

ADS Assessment and Appraisal Policy

Tim Evans
10/08/2021

Document Control Grid

| | |
|-------------------------|-------------------------------------|
| Title: | ADS Assessment and Appraisal Policy |
| File name: | assessment_and _appraisal_v1-0 |
| Location: | Internal Drive |
| Status: | LIVE |
| Version: | 1.2 |
| Last updated: | 17/09/2021 |
| Created date: | 10/08/2021 |
| Review due: | 01/09/2022 |
| Authors: | Tim Evans |
| Maintained by: | Deputy Director |
| Required Action: | Circulation |

Contents

| | |
|---|-----------|
| Glossary | 2 |
| Introduction | 4 |
| Scope of this Policy | 5 |
| The Assessment and Appraisal event | 6 |
| Relationship to Accession event | 6 |
| Location of data | 6 |
| Assessment Criteria | 7 |
| Data deposit contains no malware | 7 |
| Digital objects are in correct formats | 8 |
| Data deposit has Collection-level metadata | 8 |
| All digital objects have Core metadata | 9 |
| Digital Objects have Additional technical metadata | 11 |
| Digital objects can be opened, are valid, and can be reused | 12 |
| The data deposit has no Sensitive Data concerns | 13 |
| Content is appropriate and complete | 15 |
| Assessment Outcomes | 16 |
| Dataset can be Accessioned | 16 |
| Dataset returned to Depositor | 16 |
| Resubmitting | 20 |

Glossary

| | |
|------------------|--------------------------------|
| Accession | |
| ADS | Archaeology Data Service |
| AIP | Archival Information Package |
| DPC | Digital Preservation Coalition |

| | |
|-----------------|---|
| Metadata | Descriptive information about data |
| NFS | Networked File Storage |
| Object | The basic unit of digital information, for example a Report |
| SIP | Submission Information Package |
| VM | Virtual Machine |

1. Introduction

The Assessment and Appraisal Policy describes a defined **phase of work** undertaken on **all** data supplied to the ADS. This Assessment and Appraisal work is part of the larger process of **Ingest** - whereby data is moved into the archive for long-term preservation and dissemination. Ingest covers a multitude of smaller human and machine (i.e. automated) tasks to collect, check, and shape data given to us by the depositor into a coherent and consistent package that is called the **Accession**. The Accession is a formal transfer of data, logged with a Unique Identifier and marked with the issuing of a formal receipt from the ADS to the Depositor.

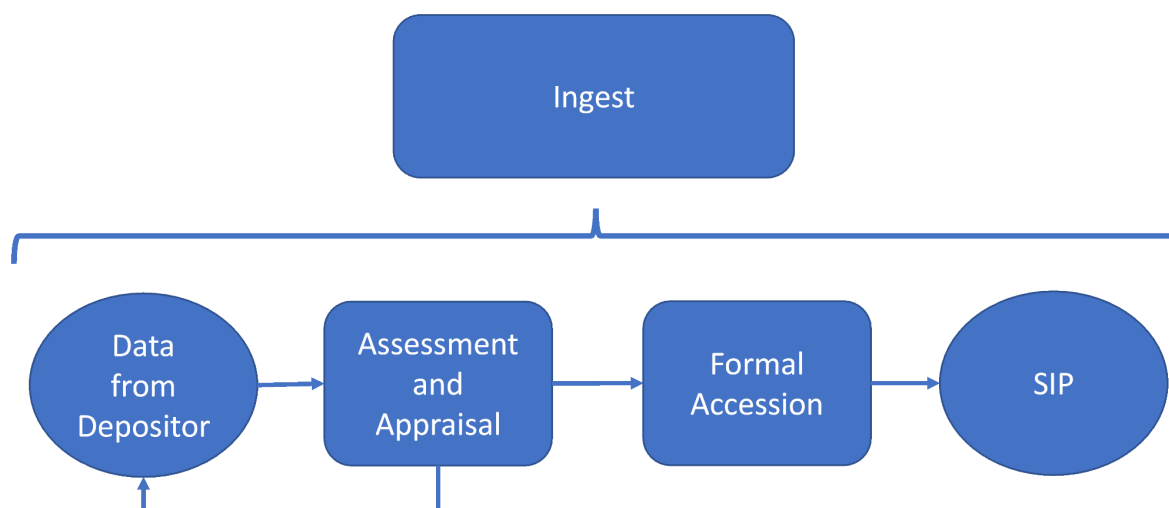


Figure 1: the role of Assessment and Appraisal within Ingest.

