should ensure that all changes are in line with current practice.

6.1.4 All updated releases should be publicised through the ADS' 'Collections History' page[94], updated resource discovery metadata[95] transferred to the *ArchSearch* catalogue[96] and the archive index[97] rerun to legislate for changes to the archive. While the adding of data to an existing collection should not require the minting of a new DOI for the collection, it is necessary to update the metadata so that the information is current and accurate.

6.1.5 Repository staff should also ensure the transfer of all additional data, metadata, and documentation to off-site 'deep storage' where the dataset is backed-up.[98]

6.2 Replacing part/all of a dataset and versioning

6.2.1 Updates involving the replacing of data are a little more complex as depositors/users may require access to previous versions of part/all the dataset. Following the OAIS model[99] the ADS retains all previous versions of a file and editions of the 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 Manual*.[100] This is a semi-automated process initiated through the CMS.

6.2.2 When a depositor wishes to submit replacement data to an existing collection, repository staff should create a new version of the AIP and DIP as outlined in the *Repository Operations Manual*.[34] The archive then enters the workflow much like a new accession and repository staff carry out the same procedures and processes as with a 'new' collection.

6.2.3 Where necessary, depositors should submit additional collection- and file-level metadata so that all documentation is current and accurate. All updated and replacement metadata should be incorporated into the CMS[32] and OMS[41] respectively, with clear documentation of all actions undertaken for the preservation and dissemination of the dataset added for the new version of the AIP and DIP. All necessary or requested changes to the archive interface under consultation with the depositor and the Collections Development Manager. An AIP check ensures that all preservation actions and processes are in accordance with current practice.[93]

---

[94] Collections History - https://archaeologydataservice.ac.uk/about/collectionsHistory.xhtml, accessed 30 June 2020.
[95] 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, accessed 30 June 2020.
[96] ArchSearch - https://archaeologydataservice.ac.uk/archsearch/basic.xhtml, accessed 30 June 2020.
[97] Archive index - https://archaeologydataservice.ac.uk/archive/, accessed 30 June 2020.
[98] The repository uses Amazon Web Services S3 Glacier cloud storage for off-site, 'deep storage'. See *Systems Overview* for wider discussion - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems, accessed 30 June 2020.
[99] https://www.dpconline.org/knowledge-base/preservation-lifecycle/oais, accessed 30 June 2020.
[100] 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, accessed 30 June 2020.

6.2.4 On release repository staff should ensure that the ‘last modified date’ and, particularly, the ‘updated year’ fields are updated within the CMS. This ensures a clear collection history for both staff and data consumers.

6.2.5 Repository staff should also ensure that all additional data, metadata and documentation is transferred to off-site, ‘deep storage’ where the dataset is preserved and stored in the long-term.[98]

Appendix 1: Data validation and consistency checks

A1.0.1 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.[101]

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

[101] This document in now no longer available.

Appendix 2: Data types

A2.0.1 The ADS uses dedicated data types to classify digital objects (files) within the OMS.[41] 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

A2.0.2 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

Appendix 3: Relationship types

A3.0.1 The ADS adheres to the PREMIS concept of a relationship, i.e. "a relationship in which one object provides documentation for another".[102] Agreed ADS Relationship Type List:

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

Table A3.0.1: Relationship types used by the ADS.

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

[102] http://id.loc.gov/vocabulary/preservation/relationshipType/collection_PREMIS.html, accessed 30 June 2020.

Appendix 4: Resource type

A4.0.1 To facilitate the internal management of data the repository uses a discreet 'resource type', based on the *FISH Resource Description Thesaurus*[103], which characterises the content of data. These add granularity to the 'data type' classification.[104]

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

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

Appendix 5: Process type

A5.0.1 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

Appendix 6: Process documentation

A6.0.1 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.

Table A6.0.1 Preservation actions and processes.

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.

Appendix 7: Deposit Email and Data Receipt

A7.0.1 All datasets received by the ADS are acknowledged by a returned pro forma email on receipt of each accession.[105] Example deposit email:

> Subject heading: Deposit of [PROJECT TITLE]
>
> Dear [depositor]
>
> Thank you for depositing the digital data associated with the above project, all files have been copied to our servers and will be stored until work is started on the collection. Please retain your own copies until we have notified you that your collection is ready to be released and you have checked it.
>
> What happens now?
>
> Please check the list attached to ensure that we have received all files intended for archiving. If you notice any errors or omissions, please let us know straightaway. Addition or substitution of files once work has begun on your collection may incur charges to cover the costs of additional work.
>
> If you have yet to sign our licence agreement, please do so, as work will not commence on your collection until this has been received.
>
> Should you have any queries concerning your collection, or the work we will be doing for you, then please do not hesitate to contact us.
>
> Yours sincerely,
>
> ADS

A7.0.2 Accompanying this email is a list of all files included within the accession; depositors are encouraged to check this 'deposit receipt' to ensure that the repository has received **all** the intended files.

