ads

ARCHAEOLOGY DATA SERVICE

http://archaeologydataservice.ac.uk

ads ARCHAEOLOGY DATA SERVICE

The Archaeology Data Service

- Set up in 1996
- Based at the University of York
- Only accredited UK **digital data** repository for archaeology

Remit:

"Support research, learning and teaching with free, high quality and dependable digital resources"

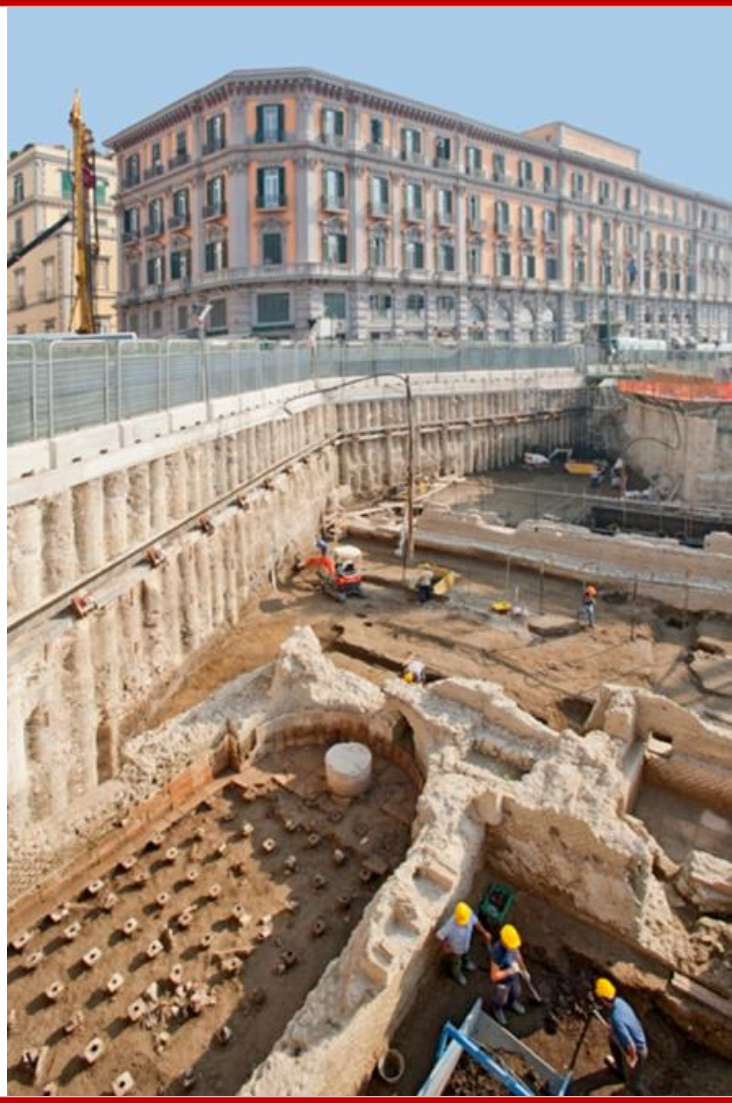
Outline

Part One

- Why is preserving data important
- Behind the scenes at ADS
- ADS Resources

Part Two

- Data Management Planning
- Data Practical



ARCHAEOLOGY DATA SERVICE

Why was ADS established?

- Archaeology is **destructive**
- Comprehensive **records** of field work are imperative
- The use of computers in archaeological fieldwork recording and research has become **routine**

Digital Data

Born Digital

Data created in digital format

Image © Oxford Archaeology (North)