The rationale for Assessment and Appraisal is to ensure that datasets are well formed, documented, and meet all legal requirements. The event:

- Acts as a precursor to formal accession procedures, and ensures a minimum standard required for a consistent SIP.
- Ensures a base level of technical integrity in the dataset; files can be opened, verified, and free of malware.
- Ensures sufficient collections level metadata is present, so that dataset can be managed within the ADS catalogue.
- Ensures every digital object has a core level of metadata, that allows it to be managed and discovered within the ADS catalogue.

- Ensures that digital objects are documented with pertinent technical metadata that allows them to be understood and re-used.
- Ensures that the data formats supplied meet the highest standards for long-term preservation and re-use.
- Ensures that all due care and consideration for sensitive data has been taken by the depositor.
- Prevents excessive or ad-hoc revision during the rest of the Ingest phase.
- Ensures decisions made about the final SIP are documented in a consistent and transparent manner.
- Ensures the highest possible standard of data deposition in-line with the ADS Collections Policy.

2. Scope of this Policy

The Assessment and Appraisal Policy sits within an existing network of Policy documents, which where relevant are cited and linked to herein. These include:

- The ADS Collections Policy
- The ADS Sensitive Data Policy
- The ADS Ingest Manual
- The ADS Deaccession and Data Disposal Procedures

The key practical relationship is with the **ADS Ingest Manual**, which provides an overview of the process of accepting an archive, and formal accessioning of a SIP. As described below, Assessment and Appraisal forms a distinct part of Ingest for each deposit of data, but is not in itself an accession.

It is the intention that having a separate and defined policy for this stage allows the ADS to be consistent in approach, and for all Depositors to be aware of the criteria by which data are reviewed.

3. The Assessment and Appraisal event

3.1. Relationship to Accession event

The purpose of Assessment and Appraisal is to ensure that data is fit to be accessioned, and that any major issues are identified and wherever possible resolved before the creation of the SIP. The outcome of the event should only ever be one of the following:

- Data meets expected criteria, and can be formally accessioned and the creation of a SIP.
- Data does not meet expected criteria, and is returned to the depositor with comments on how to proceed. No accession takes place.

It should be clearly stated that this event takes place **prior** to the formal accession of any data onto ADS preservation systems. The ADS takes no responsibility for the long-term maintenance of datasets which have not been accessioned. The ADS have several defined steps to ensure that Depositors are provided with guidance and information on best practice. A failure to respond to guidance and instruction arising from Assessment and Appraisal is covered in the ADS Deaccession and Data Disposal Procedures.

If the dataset is approved for accession, the depositor will be notified of the formal receipt of the SIP via procedures covered in the ADS Ingest Manual.

3.2. Location of data

The Assessment and Appraisal phase takes place outside of main ADS preservation systems. Where data is supplied manually (e.g. via internet transfer, email attachment or physical media), data is moved to a designated space within the ADS NFS demarcated as non-accessioned holdings. In singular cases, such as a very large dataset, a local file store may be used for ease of access.

For the ADS-easy and OASIS applications, files uploaded by the external user are held on a separate VM or section of the ADS NFS, where they reside until formally accessioned as part of the ingest process for those applications. Again, it should be clearly noted that data uploaded to ADS-easy and OASIS has **not** been formally accessioned, and thus cannot be considered preserved until the full creation of an AIP.

For the sake of consistency in this and other ADS documents, all of these spaces named above are referred to as **Temporary Holdings**.

4. Assessment Criteria

The following criteria will form the basis of an Assessment and Appraisal event undertaken by the ADS on **all** datasets submitted. The full outcomes are covered in Section 5 of this document.

