Archiving and Repositories

Holly Wright
Data archiving in the digital lifecycle

Where data archiving fits

- At the start of the project - identifying the archive or repository to be used
- Before starting data capture – check archive requirements
- During the project – documentation + deposit
- On completion – deposit
- After the project – enabling access and reuse
Why Deposit?

1. Ensure Preservation
2. Provide Access
3. Professional Recognition
4. Follow Professional Standards
5. Meet Governmental Requirements
6. Meet Funding Agency Requirements
Data Deposit Requirements

AHRC

- Submission of a ‘Technical plan’ replaces the Technical Appendix as of 2012

- is essential:
  “where digital outputs or digital technologies are an essential part to the planned research outcomes”.

This should give a summary of those outputs, explain the technical methodology, technical support / experience, and address preservation, sustainability and use.
Archiving

What happens to data once its deposited?
Behind the Scenes at the ADS

- Follow the Open Archival Information System (OAIS) reference model
  - International ISO standard 14721

- Ensure the multiple and regular backups and the renewal of storage media
  - Virtual Server
  - Tape backup at University of York & Hull
  - Deep Store
  - Five year rotation strategy

- Use data migration strategies
Open Archival Information System (OAIS) reference model
Open Archival Information System (OAIS) reference model
Deposit Evaluation

- Intellectual content & potential interest in their re-use
- Viability of data: management, preservation and dissemination
- Are we a suitable archive?
- Authority to deposit the data
- Material is ‘complete’
- Digital form in preferred file format – consult repository websites
- Sufficient project documentation and file Metadata
Submission Information Package (SIP)

- **Virus check**
- Media and file **readability** check
- Data resource **integrity** check
- Check file **formats** suitable for deposit
- Documentation **completeness** check
- Data **validation** and **consistency** checks
- Web interface text check
- Copy to data server

© Digital Preservation Business Case Toolkit
Submission Information Package (SIP)

- 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
- Scan paper documentation
- Acknowledge receipt of data
- Store original media
Archiving
Archival Information Package (AIP)

- Check **licence**
- Check copyright and confidentiality clearance
- **Consistency** checks
- Selecting 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
### Archiving

## Information Packages

<table>
<thead>
<tr>
<th>Delivery</th>
<th>Preservation</th>
<th>Presentation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png or Joint Photographic Expert Group .jpg</td>
<td>Any EXIF &amp; IPTC metadata will also need preserving</td>
</tr>
<tr>
<td>Portable Network Graphics .png</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>Joint Photographic Expert Group .jpg / .jpeg</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Joint Photographic Expert Group .jpg / .jpeg</td>
<td>*</td>
</tr>
<tr>
<td>Graphics Interchange Format (Compuserve) .gif</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>Bit-Mapped Graphics Format (Microsoft) .bmp</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>PhotoCD .pcd</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>Photoshop (Adobe) .psd</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>CorelPaint .cpt</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>Portable Network Graphics .png</td>
<td>*</td>
</tr>
<tr>
<td>Adobe Digital Negative .dng</td>
<td>Adobe Digital Negative .dng and .tif</td>
<td>Adobe Digital Negative .dng and Joint Photographic Expert Group .jpg</td>
<td>*</td>
</tr>
<tr>
<td>JPEG2000 .jp2 / .jpx</td>
<td>Uncompressed Baseline TIFF v.6 .tif</td>
<td>JPEG2000 .jp2 / .jpx</td>
<td>*</td>
</tr>
</tbody>
</table>
Archiving

Migration

Requires active and continuous management
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 do ADS disseminate data?

All archives are freely available through the web interface.

http://archaeologydataservice.ac.uk/
Dissemination

ADS Resources

1.3 million metadata records

Archives

- 22 Journals and Series
- 35,000+ Grey Literature reports
- 900+ Project Archives
- Six specialist Bibliographies
- 19 Doctoral Theses

Specialised Websites

- England’s Rock Art
- Image Bank
Dissemination

Project Archives

University-led research projects

Major infrastructure programmes
Project Archives

• Introduction
• Overview
• Downloads and/or Search
• Metadata
• Usage Statistics
• Resource Identifiers
Acknowledgements

ARIADNE is a project funded by the European Commission under the Community’s Seventh Framework Programme, contract no. FP7-INFRASTRUCTURES-2012-1-313193.

These training materials are based on teaching materials for research data management in archaeology created by Lindsay Lloyd Smith (2011) in the JISC-funded DataTrain project based at Cambridge University Library with contributions from the Archaeology Data Service.