Historic England

Historic England Case Study: Data Management Plans

Claire Tsang

Archaeological Archives Curator
Claire.Tsang@historicengland.org.uk
@ClaireTsang5

Hugh Corley

Digital Archaeologist
Hugh.Corley@historicengland.org.uk
@Hscorley

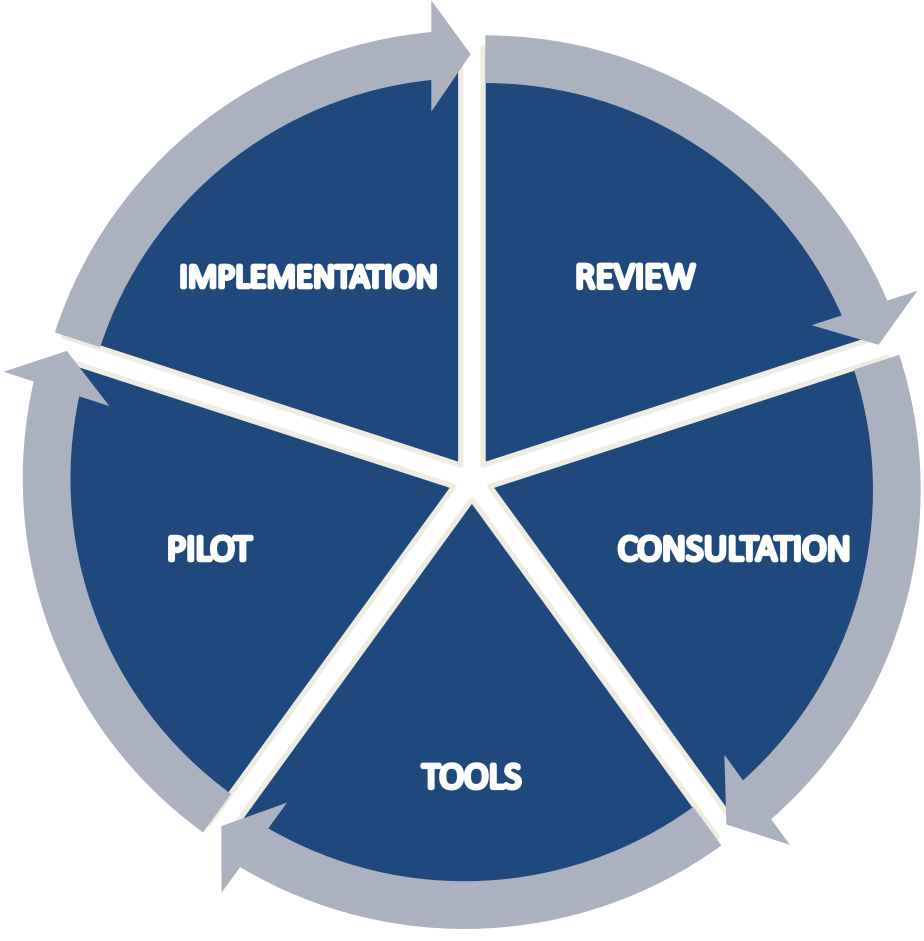
@HE_Archaeology

Historic England

Why?

Historic England

ADAPt: Review & Consultation



- Data holdings
- Internal practice/tools
- External practice/tools

Historic England

HE Data Management Toolkit

Data Management

- Project Management and Archive Checklist
- Data Transfer Agreement
- Data Management Plan
- Evaluation Decision Tree
- Selection and Appraisal Criteria
- *NEW* Data Calculator

File Creation

- File Format Standards
- File-naming Conventions
- Guidance: Preparing reports for deposit
- Guidance: How to create a PDF/A-1a/b
- Guidance: Databases and Spreadsheets

Image Guidelines

- Image Capture guidelines
- Image Flowline
- Image Metadata Index

Folder Templates

Metadata Forms and Library

Historic England

ADAPt Procedures

In summary, staff should:

- Produce a Data Management Plan (DMP)
- Follow the standard file naming procedure
- Store files in appropriate and recognisable locations
- Regularly add metadata to digital files
- Continually review data in accordance with the Selection & Appraisal Policy
- Maintain contact with the Archaeological Archives Team.