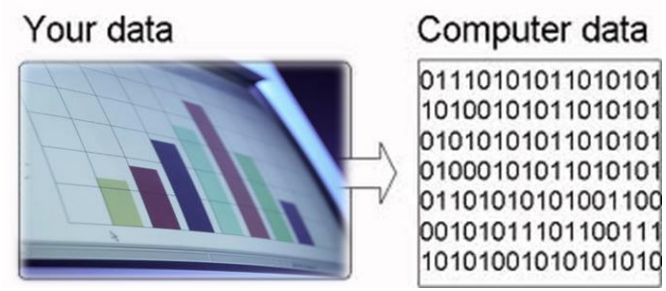
Digitised Data

Hardcopy converted to digital format

Image © State Library of New South Wales 2015

BEWARE: Digital Data is Fragile

Digital data is encoded and requires software & technology to present content



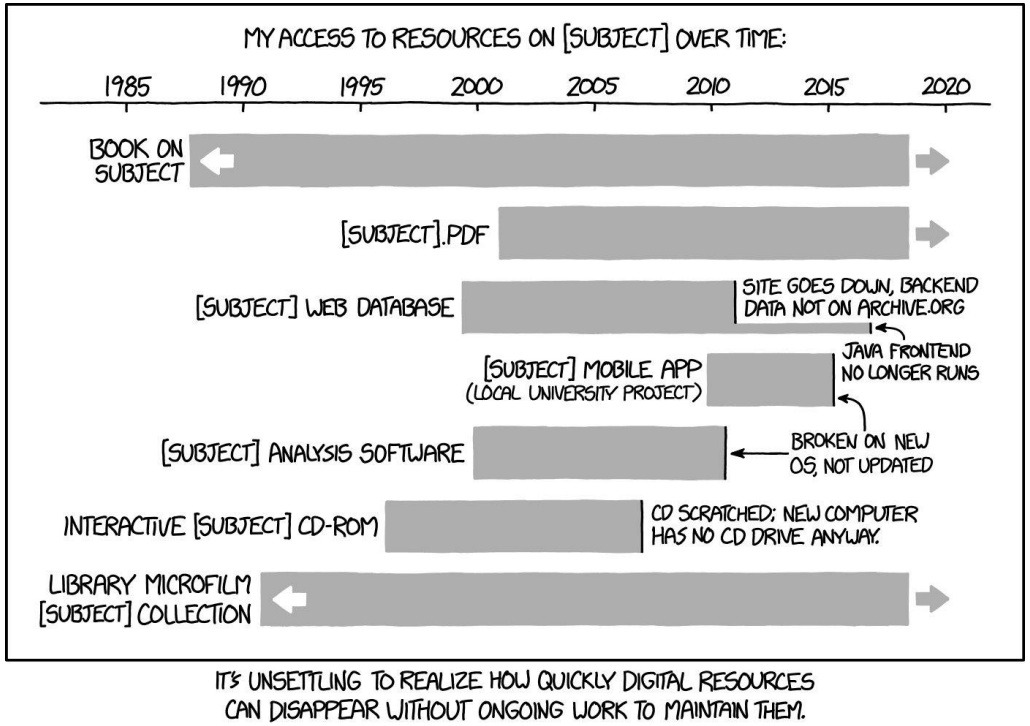


Image Copyright: https://xkcd.com/1909/

Why is Digital Data Fragile?

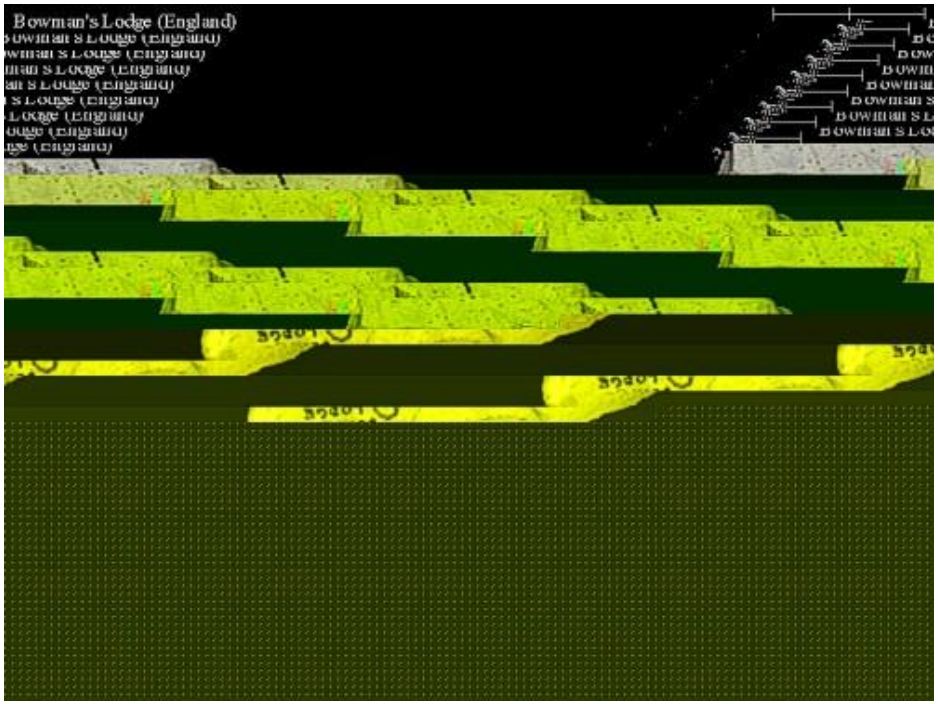
https://youtu.be/8dhp_20j0Ys

Why is Digital Data Fragile?