4.1. Data deposit contains no malware

It is essential that the ADS are confident that no virus or malware is included in the deposited data.

Deposits made through ADS-easy and OASIS, are submitted to programmatic virus checking while on NFS or VM space. Additional virus scanning as part of Assessment and Appraisal is thus not required.

Virus scanning is **mandatory** for all manual deposits. The use of virus and malware prevention software ensures a clean dataset before any transfer of data to Temporary Holdings.

On the detection of viruses or malware, the ADS will follow the policies and guidelines detailed in the Security Overview (any infected data are isolated and disinfected).

If a virus is detected in any part, the dataset **cannot** be accessioned.

The Depositor will be instructed to supply a clean copy of the relevant object(s). See Section 5.2 for details.

4.2. Digital objects are in correct formats

The ADS preservation-by-migration approach depends on having a normalised set of data formats which are either defined as suitable for long-term preservation, or else flexible enough that a migration workflow could be established where required.

These formats have also been set specific to different types of data, with particular formats offering the optimal solution for the preservation of the significant properties of an object. For example, a vector data format for geographic information system (GIS) software is best deposited as a Shapefile or Geographic Markup Language, and not a PDF file.

Accordingly, all objects deposited with the ADS should meet the requirements specified in the **ADS Guidelines for Depositors**, and relevant to the type of dataset.

Data submitted through ADS-easy and OASIS is subject to programmatic controls to restrict submissions to appropriate formats. However, **additional** manual checks will ensure that they adhere to the repository's policies and guidelines.

If a discrepancy is noted the dataset **cannot** be accessioned.

The Depositor will be instructed to supply objects in suitable formats. See Section 5.2 for details.

4.3. Data deposit has Collection-level metadata

All datasets deposited should be accompanied by higher-level metadata, the requirements for which are specified in the **ADS Guidelines for Depositors**.

In the case of a formal update (i.e. a new version), a fresh set of Collection metadata should be supplied by the Depositor that reflects the history and provenance of the revised/new dataset.

The ADS-easy system collects this information within the online form, and should ensure consistency with the Guidelines for Depositors. However, an assessment of the quality and accuracy of this metadata should still be made to ensure it is fit for purpose.

The ADS digital archivist will check to ensure that the following mandatory fields have been completed:

- The Collection **Title**
- The Collection **Description**
- The Language(s) used in the collection
- The Data Creation Dates
- The Primary Contact.
- The Copyright Holder(s)
- The Licence Holder
- The Data Creator(s)
- Keywords: at least three keywords

Although the ADS will also check for common spelling mistakes in metadata, the presence of typos will not impact the final decision as these can be cleaned during the creation of the AIP and Cataloguing.

If a required metadata field is not provided then the dataset **cannot** be accessioned. The ADS will not make efforts to complete missing metadata.

The Depositor will be instructed to re-supply the metadata template with the required fields completed. See Section 5.2 for details.

4.4. All digital objects have Core Metadata

Every digital object deposited should be accompanied by the ADS Core Metadata Terms (ACT), and using the exact template provided by the ADS and available on the **Guidelines for Depositors**.

These **essential** elements are required for ADS to perform a number of key functions for digital preservation, including rights management and facilitating discovery and re-use.

The Digital Archivist appraising the deposit will check that **each object** has:

- Filename
- Title
- Description
- Tag
- Creator
- Copyright Holder
- Creation History
- Language
- Year(s) digitally created

If the metadata template is not used, or if a Term is not filled in then the dataset **cannot** be accessioned.

For each data type, a representative sample of 10% of objects will be cross-checked by the digital archivist to ensure the veracity of the Core Metadata.

The Digital Archivist appraising the deposit will check for:

- The title meets ADS Guidelines by being short and descriptive. For example "image.jpg" is not permitted
- The description meets ADS Guidelines by allowing users to easily understand the contents and purpose of the object. For example "Photo" is not permitted.
- The title and description **match** the content of the object. For example where a photograph is described as a flint arrowhead, but is in fact a picture of a horse.

If the Digital Archivist appraising the deposit identifies multiple cases of a breach of requirements, then the dataset **cannot** be accessioned.