ADAPt File-naming Convention

General format

- File format is made up from up to six elements: Identifier-Event-Function-Product-Status-Version
- File name elements are to be divided by a hyphen
- No spaces are to be used within file name elements e.g. CloseSpacesUseCamelCases

Elements

Identifier: 'HE' prefixes project numbers

Event: Codes for events within a project are provided by the Archaeological Archives Team. The format is 'year of event' and a sequential letter.
HE490-2012a *and* ***HE490-2012b*** for an evaluation followed by an excavation. If a file contains data from all events, leave blank.

Function: A short CamelCase description of file contents which describes. Any extended notes in Comments tag in file metadata.

Product: Use for supporting products, i.e. a report diagram.

Status: The stage the file has reached in its life cycle. Set list of terms:
Reports and report products:

- Delete — Temporary copy of file to be deleted
- Backup — Backup copy of file
- Draft — Working copy of a file
- Final — Final copy of file

Databases and Spreadsheets:

- Active — Database - records still being added to/edited
- Reference — Database - records stable

Version: For drafts two digits prefixed with v, i.e. v01 and onward
Interim changes and comments identified by initials after current version number, i.e. v01LHS. Final versions include only the necessary digits.
HE490-2012a-EvaluationReport-draft-v0

Reports

Function examples:

ProjProposal	ArchiveSumRep	*Technique/Material*AssRep
PD	UPD	*Technique/Material*AnlRep
InterimRep	FinalReport	*Technique/Material*AncillaryRep

Report Products

Function: Name of the report the table or image supports
Product: Name of the supporting product
HE5785-2014b-GlassAssRep-Diagram3-Draft-v01

Databases and Spreadsheets

Function examples: Brief description of file contents:

ContextsFinds	RecNumAllocUsed	BoxList
AnimalBone	OrigObjectRecords	AnlSampleRecords

Any extended notes in Comments tag in file metadata
HE490-2012a-DatabaseTables-Active-v01

Images

Site/Original Photography
Identifier-RecordNumber: **HE4584-12523**

Scanned Site Drawings
Identifier-sheet: **HE4584-Sheet103**

Scanned Context/Sample/Small Find Sketches
Identifier-Context/Sample/Small find number: **HE4584-56481**

Specialist Images
Identifer-Technique-RecordNumber: **HE4584-SEM-RecordNumber, HE-HE4584-Xray-RecordNumber**

This document provides a standardised file-naming convention that will make file names more transparent. It will allow everyone to understand, at a glance, what a file contains, where it has come from and what it relates to. Required deviation from the convention should be undertaken in this ethos only.

Historic England

File naming convention

File name elements are to be divided by a hyphen / dash
No spaces are to be used within file name elements