- **Deterioration of the storage medium**
  - Degrade – Bit rot!
  - Can be easily damaged
  - Can be easily overwritten

Bit Rot cartoon © Digital Preservation Business Case Toolkit

Case Study: ADS



What happened to an image as it was removed from a CD.

What the image was supposed to look like.

Case Study: NASA!

- Lost original Apollo 11 data tapes
  - "original" in directly transmitted from Moon.
- Erased and reused.
- High-quality broadcast versions were found.
- NASA restored the found footage.
- Rereleased in HD for 40th anniversary of Apollo 11.

https://www.nasa.gov/feature/not-unsolved-mysteries-the-lost-apollo-11-tapes

Why is Digital Data Fragile?

- Deterioration of the storage medium
- **Obsolescence of the storage medium**

5.25" Floppy

Bit Rot cartoon © Digital Preservation Business Case Toolkit

Media Types

- Experience rapid change


3.5" Floppy


xD Picture Card


QIC DC600


DC4_120


9-Track Reel


Punch Tape


8" Floppy


DVD-ROM


G2000 Tape


4mm Tape


Smart Media


CD-ROM


Floptical Disk


5.25" Floppy


Travan


5.25" Optical Disk


DG90M Tape


Memory Stick


Rectangular Hole Punch Card


IBM 3480


8mmD-eight


Cassette tape


Click!


Sparq Disk Cartridge


MultiMedia Card


12" Optical Disk


cloud


Jaz Disk


SD Memory Card


Zip Disk


DLT Tape

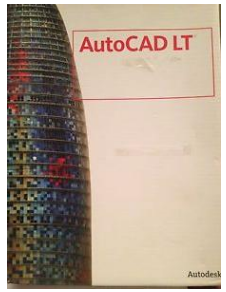

CompactFlash


Ditto Max

Why is Digital Data Fragile?

- Deterioration of the storage medium
- Obsolescence of the storage medium
- **Obsolescence of the software**

5.25" Floppy



Bit Rot cartoon © Digital Preservation Business Case Toolkit

Case Study: NASA again!

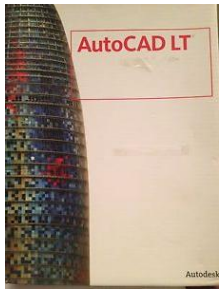
- NASA sent two Viking Landers to Mars in 1975
- Data recorded on magnetic tape
- Climate controlled environment
- In the 1990s they could not decode the formats used
- Had to track down old printouts and retype everything

Photos: Courtesy NASA/JPL-Caltech

Why is Digital Data Fragile?

- Fragility of the storage medium
- Obsolescence of the storage medium
- Obsolescence of the software
- **Obsolescence of the hardware**