In either a case of missing, or inaccurate core metadata, the Depositor will be instructed to re-supply the metadata template with the required or corrected fields completed. See Section 5.2 for details.

4.5. Digital Objects have Additional Technical Metadata

Certain data types require additional metadata to document and explain technical settings, structure, or other specific elements that are essential for long-term support and re-use. For example, a relational database needs the relations and structure defined so that another user can rebuild it. In ADS terms, this is referred to as **Additional Technical Metadata**.

The **Guidelines for Depositors** stipulate the metadata required for each data type. With advice and recommendations provided in templates or pro-formas.

The Digital Archivist appraising the deposit will check that each object, where required, has the minimum standard of Additional Technical Metadata and that the relationship between object and metadata is clearly defined. Unlike Core Metadata, Additional Technical Metadata **does not** have to be in an ADS template. And the following guidelines are followed:

- Exports from users' own systems are acceptable.
- Overarching documents that cover multiple data types are acceptable.
- Appendices in reports are acceptable metadata, as long as the reports themselves are archived as part of the **same** Collection. Either in the same Accession or in a previous Accession. Reports accessioned via OASIS **cannot** be used as metadata for an object in a separate collection. In these cases information within the report - or a copy of the report itself - must be included within the accession within the same collection.
- Linking to documents hosted on another website, or within another ADS Collection is not permitted. The metadata **must** be supplied with the deposit.
- For formal updates (such as a new version of a database), the metadata must be supplied with the dataset. Referring to previous metadata is **not** permitted.

If an object is **missing** the Additional Technical Metadata stipulated by the Guidelines for Depositors then the dataset **cannot** be accessioned.

For each data type, a representative sample of 10% of objects will be cross-checked by the digital archivist to ensure the veracity of the technical metadata. The Digital Archivist appraising the deposit will specifically check for key elements including, but not limited to:

- Fields within tabular data (e.g. databases, spreadsheets, geophysical survey data, etc.) are documented and the same data type in both the object and metadata.
- GIS projection is the same in the object and metadata.
- Layers within CAD files are fully listed and documented.

If the significant properties for the data type outlined in the Guidelines for Depositors are not covered by the metadata provided then the dataset **cannot** be accessioned.

The Depositor will be instructed to re-supply the metadata with the required fields elements present. See Section 5.2 for details.

4.6. Digital objects can be opened, are valid, and can be reused

It is essential that digital objects deposited with the ADS can be opened and are not limited in functionality via password protection or other forms of content locking.

For each data type, a representative sample of 10% of objects will be opened by the digital archivist to ensure that they open correctly. Certain data types will also undergo further visual checking to ensure the content is valid and fit for purpose:

- Database and Spreadsheet: to ensure that columns and headers are correctly formed.
- GIS: objects display correctly and with no error warning.

- Images: fit the relevant baseline profile (for example TIFF have no LZW compression).
- Documents: track changes or other editing/commenting functions not present.
- All objects: there is no password protection or locking of content that prohibits normalisation of the format, or digital re-use of the data itself.

If any objects cannot be opened, or issues are spotted in manual validation then the dataset **cannot** be accessioned.

The Depositor will be instructed to re-supply the objects with corrections. See Section 5.2 for details.

4.7. The data deposit has no Sensitive Data concerns

It is essential that the depositor has reviewed their dataset for Sensitive Data issues, and ensured any such issues have been addressed.

The collective term of Sensitive Data covers personal data, confidential data and sensitive personal data. In the historic environment 'sensitive data' can also be sensitive due to the nature of the historic environment investigations. These definitions are explored and defined fully in the ADS Sensitive Data Policy.

When data is deposited with the ADS, depositors must sign an [ADS Deposit Agreement](#) before any data is accessioned into the ADS archive. This [Deposit Agreement](#) states that the depositor has given all fair processing notices and/or, as applicable, consents have been obtained and that the data collection does not and will not contravene any laws. **Data therefore should not enter the ADS archive unless all due care and consideration for sensitive data has been taken by the depositor.**

The Digital Archivist appraising the deposit will check 100% of raster images, and 10% of all other data types for any Sensitive data flags:

- The dataset contains images of children.
- The dataset contains administrative or organisational emails or other data that could be considered confidential.