ProjectIdentifier-Function-Product-Status-Version

e.g.HE5785-GlassAssRep-Draft-v01

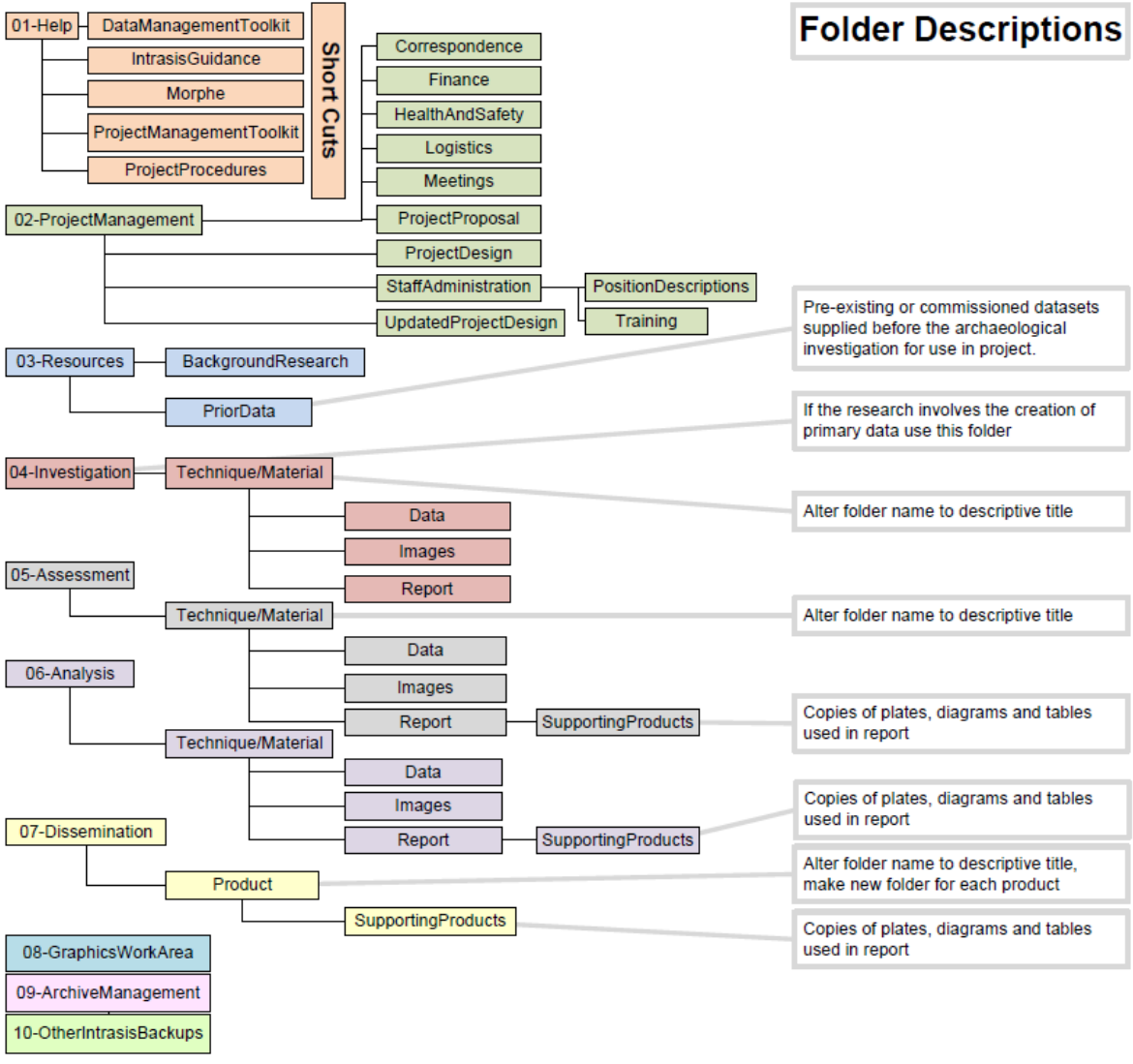
HE5785-GlassAssRep-Diagram3-Draft-v01

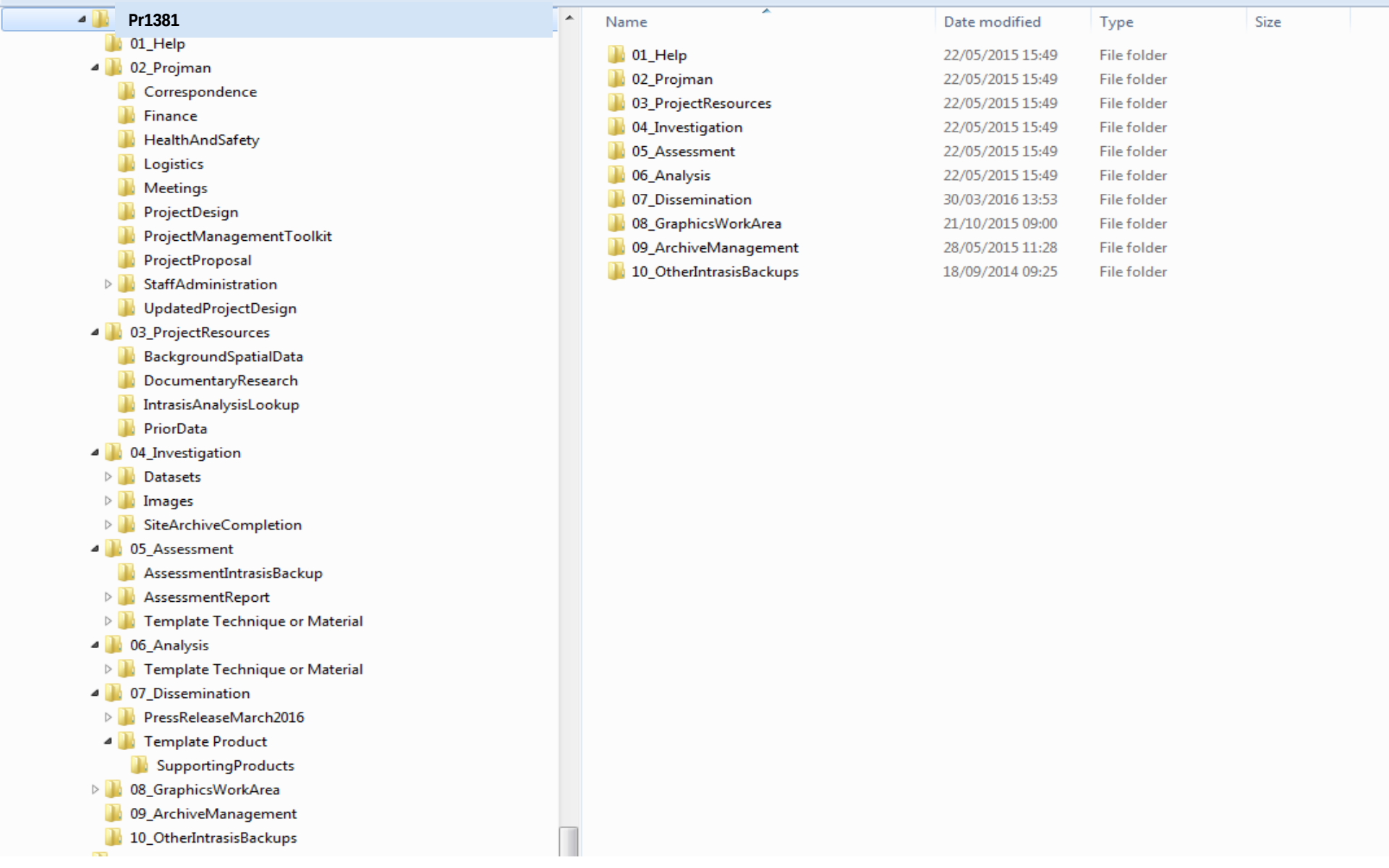
Classification	Creation format	Version	Extension	Archive format	Version	Ext.
Text	Microsoft Word		.doc/.docx	Microsoft Word		.docx
	plain text files	n/a	.txt	plain text files	n/a	.txt
	Hypertext Markup Language		.html/htm	Hypertext Markup Language		.html
	Macro enabled Microsoft Word Document OOXML		.docm	Macro enabled Microsoft Word Document OOXML		.docm
	Rich Text Format		.rtf	Rich Text Format		.rtf

ADS – prefer/recg/accept	Digital Archiving Protocol	Additional Documentation	Comments	Access format	Version	Ext.
Preferred	Preserve Creation format.	software, version and platform		Adobe Portable Document Format (*)		PDF
Accepted	Preserve Creation format	text encoding		Adobe Portable Document Format (*)		
Accepted	Preserve Creation format	Software used in creation, doctype with HTML schema		Adobe Portable Document Format (*)		
Accepted	Preserve Creation format			Adobe Portable Document Format (*)		
Accepted	Preferably convert to .docx files but can preserve creation format	Software and version		Adobe Portable Document Format (*)		