5.25" Floppy



Bit Rot cartoon © Digital Preservation Business Case Toolkit

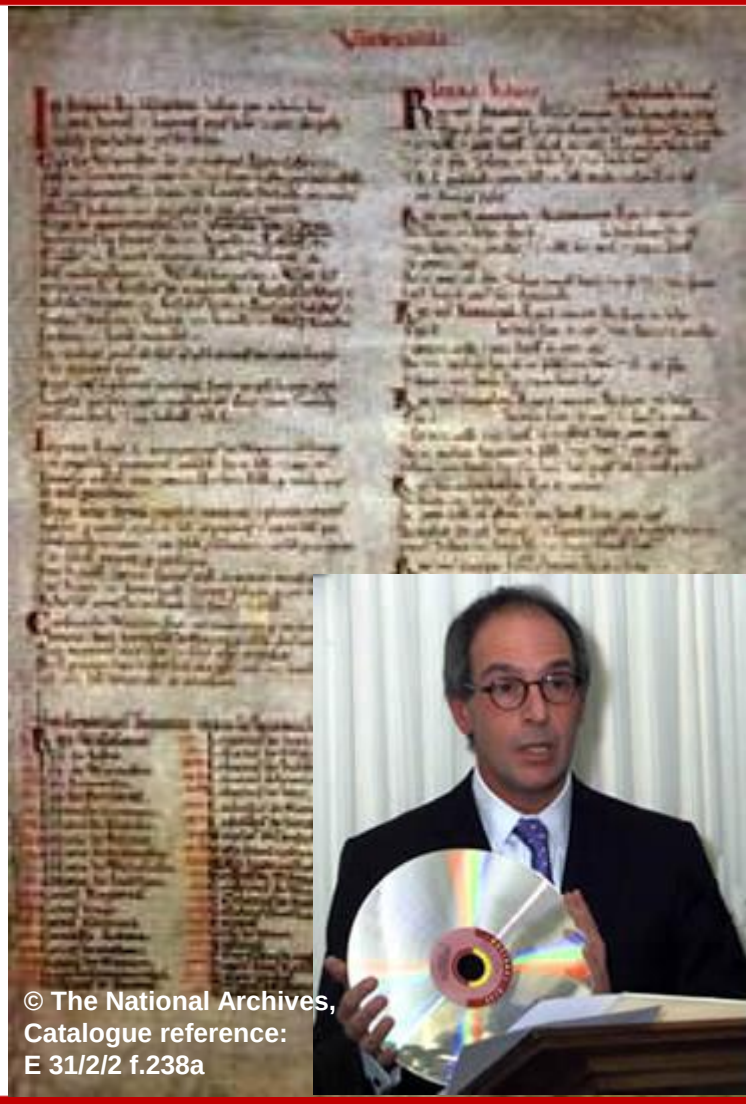
Technology

- Hardware experiences rapid change

ads ARCHAEOLOGY DATA SERVICE

Case Study: BBC Domesday Project

- 1986
- photographs, maps, etc
- 30cm laserdiscs
- BBC Microcomputers
- In 2006 the laserdiscs were **obsolete** as was the hardware
- Rescue projects launched by The National Archives and Leeds University

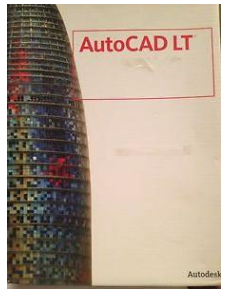


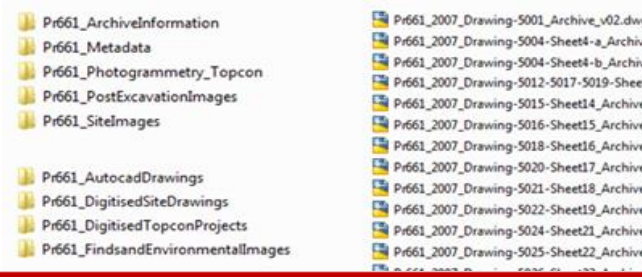
http://www.bbc.co.uk/history/domesday/story

Why is Digital Data Fragile?

- Fragility of the storage medium
- Obsolescence of the storage medium
- Obsolescence of the software
- Obsolescence of the hardware
- **Failure to document the data adequately**

5.25" Floppy





Bit Rot cartoon © Digital Preservation Business Case Toolkit

Case Study: Newham Museum Archaeological Service

Active in archaeological fieldwork across North East London for several decades closed abruptly in 1998 with only a few days notice.

Staff left, computers were sold, a desperate salvage operation began

The result? Two shoe boxes of floppy disks.

Case Study: Newham Museum Archaeological Service

Archive:

- approx. 150 excavations
- 6432 individual files
- 1500 excavation reports
- 700 database files
- 1200 geophysics files
- 200 separate projects

Image © www.digitalbevaring.dk

Case Study: Newham Museum Archaeological Service

Archive:

- approx. 150 excavations
- 6432 individual files
- 1500 excavation reports
- 700 database files
- 1200 geophysics files
- 200 separate projects

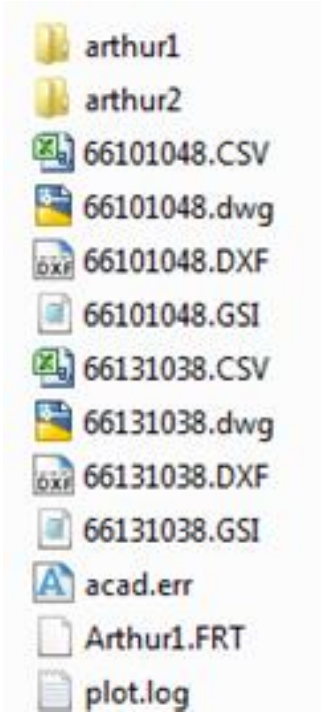
The Problem?

