Deaccession and Data Disposal Processes at the Archaeology Data Service

#DPClinic February – Deletion, disposal and deaccessioning in digital preservation

23rd February 2024

Olivia Foster
ADS Policy and Procedures

- Public facing policy
- Internal procedures

Deaccession and data disposal policy

The Deaccession and Data Disposal Policy provides a framework for the deaccessioning and disposal of data from the ADS archive collections.

Prior to 2022 elements of this policy could be found in the Appraisal and Deaccession Policy which details a framework of policies and procedures put in place to retrospectively appraise datasets, deaccession data and support the sustainable management of ADS' collections.

https://archaeologydataservice.ac.uk/about/policies/ads-policies-and-procedures/appraisal-and-deaccession-policy/
“the core activity of the ADS is the long-term digital preservation of the data entrusted to our care”

So why are we deleting data?
Scenarios for data disposal

Data supplied for costing/appraisal only

Replacement data provided

No longer wish to/able to deposit

Incomplete dataset

Issues identified during archiving

GDPR/Copyright issue

Depositor wants data to available exclusively elsewhere
May be asked to remove data at any stage of the archiving process

- Removal of non-accessioned data and incomplete projects
- Removal or replacement of individual files
- Complete removal of a released dataset
1. Documentation of dataset prior to disposal

- Documentation of correspondence concerning the removal of the dataset
  Ensure there is a clear record of why data is being removed

- Record contents of the dataset by creating a ‘data log’ of the SIP
  Ensure there is a clear record of data to be removed
  - checksums, filenames and file sizes

- Return data to depositor if required

---

<table>
<thead>
<tr>
<th>File &amp; Path</th>
<th>File Size</th>
<th>File Type</th>
<th>MD5 Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S21HREC1_context_sheets.pdf</td>
<td>3579469</td>
<td>Acrobat PDF 1.6 - Portable Document Format</td>
<td>4d55849a55e97bd4c3c7bc93f1a3f465</td>
</tr>
<tr>
<td>1S21HREC1_ATOM_text_metadata.xlsx</td>
<td>22576</td>
<td>Microsoft Excel for Windows</td>
<td>b3e2c855299c6a5a5bea15454e567c4d</td>
</tr>
<tr>
<td>DSCF6848.jpg</td>
<td>7248286</td>
<td>Exchangeable Image File Format (Compressed)</td>
<td>5261b4307cdd183489b70940e21e718c</td>
</tr>
<tr>
<td>DSCF6851.jpg</td>
<td>7497963</td>
<td>Exchangeable Image File Format (Compressed)</td>
<td>e9f0e5305c5c4c13f8be8c5eb6f59310</td>
</tr>
</tbody>
</table>
2. Identify where the dataset is published

- Remove any download links for the files from the ADS website
- Update interfaces to explain data has been removed
  - A clear message outlining what has happened to the resource, e.g.
    - Resource identified as duplicate of {identifier for other resource}
    - Redacted or requested to be removed by the Publisher/Author
    - Document contains Personal or Sensitive data
  - Details of any new persistent identifiers for the dataset alongside terms of access for the receiving repository
- Collection metadata (a full description of the resource) is retained
2. Identify where the dataset is published

What shouldn’t be removed?

- Persistent identifiers
- Collection level metadata

<table>
<thead>
<tr>
<th>Title:</th>
<th>Carlisle Northern Development Route Cumbria Post-excavation Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series:</td>
<td>Oxford Archaeology North unpublished report series</td>
</tr>
<tr>
<td>DOI</td>
<td><a href="https://doi.org/10.5284/1109723">https://doi.org/10.5284/1109723</a></td>
</tr>
<tr>
<td>Note:</td>
<td>The document represented by oxfordar2-515252_211409.pdf has been deaccessioned from ADS Archives on 08/11/2023. This request was made due to the report being duplicated in the associated project archive (<a href="https://doi.org/10.5284/1106759">https://doi.org/10.5284/1106759</a>)</td>
</tr>
</tbody>
</table>

The FAIR principles. Image © SangyaPundirg
3. Identify where the dataset is stored

- Locate where data is stored
  - Consult archiving procedures to see storage locations
- Remove all copies of data from ADS servers
- ‘Admin’ files (email correspondence, data log etc.) are retained.
3. Document the data removal process

- Update DROID signatures for the collection
- Update manifest file (filenames and checksums)
- Record deletion of files as a process in the Collection Management System (CMS)
Result of deaccessioning work

- Deletion process documented and ‘admin’ data retained (correspondence, data log of dataset prior to deletion, manifest file documenting dataset after deletion)
- Webpages that display original collection level metadata, explain what has happened to the resource and if it is available elsewhere.

Council for British Research in the Levant Research Reports

Introduction
Metadata
Usage Statistics

Data copyright © Council for British Research in the Levant unless otherwise stated

This is a landing page for doi:10.5284/1053684 (ADS archive arch-3514-1)

This dataset has been deaccessioned at the request of the depositor. The ADS has no files to download for this collection

For more information on CBRL publications, please visit the Council for British Research in the Levant Website.
Thoughts for the future …

- Removal of sensitive data?
- Embargoes?
- Previous versions of datasets?
Thank you