Historic England

Folder Structure




Pr1381
01_Help
02_Projman
Correspondence
Finance
HealthAndSafety
Logistics
Meetings
ProjectDesign
ProjectManagementToolkit
ProjectProposal
StaffAdministration
UpdatedProjectDesign
03_ProjectResources
BackgroundSpatialData
DocumentaryResearch
IntrasisAnalysisLookup
PriorData
04_Investigation
Datasets
Images
SiteArchiveCompletion
05_Assessment
AssessmentIntrasisBackup
AssessmentReport
Template Technique or Material
06_Analysis
Template Technique or Material
07_Dissemination
PressReleaseMarch2016
Template Product
SupportingProducts
08_GraphicsWorkArea
09_ArchiveManagement
10_OtherIntrasisBackups
Name
Date modified
Type
Size
01_Help 22/05/2015 15:49 File folder
02_Projman 22/05/2015 15:49 File folder
03_ProjectResources 22/05/2015 15:49 File folder
04_Investigation 22/05/2015 15:49 File folder
05_Assessment 22/05/2015 15:49 File folder
06_Analysis 22/05/2015 15:49 File folder
07_Dissemination 30/03/2016 13:53 File folder
08_GraphicsWorkArea 21/10/2015 09:00 File folder
09_ArchiveManagement 28/05/2015 11:28 File folder
10_OtherIntrasisBackups 18/09/2014 09:25 File folder

Project Manager	
Project Number	
Project Name	
Author(s)	
Origination Date:	
Reviser(s)	
Date of last revision	
Project stages covered	
Version	
Status	
Summary of Changes	
File Name/Location	
Related Policies	Please document any deviation from ADAPt. List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection/Creation	
Data to be Collected/Created	Specify what data will you collect or create. Indicate what type of data, formats your data will likely be (e.g. Word Documents, Excel Spreadsheets, Intrasis) and volume. Outline and justify any deviation from ADAPt in choice of format and consider the implications of data format and data volumes in terms of cost, storage, backup and access.
	Standard text: Intrasis will be the recording system used on site. Images will be created according to standards set out in ADAPt and the Intrasis Guidance, currently outdating the Recording Manual. All file formats created will meet the standards set out in ADAPt. Drawings will be created by hand on polyester drawing film, using 4H Staedtler Mars Lumograph pencils and scanned in by the Archaeological Archives Team at standards set out in ADAPt and site sketches will be scanned to standards set out in the Intrasis Guidance. *The documentary archive will consist of:* *Text: Various Word Documents, including Project Design, Assessment Reports, Site Archive Completion Report.* *Databases: Intrasis dataset, all data generated during site archive completion and assessment will be held in the Intrasis database.* *Images: Hard copy drawings, digital images, including x-rays, site photography, scanned drawings, c. 200.*
How Data will be Collected/Created	Document standards and methodologies used, quality assurance processes. Document any deviation from ADAPt on version control, file naming, folder structure and usage
	Standard text: The data will be created according Recording Manual, Intrasis Guidance, and ADAPt
Relations	If the data collection was derived in whole or in part from published or unpublished sources, whether printed or machine-readable, please give references to the original material. Please give details of where the sources are held and how they are identified there (e.g. by accession number). If the collection is derived from other sources please indicate whether the data represent a complete or partial transcription/copy and the methodology used for its computerisation. Clearly state whether the data represents a clear enhancement of this previous dataset.
Documentation and Metadata	
Metadata	Document any deviation from ADAPt
	Standard text: Metadata will be created to the standard set out in ADAPt
Documentation	Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it.
Ethical and Legal Compliance	
Data Security Issues	Document if there are any issues such as: Does the data need to be embargoed to protect the site(s)? Is some of the data protected under data protection legislation? Does the dataset contain commercially sensitive data? Detail how any issues are to be dealt with.
Intellectual Property Rights	Document if any data is not Historic England Copyright and how this will be managed, any restrictions on the use of the data. Example: OS data.
	Standard text: The data and reports created by any external specialists will be Historic England Copyright; this will be managed through their contracts.
Data Storage	
Storage and Backup	Document any deviation from the use of the FC network and Site Procedures for backing up data and transfer to network.
	Standard Text: Data will be stored on the FC Network and during excavation will be managed in line with Project Procedures for backing up data and transfer to network.
Access and Security	Document any deviation from the use of the FC network to provide access to data and any security measures required, including the use of Sharepoint.
	Standard Text: Data will be made available to the project team through the FC Network. There are no security issues.
Selection and Preservation	
Preservation Plan	Document what data will be preserved and the repositories for the archive.
	Standard Text: The physical archive, documentary and material, will be transfer to the XXXXX. In addition any a copy of any digital files they wish to receive will also be transferred to them. The documentary archive will contain any hard copy data reports the repository wish to receive. *The digital image and text archive will be archived with the EHA and the Intrasis dataset will be archived with the ADS. All other archiving decisions will be made in discussion with the EHA at UPD stage.*
Data Sharing	
Data Sharing Plan	Document any addition to the dissemination of the data, other than deposition with the repository, HER and publication. Consider how this may be affected by file types, size complexity and sensitivity of the data. Consider how data will be shared with externals during the project (i.e. those who can not access the FC network).
	Standard Text: The data generated from this project will be made publically available through the digital repositories. Awareness of the work will be raised through publication, and documentation with the HER and EHA Excavation Index.
Data Sharing Restrictions	Document any restrictions on the use of the data after project completion.
	Standard Text: There are no restrictions on the use of this data after project completion.
Responsibilities and Resources	
Responsibilities	Document any deviation to the responsibilities set out in ADAPt
	Standard Text: The Project Manager and Archaeological Archives Curator are responsible for ensuring the data management plan is followed.
Resources	Consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.
	Standard Text: Resources required to deliver this plan are covered by standard Historic England resources and the project design. The costs of deposition of the archive are covered by the Historic England storage and digital storage grants.