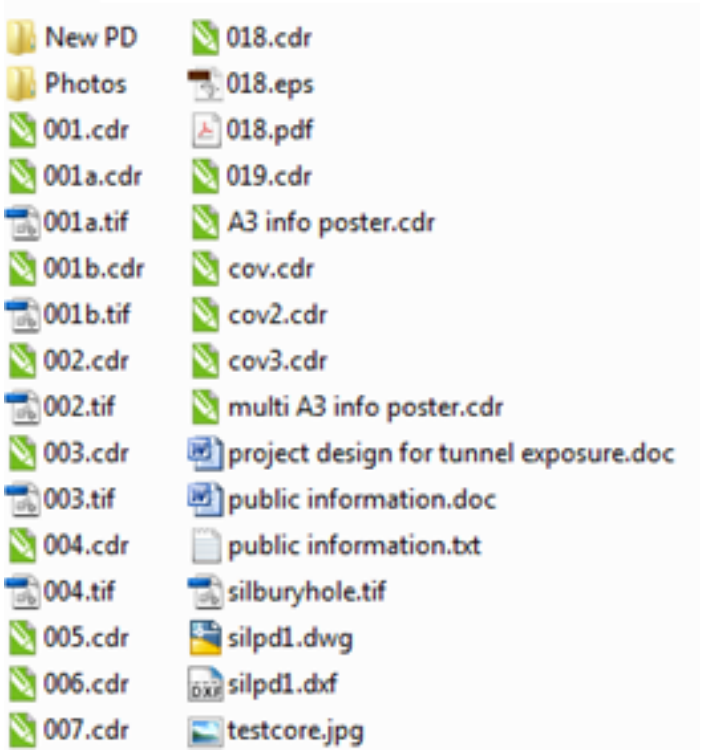
- No link between database/files and excavations
- Missing key to codes
- Little metadata
- Some files simply couldn't even be opened (i.e., CAD)