- The dataset contains financial information.
- The dataset contains personal addresses and/or telephone numbers.
- The dataset contains IP information.
- The dataset contains sensitive data related to an **identifiable** individual, which includes:
 - ◆ racial or ethnic origin
 - ◆ political opinions
 - ◆ religious or philosophical beliefs
 - ◆ trade union membership,
 - ◆ genetic data
 - ◆ biometric data
 - ◆ information about criminal convictions,
 - ◆ health
 - ◆ person's sex life or sexual orientation.

If the dataset contains images of children then a consent form should be included in the dataset record or referenced in the dataset's Data Management Plan or other documentation provided.

If the dataset contains administrative or organisational data that could be considered confidential, then a consent form should be included in the dataset record or referenced in the dataset's Data Management Plan or other documentation provided.

If the dataset contains personal information then a consent form should be included in the dataset record or referenced in the dataset's Data Management Plan or other documentation provided.

If the dataset contains sensitive personal data then a consent form should be included in the dataset record or referenced in the dataset's Data Management Plan or other documentation provided.

If any Sensitive data flags are spotted and a clear and obvious form of consent is not present, then the dataset **cannot** be accessioned.

The Depositor will be instructed to either supply the relevant consent form, or provide instruction to remove the highlighted objects. See Section 5.2 for details.

4.8. Content is appropriate and complete

Except for the checks on data format and validity noted above, the ADS makes no comment on the intrinsic quality of the data deposited. However, Assessment and Appraisal should ensure that what is deposited is not incomplete or inappropriate.

For each data type, a representative sample of 10% of objects will be checked for:

- Any objects that do not belong to the named collection. For example reports from another project or site.
- Content that clearly does not relate to the collection or provide historical, social or geographical context (see note below).
- Any objects obviously missing data. For example spreadsheets or database tables with no rows, shapefiles with no points, lines or polygons, or reports that have been truncated or obviously missing pages.
- Any clear and obvious duplication of data objects.
- Any objects that are draft or working versions that have been deposited in error i.e. not clearly marked as such in the accompanying metadata.

In addition, datasets produced from specific techniques will be expected to conform to accepted community best practice on content and standards. The ADS archivist will also assess datasets from laser scanning and geophysical survey, to ensure that the elements deposited **match the requirements** for the data type specified in the ADS Guidelines for Depositors.

If any issues are identified then the dataset **cannot** be accessioned. The Depositor will be instructed to either supply the relevant consent form, or provide instruction to remove the highlighted objects. See Section 5.2 for details.

A note on images: ADS Policy is that images reflecting the wider context of a project are an important part of the archive. This includes informal “working shots” of staff and the local environment. However, Appraisal by the Digital Archivist should flag excessive numbers, or else wholly inappropriate images. Immediate arbitration is passed to the ADS Collections Development Manager. In the case that images are deemed unsuitable, they will be included in the report back to the depositor.

5. Assessment Outcomes

5.1. Dataset can be Accessioned

If the Digital Archivist has not noted any issues, then the dataset can be formally accessioned as per ADS procedures. No further action is required.

5.2. Dataset returned to Depositor for Correction or Clarification

If any issues have been identified, then the ingest stage pauses, and **no data** is accessioned. The Digital Archivist responsible for appraising the data will contact the original Depositor by email notifying them that action is required, and with a short pro-forma report attached to highlight issues and a description of required action or solution. Depending on the issue, this action will either be a full or partial (re)submission from the Depositor, or a confirmation from the Depositor on an ADS action (e.g. removing files).

The email will contain instructions on how to re-submit and the timeline for doing so.

Email title: *"Action required - issues with data submitted to ADS for Project #project_name#"*

Email text:

"Dear #depositor_name#"

Assessment of the data supplied for this project on #current_date# has highlighted issues which prevent the ADS being able to be formally accessioned.

A full report of these issues is attached to this email.

