An Overview of ADS Systems

ClfA GeoSIG Workshop
Birmingham 2015

Louisa Matthews
Collections Development Manager

http://archaeologydataservice.ac.uk
Overview

• History – why archive?
• Behind the scenes at the ADS
• Geophysics consultation 2010
• Background to charging
• How to calculate costs
The ADS: Who we are and what we do

- Founded 1996
- Department of Archaeology, University of York
- Collections
  - 1,100,000 metadata records
  - 34,535+ unpublished fieldwork reports
  - 700+ rich archives
- Guides to Good Practice
- DPC Decennial Award 2012
Our original philosophy

The ADS supports research, learning and teaching with freely available, high quality and dependable digital resources. It does this by preserving digital data in the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, it provides technical advice to the research community, and supports the deployment of digital technologies.
Why were we worried about digital archives?

“The rate of change in computing technologies is such that information can be rendered inaccessible within a decade...”

- Joint Information Systems Committee: Why Digital Preservation?
Problems

Hardware failure / obsolescence

“...erm, does anyone have a floppy disk drive...?”

Format failure / obsolescence
Lack of Documentation

“Have we got the box with the picture on?”

Access Failure

“Dear former employer / supervisor,

I need my project from a few years ago as my laptop met with an accident, can I have a copy please...?”
How we enact the vision

How do you know you can trust the way we preserve our data?

- We hold a Data Seal of Approval
- We follow the Open Archival Information System (OAIS) reference model

Datasealofapproval.org
Our archive policies

Common to traditional archives:
- A collections policy
- Selection and retention, review and disposal
- Sensitive data DPA/FoI requests
- Human Tissue Act 2004
- Licensing and copyright
- Terms and conditions of use
- Charging policy

PLUS:
- Preservation Policy
- ADS Preservation Policy
- ADS Repository Operations
- ADS Ingest Manual
- Copyright Infringement Policy
Collections Policy

• Accept **all archaeological data types** (but in preferred formats)
• **Do not impose** chronological, geographical or thematic limitations
• Depositor usually should have a **UK connection**
• Collaborate with **local and national agencies**
• Collaborate with **funding bodies**
Deposit Evaluation

• Intellectual content & potential interest in their **re-use**
  - Who will be interested in the data in the future?

• **Viability** of data management, preservation, and distribution
  - Assessment of data structure and format
  - Nature and completeness of documentation
  - Technical and cost benefit issues

• Other suitable archives?
  • - Might the data be better deposited elsewhere, with better suited expertise or re-use potential?
  • - Prevent duplication of archiving efforts within archaeological community
Deposit Requirements

- **Authority** to deposit the data
  - Permissions obtained in terms of Rights over the data.
  - Able and willing to grant repository a licence to disseminate the data.

- Material is ‘**complete**’
  - ADS: finished project archives that will not be added to.
  - Individual digital entity is complete, i.e. not draft a version of paper.

- Digital form in preferred **file format** – consult repository websites
  - Most common file formats accepted
  - Open formats preferred

- Sufficient project documentation and file metadata
Dataset-level metadata for data collections deposited with the ADS

Please complete this form as fully as possible with details of your dataset. This data will form the basis of an entry about your dataset in the ADS Catalogue, and underpins the computerised searching process that allows users to discover and retrieve information.

1. **Title** - the title (and any alternatives) for the dataset.

2. **Introduction** - a brief summary (200-300 words) of the main aims and objectives of the project. This will appear on the first page of the interface for your dataset.

3. **Overview** - a summary description (200-300 words) of the content of the dataset. This will appear on the Overview page of the interface for your dataset.

4. **Interface Images** - the file names, related caption and copyright information if appropriate for two images to illustrate the Introduction and Overview pages. The image files must be supplied to ADS with this Project-level metadata form.

5. **Subject** - suggest keywords for the subject content of the dataset. If possible, please use existing documentation standards (e.g. The RCHME thesaurus of Monument Types, the MDA Archaeological Object Name thesaurus) and indicate which standard you are following. If you use a documentation standard unique to your organisation, it would be extremely helpful if you could send a copy of it with your dataset.