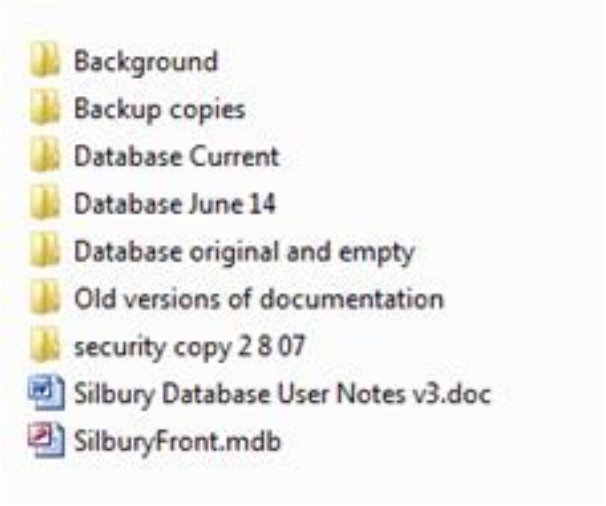
Case Study: Silbury Hill

Silbury Hill from the eastern bank of the Winterbourne © English Heritage

Case Study: Silbury Hill

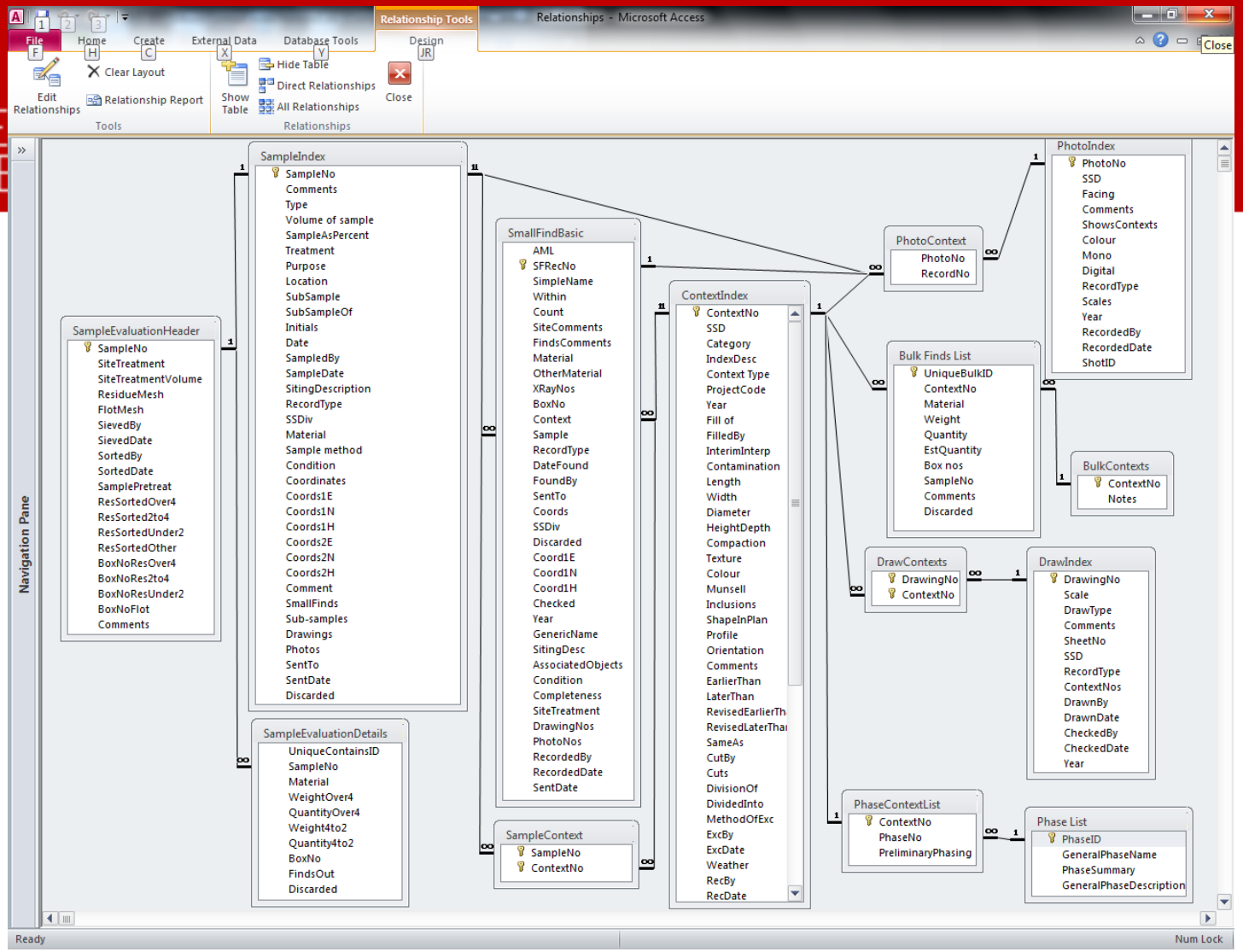






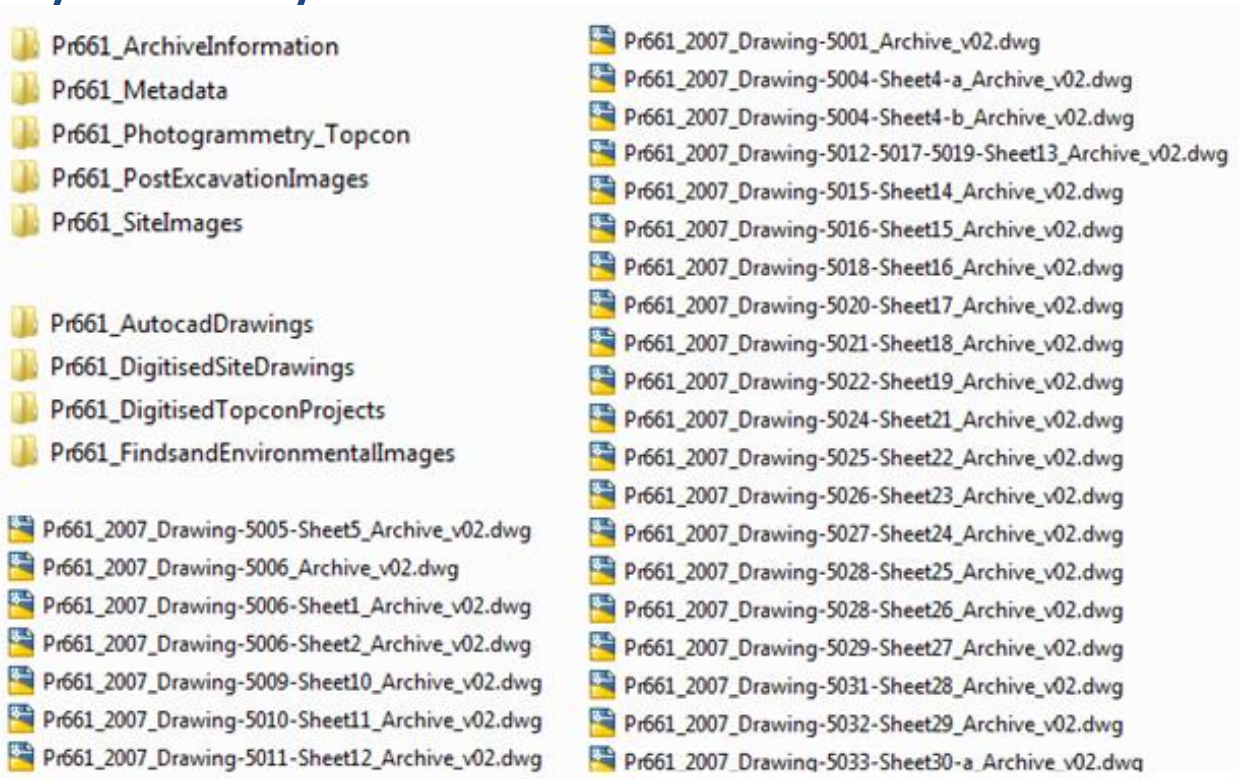
http://archaeologydataservice.ac.uk/blog/2013/08/jenny-ryders-day-of-archaeology-at-the-ads-a-silbury-hill-update/