The Data Management Plan (DMP) is meant to help plan and resource the management of the data generated through research. The plan is meant to be initiated at the project planning stage, however, it is not expected that you will be able to answer all of the questions initially. The DMP applies to the data you create where the Archaeological Archives Team will be responsible for assisting you in depositing your data with a repository. This includes both data specifically for your research or where you have enhanced the data of others. This will help identify any

Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
What data will you collect or create?	**Questions to consider:** - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
How will the data be collected or created?	**Questions to Consider:** - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.
Documentation and Metadata	
What documentation and metadata will accompany the data?	**Questions to consider:** - What information is needed for the data to be to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? **Guidance:** Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.
Ethics and Legal Compliance	
How will you manage any ethical issues?	**Questions to consider:** - Have you gained consent for data preservation and sharing? - How will you protect the identity of participants if required? e.g. via anonymisation - How will sensitive data be handled to ensure it is stored and transferred securely? **Guidance:** Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.
How will you manage copyright and Intellectual Property Rights (IPR) issues?	**Questions to consider:** - Who owns the data? - How will the data be licensed for reuse? - Are there any restrictions on the reuse of third-party data? - Will data sharing be postponed / restricted e.g. to publish or seek patents? **Guidance:** State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.
Storage and Backup	
How will the data be stored and backed up during the research?	**Questions to consider:** - Do you have sufficient storage or will you need to include charges for additional services? - How will the data be backed up? - Who will be responsible for backup and recovery? - How will the data be recovered in the event of an incident? **Guidance:** State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.
How will you manage access and security?	**Questions to consider:** - What are the risks to data security and how will these be managed? - How will you control access to keep the data secure? - How will you ensure that collaborators can access your data securely? - If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems? **Guidance:** If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.
Selection and Preservation	
Which data should be retained, shared, and/or preserved?	**Questions to consider:** - What data must be retained/destroyed for contractual, legal, or regulatory purposes? - How will you decide what other data to keep? - What are the foreseeable research uses for the data? - How long will the data be retained and preserved? - **Guidance:** Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.
What is the long-term preservation plan for the dataset?	**Questions to consider:** - Where e.g. in which repository or archive will the data be held? - What costs if any will your selected data repository or archive charge? - Have you costed in time and effort to prepare the data for sharing / preservation? **Guidance:** Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.
Data Sharing	
How will you share the data?	**Questions to consider:** - How will potential users find out about your data? - With whom will you share the data, and under what conditions? - Will you share data via a repository, handle requests directly or use another mechanism? - When will you make the data available? - Will you pursue getting a persistent identifier for your data? **Guidance:** Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.
Are any restrictions on data sharing required?	**Questions to consider:** - What action will you take to overcome or minimise restrictions? - For how long do you need exclusive use of the data and why? - Will a data sharing agreement (or equivalent) be required? **Guidance:** along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a non-disclosure agreement would give sufficient protection for confidential data.
Responsibilities and Resources	
Who will be responsible for data management?	**Questions to consider:** - Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? - Who will be responsible for each data management activity? - How will responsibilities be split across partner sites in collaborative research projects? - Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners? **Guidance:** Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.
What resources will you require to deliver your plan?	**Questions to consider:** - Is additional specialist expertise (or training for existing staff) required? - Do you require hardware or software which is additional or exceptional to existing institutional provision? - Will charges be applied by data repositories? **Guidance:** Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.