---

**Preparing Datasets for Deposit**

**ADS CHECK LIST**

**PLANNING STAGE**
- Create a Data Management Plan (DMP)
- Establish the data types you will produce during your project
- Decide on the file formats you will use
- Contact the ADS about your deposit or consider using ADS easy
- Ascertain archive costs and include them and preparation time in tender/funding bid
- Determine all metadata requirements

**DATA COLLECTION STAGE**
- Put secure backup systems in place (remember to regularly backup data)
- Ensure all project members are aware of DMP and are following the same procedures
- Collect data (remember to regularly review data)
- Use file structure and naming rules determined in DMP
- Collect required file-level metadata using ADS templates
- Collect and record any licenses/copyright required

**POST ANALYSIS STAGE**
- Carry out selection and retention strategy
- Tidy data – check file-level metadata, file formats, file structure, file names
- Prepare any accompany documentation – i.e database relationships, code lists
- Prepare project-level metadata
- Sign deposit license
- Transfer dataset to ADS
1. **Traditional deposit**, one to one for large datasets
2. **ADS-easy** for small to medium data sets
3. **OASIS** for grey literature
Digital Preservation Standards

Behind the Scenes at the ADS

- Ensure the multiple and regular backups and the renewal of storage media
  - 23 Virtual Machines
  - Tape backup at University of York
  - Deep Store – UK Data Archive
  - Five year hardware rotation strategy

- Use data migration strategies

- Follow the Open Archival Information System (OAIS) reference model
Digital Preservation Standards

- **Submission Information Package (SIP):**
  Original data supplied by the producer (creator or depositor) including documentation to facilitate archiving and reuse.

- **Archival Information Package (AIP):**
  Data generated from the SIP and the long-term Preservation package managed within the OAIS. Including administrative, technical and reuse documentation.

- **Dissemination Information Package (DIP):**
  Data generated from the SIP/AIP and made available to consumers (users) including documentation to facilitate use.
Submission Information Package (SIP)

- Virus check
- Media and file readability check
- Data resource integrity check
- Documentation completeness check
- Data validation and consistency checks
- Web interface text check
- Copy to data server
- Authenticate original version
- Replace spaces with underscores
- Log details of SIP in Collections Management System
- Create checksums
- Run Droid to generate file level metadata
- Store licence in AIP directory
- Acknowledge receipt of data
- Store original media
Archival Information Package (AIP)

- Select **preservation** and **dissemination** file formats
- Develop a **conversion** plan
- Convert the files
- Validate file conversion
- **Metadata** update
- Create and store **checksums** for the AIP
- Submit AIP for checking

© Digital Preservation Business Case Toolkit
Dissemination Information Package (DIP)

- **Convert** the files
- **Validate** file conversion
- Create web **interface**
- Allocate permanent urls / **DOIs**
- Pre-Release interface
- Make any depositor changes
- Release Archive
- Publicise Archive
How to we manage all this?

- The Collections Management System
- Knowledge Base

The Collections Management System (CMS) is a tool used by organizations to manage their collections. It provides a way to organize and track the items in a collection, as well as to document the provenance, condition, and other relevant information about each item.

Knowledge Base (KB) is a repository of information that is used to support the management of collections. It contains standard operating procedures (SOPs), guidelines, and other information that is useful for managing collections.

The Data Procedures are our in-house manuals. Every effort should be made to ensure that they are kept up-to-date, so everyone is doing the same thing. If you feel that a Procedure document needs overhauling/significant editing, raise the issue at the nearest CATS meeting and time will be allocated. Each Procedure document has a document control section at the start and comments etc. should be logged there.

- The template for ADS Data Procedures
- ADS Data Procedures: Audio
- ADS Data Procedures: CAD and vector graphics
- ADS Data Procedures: Databases
- ADS Data Procedures: Geophysics (including GPR)
- ADS Data Procedures: GIS
- ADS Data Procedures: Moving Images
- ADS Data Procedures: Photogrammetry (NEEDS REVISING)
- ADS Data Procedures: P3M & RTI (DRAFT)
- ADS Data Procedures: Raster Images
- ADS Data Procedures: Spreadsheets
- ADS Data Procedures: Statistics
- ADS Data Procedures: Harris Matrices (DRAFT)
- ADS Data Procedures: Binary and Plain Text
- ADS Data Procedures: Virtual Reality (NEEDS REVISING)
- ADS Data Procedures: LIDAR (NEEDS REVISING)
- ADS Data Procedures: Websites
What did we learn in 2010?