Case Study: Silbury Hill



http://archaeologydataservice.ac.uk/blog/2013/08/jenny-ryders-day-of-archaeology-at-the-ads-a-silbury-hill-update/

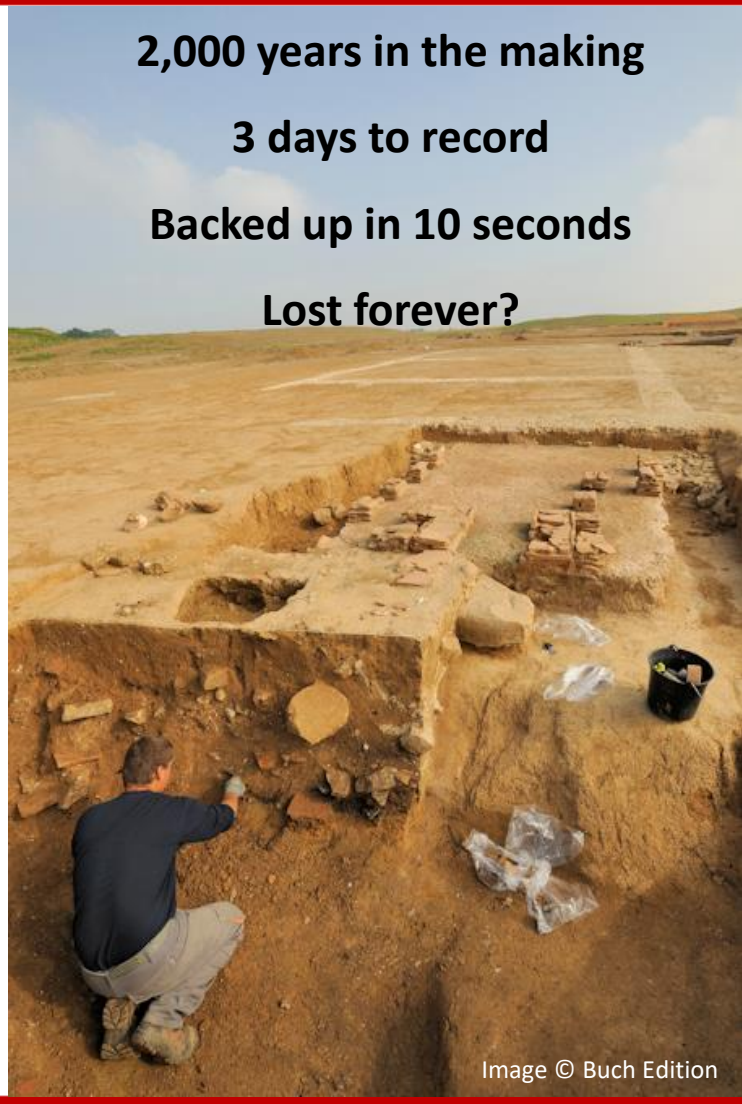
Case Study: Silbury Hill



http://archaeologydataservice.ac.uk/blog/2013/08/jenny-ryders-day-of-archaeology-at-the-ads-a-silbury-hill-update/