Please respond to this email, addressing each point within the text of your response. For issues where new/additional files are required then either attach them to your response, or use an internet transfer method of your choice (such as Dropbox).

*Until these issues are resolved the data you have supplied resides on ADS temporary holdings only. The data **has not** been formally accessioned into the ADS archive. If **no response** is received to this email within 6 months the deposit **will be removed** from ADS temporary holdings according to the ADS Deaccession and Data Disposal Procedures.*

*As a reminder, all data supplied to the ADS should conform to the instructions on the **ADS Guidelines for Depositors**.*

The generated report will be named according to the following convention: #ads_tracking_id#-appraisal_report-#dd_mm_yyyy# and provided in an XLSX format.

A copy of the email and report will be stored in the ADS CMS record.

Malware present

| Issue | Relates to | Required Action | Archivist Note |
|-----------------|-------------------|------------------------|-------------------------------|
| Malware present | #object_name# | Resubmit object | Spotted during scan on #date# |

Incorrect formats

| <i>Issue</i> | <i>Relates to</i> | <i>Required Action</i> | <i>Archivist Note</i> |
|---|--------------------------|-------------------------------|-------------------------------------|
| Format not accepted | #object_name# | Resubmit object | Please use #format# for this object |
| Format does not preserve significant properties | #object_name# | Resubmit object | Please use #format# for this object |

Collection metadata missing

| <i>Issue</i> | <i>Relates to</i> | <i>Required Action</i> | <i>Archivist Note</i> |
|--|------------------------------|---------------------------------------|-----------------------------------|
| Collection metadata not deposited | Collection metadata template | Submit Collection Metadata template | Please use the template at #link# |
| Collection metadata fields missing or incomplete | Collection metadata template | Resubmit Collection Metadata template | Please fill in the creator field. |

Core metadata missing

| <i>Issue</i> | <i>Relates to</i> | <i>Required Action</i> | <i>Archivist Note</i> |
|--|--------------------------|-------------------------------|-----------------------------------|
| Core metadata not deposited | #file_path/#object_name# | Submit Core Metadata template | Please use the template at #link# |
| Core metadata fields missing or incomplete | #file_path/#object_name# | Core Metadata template | Please fill in the creator field. |

Technical metadata missing

| Issue | Relates to | Required Action | Archivist Note |
|---|--------------------------|------------------------|---|
| Technical metadata not deposited | #file_path/#object_name# | Submit metadata | Please submit required metadata for this data type as described at #link# |
| Technical metadata fields missing or incomplete | #file_path/#object_name# | Resubmit metadata | Please fill in the creator field. |

Files not valid

| Issue | Relates to | Required Action | Archivist Note |
|-----------------------|--------------------------|-----------------|----------------|
| File cannot be opened | #file_path/#object_name# | Resubmit object | |
| File contains error | #file_path/#object_name# | Resubmit object | |

Sensitive Data alert

| Issue | Relates to | Required Action | Archivist Note |
|----------------------|--------------------------|---|--------------------------------------|
| File contains #flag# | #file_path/#object_name# | Submit consent or confirm in writing that file should be removed from | Please see ADS Sensitive Data Policy |

| | | | |
|--|--|----------------|--|
| | | deposit by ADS | |
|--|--|----------------|--|

Content not suitable

| Issue | Relates to | Required Action | Archivist Note |
|------------------------------------|--------------------------|--|--------------------------------------|
| Content does not match description | #file_path/#object_name# | Resubmit correct object or confirm in writing that file should be removed from the deposit | The object actually depicts #insert# |
| | | | |

6. Resubmitting

Formal instruction on how to resubmit data is provided in the email report issued by the ADS.

Upon receiving a resubmission/response from the Depositor, the ADS Digital Archivist will compare this to the original report (see 5.2) to ensure that all issues have been addressed. Any re-submitted or new (meta)data will be assessed using the same criteria as the original data.

If all issues have been satisfactorily resolved, then the data will move forward to the Accession stage. A copy of **any** pertinent emails provided by the Depositor during resubmission will be stored according to standard ADS procedure.