Historic England

The Project Design

Description of the project

Project name

Summary description

Background

Research Aims and Objectives

Business Case

Project scope

Interfaces

Project review

Resources and programming

Project Team structure

Method statement

Stages, Products and Tasks

Ownership

Risk Log

Budget

Data Management Plans

Data management plans should cover the following:

- Description of the data to be collected / created — **WHAT**
- Standards / methodologies for data collection and management — **HOW**
- Ethics and Intellectual Property concerns or restrictions — **LEGAL**
- Plans for data sharing and access — **SHARING**
- Strategy for long-term preservation — **PRESERVATION**

Historic England

In the Beginning…

The Project Design

Description of the project

Project name

Summary description

Background

Research Aims and Objectives

Business Case

Project scope

Interfaces

Sharing

Project review

Resources and programming

Project Team structure

Method statement

How & What

Archiving

Stages, Products and Tasks

Ownership

Archive

Sharing

Legal

Risk Log

Budget

Preservation

Historic England

Data Management Plans

- Data Collection/Creation
- Documentation and Metadata
- Ethical and Legal Compliance
- Data Storage
- Selection and Preservation
- *Data Sharing*
- Responsibilities and Resources

Historic England

The DMP and the Toolkit

What data will you collect or create?

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

How will the data be collected or created?

- Standards and methodologies
- Structure, folder and file naming
- Version control
- Quality assurance

Historic England

The DMP and the Toolkit

What data will you collect or create?

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

How will the data be collected or created?

- *Standards* and methodologies
- Structure, folder and file naming
- Version control
- Quality assurance

Historic England

Success

Data Creator

No Access to…

Database

Identified and data entry resourced

New type of data not previously recorded

Recording system updated (before initiation)

Historic England

Success

Business Case: “This will create a really important resource”…

Products: “The images will not be preserved”

Historic England

DMP 2.0

Data·Collection/Creation¤

Data··to·be·Collected/Created¶ ¤

Specify·what·data·will·you·collect·or·create.·Indicate·what·type·of·data,·formats·your·data·will·likely·be·(e.g.·Word·Documents,·Excel·Spreadsheets,·Intrasis)·and·volume.·Outline·and·justify·any·deviation·from·ADAPt·in·choice·of·format·and·consider·the·implications·of·data·format·and·data·volumes·in·terms·of·cost,·storage,·backup·and·access.¤

Technique¤	Format¤	Quantity¤
FTIR¤	FTIR·machinery·creates·industry· standard·RAW·files.·Currently·the·only·way·to·archive·these·is·as·conversion·to·PDF/A1-A.·¤	¤
Conservation·Photographs¤	Tiff¤	¤
Computed·Radiography¤	Diconde·files¤	¤
Computed·Tomography¤	This·will·create·RAW·data,·images·and·3D·model¤	¤
X-ray·Fluorescence·(XRF)·analysis)¤	XRF·Spectra·Data·saved·as·Comma·Separated·Value,·required·images·can·be·saved·as·tiff¤	¤
Possible·X-ray·Defraction·(XRD)¤	XRD·machinery·creates·industry·standard·RAW·files.·Currently·the·only·way·to·archive·these·is·as·conversion·to·PDF/A1-A.·¤	¤
Human·Remains·Assessment¤	Recorded·into·Intrasis·database¤	¤
Charcoal·Assessment¤	Recorded·into·Intrasis·database¤	¤
Report¤	Microsoft·Word¤	¤

¶

¤

Historic England

Lessons learned

- Excellent tool to ensure consideration of all the implications of data creation
- Opportunity to plan and adapt current tools
- Centralise and document these decisions
- Best practice and a good habit
- You can still lie to yourself…