- Filename and file path
- File size (expressed in MB)
- File type
- MD5 Checksum[106]

[105] This includes submissions via ADS-easy and OASIS images.
[106] The string of digits representing the sum of a piece of stored or transmitted digital data, against which later comparisons can be made to detect errors in the data.

The deposit receipt includes a ‘checksum’ for each file so that depositors can ensure that all files have remained unchanged during transfer to the repository.

A7.0.3 File-naming and storage of both the deposit email and data receipt should follow the guidance outlined in the *Repository Operations*.[107]

[107] Following the naming strategy outlined in the Repository Operations (see *Appendix 4: Reserved File Names* and *Appendix 5: File naming policy*). These files should be attached to the CMS and stored within the AIP in accordance with *Repository Operations* (see section *5.5.1 Project Metadata*).

Appendix 8: Match Objects

A8.0.1 Digital archivists use the 'match objects' functionality to link all related files and data into notional 'objects', which are described by the same discovery metadata.[108] As such, an object should not to be confused with a file, a data type[109], a thematic group, or a bibliographic object, but can be thought of as the smallest unit of data that users may wish to discover balanced with the most specific metadata that our depositors are willing to create.

A8.0.2 Matched objects are generated programmatically using an automated 'computer matching' function enacted within the CMS and recorded within the OMS.

A8.0.3 Additional relationships between files and objects can be added/updated manually. In some instances, particularly where the deposition is large, 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.

A8.1 Example 1: a simple object

A8.1.1 In this example a depositor the SIP includes a .pdf format. According to current procedures, this file will be preserved and disseminated in the same format submitted to the repository. Each of these files will be stored in accordance with the *Repository Operations Manual*,[34] but the discrete elements are 'grouped' together into a single object.

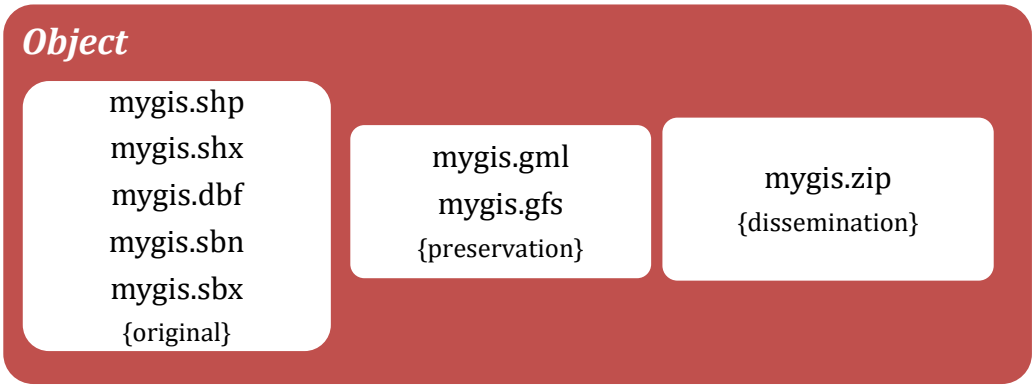
A8.2 Example 2: a simple object, with different

A8.A8.2.1 A depositor includes the my_other_report.doc within the accession. Within current data procedures, normalisation of the original file to the .docx format ensures preservation, while the conversion to .pdf allows the dissemination of the document. Each of these files will be stored in accordance with the *Repository Operations Manual,*[34] but the discrete elements are 'grouped' together into a single object.

[108] Documentation of the 'match objects' process can be found within the ADS wiki, internal access only.
[109] See Appendix 2: Data types, above.

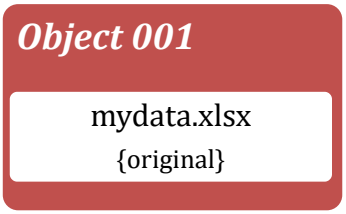
A8.3 Example 3: single dataset outputted across multiple files

A8.3.1 A depositor has included some GIS files within the deposited dataset. According to the data procedures, the preservation format(s) should be .gml and .gfs, with the original shapefile (and associated files) added to a compressed, zipped archive (.zip) for dissemination. Each of these elements will be stored in accordance with the *Repository Operations Manual*,[34] but with each file 'grouped' together into a single object.