• There is general support for the digital archiving of geophysical surveys

• Archaeologists are keen to access archived data - this is still the case – but there isn’t enough available

“I haven’t reprocessed any geophysical data - in the end I had trouble getting access to raw data (so encouraging deposition with ADS would be, in my opinion, an excellent idea!) Instead I have had to rely on reinterpreting older plots from grey-lit and publications etc., which obviously isn’t ideal.”

- P Townend, PhD Candidate, University of Sheffield

“I have used a variety of different sources to get the data. These include published journals and books, local planning authority websites... the main port of call for me in getting the difficult ones is the HER, although I have increasingly been using the ADS in order to obtain them. However, it is frustrating to know a survey has been carried out, but not be able to find it on the ADS”

- J Lyall, Geophysicist
What did we learn in 2010?

- Principle barriers to archiving identified were
  - Lack of consistent application i.e. competitive advantage to those not archiving
  - Cost

Image: Gordonhunter
What did we learn in 2010?

• Another barrier was the time taken to compile metadata.

• Better visibility of archived data as a geophysics collection might encourage more deposition (peer pressure).

• Differing opinions on what should be preserved.
How did we act on what we learned?

- Encouraged incorporation of digital archiving into standard briefs / WSIs
- Consistent costs *between* contractors
  - Identification of survey area as a ‘constant’ (the 10ha charge).
    - Could we identify other constants today?
- ADS-Easy and costing calculator (more later).
## Background to archiving costs

<table>
<thead>
<tr>
<th>Traditional Archives</th>
<th>Charging policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management and Administration</td>
<td>1 day Management = £300. 1 day admin = £280.</td>
</tr>
<tr>
<td>Ingest (cf. SIP checklist, AIP checklist)</td>
<td>11-100 files = 3 days = 3x280 = £840</td>
</tr>
<tr>
<td>Dissemination (DIP)</td>
<td>Included in above for ‘download only’.</td>
</tr>
<tr>
<td>Storage</td>
<td>£0.50 per MB</td>
</tr>
<tr>
<td>Refreshment</td>
<td>£0.30 per MB</td>
</tr>
<tr>
<td>Overheads</td>
<td>25%</td>
</tr>
<tr>
<td>VAT</td>
<td>20%</td>
</tr>
</tbody>
</table>
The ADS Charging Policy is complicated, because of the complicated nature of what we do. The costing calculator rationalises the variables for ‘straightforward’ projects.

• Admin and management became a flat fee of £200 rather than per day charge.

• Per file/unit charge (Geophysics based on units of 10 hectares) for storage, refreshment etc.
  • Risks underwritten by us.

• Ingest and dissemination – still variable dependent on complexity of file level metadata and file handling.
• For UK contracting units, in the absence of specific guidance (i.e. within a project brief) the ADS recommendation is to deposit a geo-rectified TIFF of high quality and a pre-processed composite file(s) of raw data.

• Check list:
  • Geo rectified TIFF
  • Unprocessed data files – grids?
  • OASIS Number (if report uploaded) or report plus collection level metadata
  • File level metadata
Costs – the costing calculator

• Costs for Geophysics data only (Costing Calculator ‘Traditional Methods’ fee) and including VAT.
  • 10ha = £336
  • 20ha = £432
  • 30ha = £528
  • 40ha = £624
  • 50ha = £720
  • 60ha = £816
  • 70ha = £912
  • 80ha = £1,008

  Over 80ha please contact us directly.
Things to remember

• Archiving of your report via OASIS is free - don’t resubmit with your archive, just cite the OASIS number.

• We can embargo sensitive archives (e.g. pre-application work) so you can submit and cost at the time of the project, but release once the project is in the planning system. We can’t embargo indefinitely though.
How can we build on what we have learned?

• Pioneering technology often means things aren’t right first time, so review and reassessment is important (we’ll be here again in another 5 years!)

• Can we work together to identify commonalities in the data capture process that might simplify the archiving process?
An Overview of ADS Systems

ClfA GeoSIG Workshop
Birmingham 2015

Louisa Matthews
Collections Development Manager

http://archaeologydataservice.ac.uk