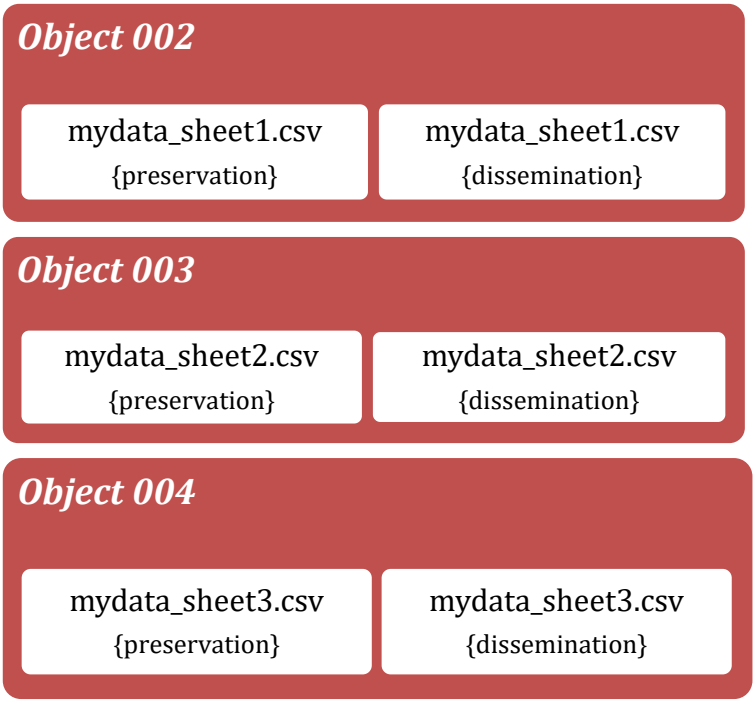


A8.4 Example 4: a simple object, comprising many preservation or dissemination elements

A8.4.1 A deposited archive includes some spreadsheet data accessioned in the .xlsx format. In line with current data procedures, each of the tables that make up the spreadsheet will be preserved and disseminated in the .csv format. Each of these elements will be stored in accordance with the *Repository Operations Manual*,[34] but with each table 'grouped' together into separate objects. All relationships between the deposited spreadsheet/object and the tables/objects will documented within the 'parent-child' table with the appropriate PREMIS relationship type.[110]



[110] See Appendix 3: Relationship types, above.

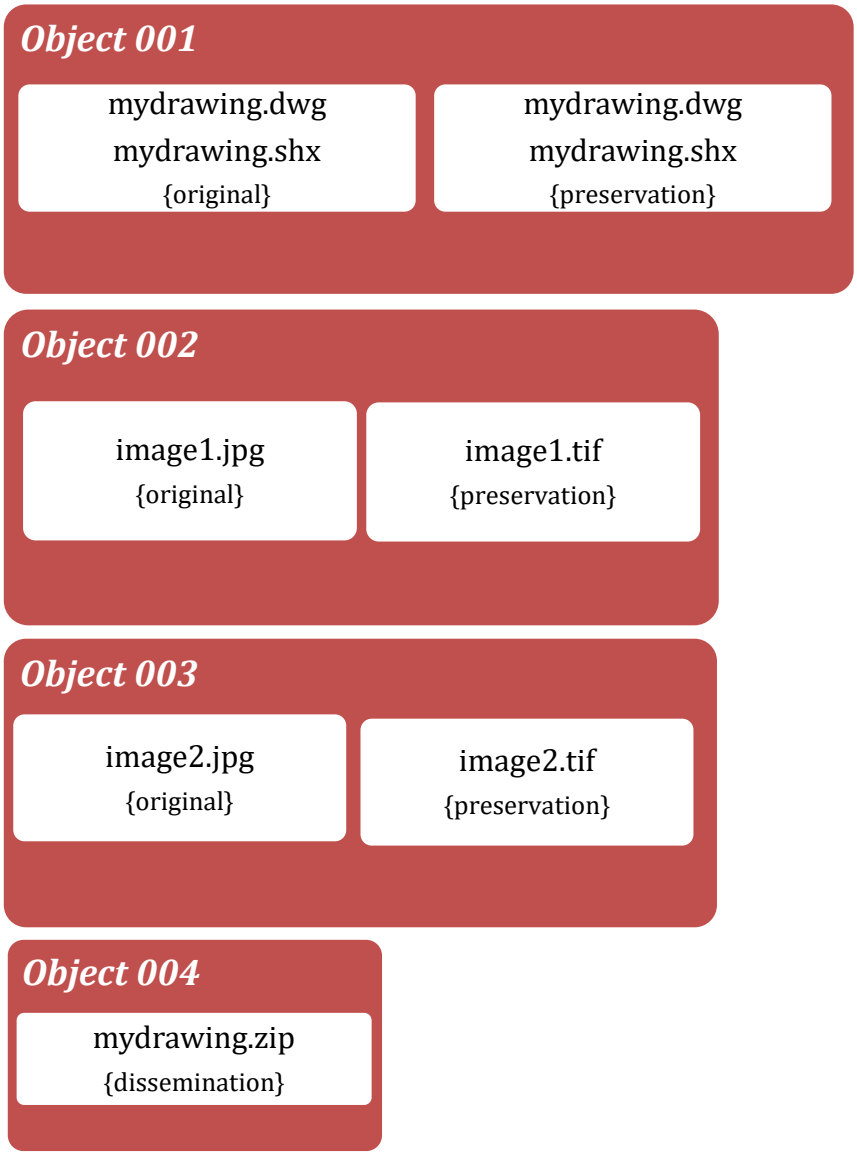


A8.5 Example 5: Data with embedded elements stored in separate files

A8.5.1 A depositor includes a vector drawing (.dwg) within the SIP that has embedded images (.jpg) and fonts (.shx). For preservation, the .dwg file is normalised to the .dwg (2018) format, with the images converted to the uncompressed .tif format, while the fonts in the original .shx format according to current procedures. For dissemination, a .zip archive of the original files is utilised. Each of these elements will be stored in accordance with the *Repository Operations Manual*,[34] with all files 'grouped' together into a single object.

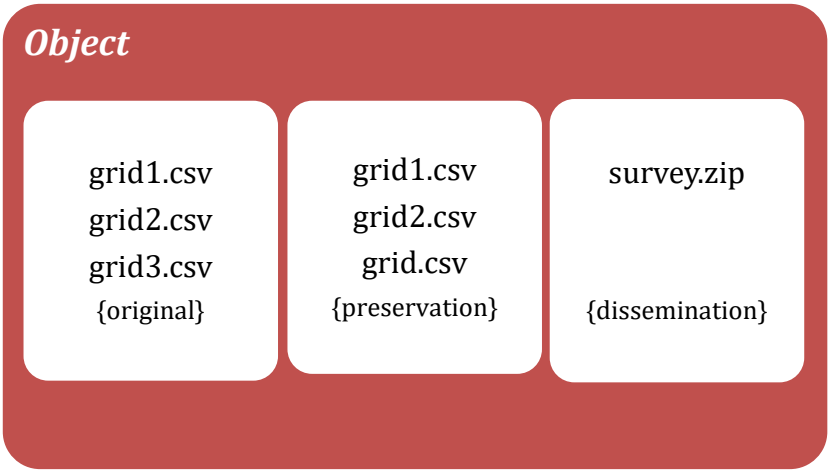
A8.5.2 Where the associated images have dedicated metadata in their own right, then elements may be broken up into discreet objects with relationships between the discrete

objects established within the 'parent-child' table using the appropriate PREMIS relationship type.[110]



A8.6 Example 6: Outputs of digital research spread over multiple files

A8.6.1 Someone deposits a geophysical survey which was undertaken on a gridded system and outputted as discrete files for each grid in the .csv format. Preservation of the dataset, according to current practice, would maintain the discreet files in .csv format. All data would be shared in a zipped archive (.zip) for dissemination. Each of these elements will be stored in accordance with the Repository Operations Manual,[34] with all files 'grouped' together into a single object.


Object
grid1.csv
grid2.csv
grid3.csv
{original}
grid1.csv
grid2.csv
grid.csv
{preservation}
survey.zip
{dissemination}

Appendix 9: Collection Level Metadata Requirements

A9.0.1 A template for the current collection level metadata requirements is provided in the *Guidelines for Depositors*.[111]

[111] See https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, specifically https://archaeologydataservice.ac.uk/resources/attach/ADS_collection_level_metadata_template.doc, accessed 01 July 2020. Details on completing the metadata form, with a completed example are also provided https://archaeologydataservice.ac.uk/resources/attach/ADS_collection_level_metadata_example.pdf.