Protecting Digital Data

- Recognise data is as **fragile** as the archaeological record we excavate
- **Stop** archiving data as **objects** rather than computerised information



Protecting Digital Data

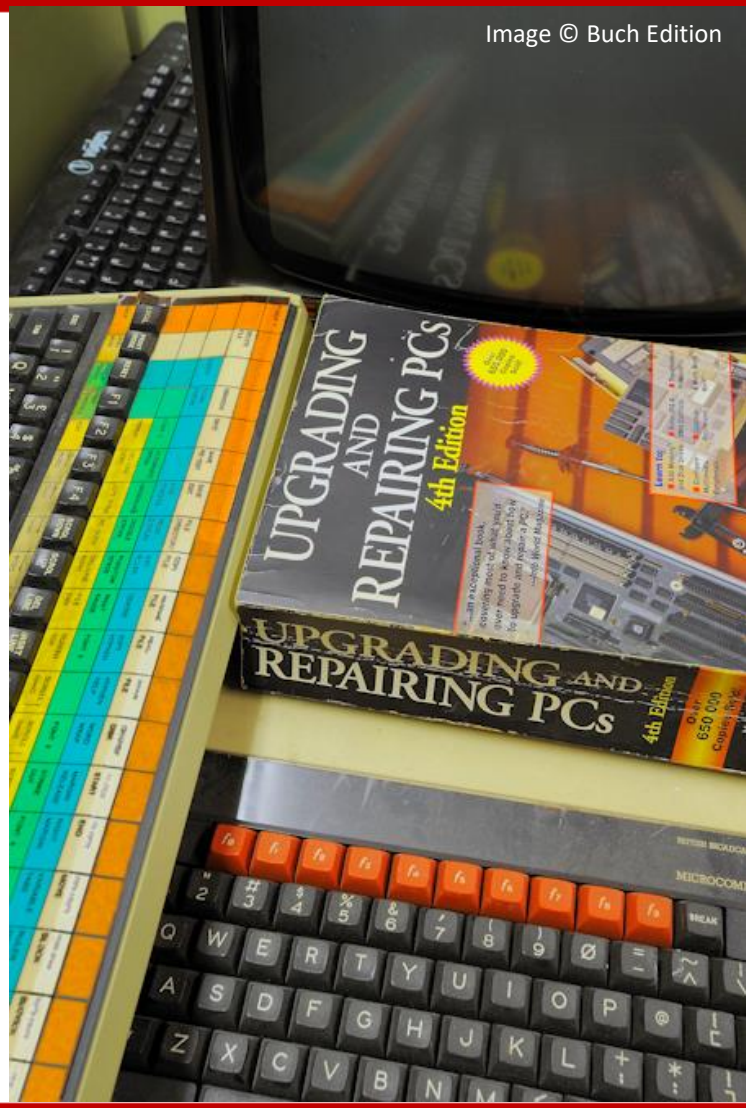
- Recognise data is as **fragile** as the archaeological record we excavate
- **Stop** archiving data as **objects** rather than computerised information
- Create **Data Management Plans**
- Professionally **archive** digital material

ADS Role: Digital Preservation

3 Methods

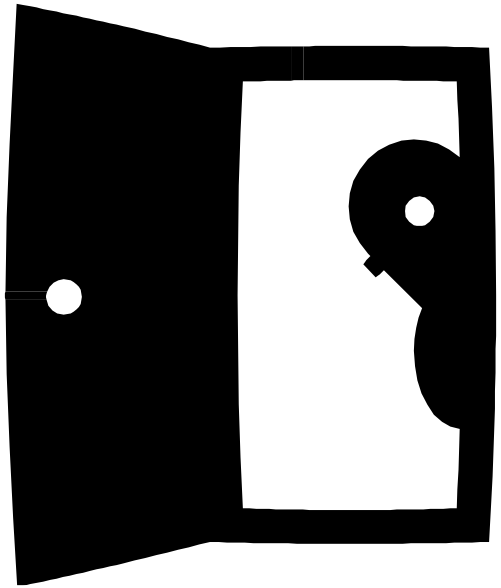
- The Hardware Museum
- The Software Emulator /Virtualisation
- Migration

NB much more intervention is needed than conventional archives



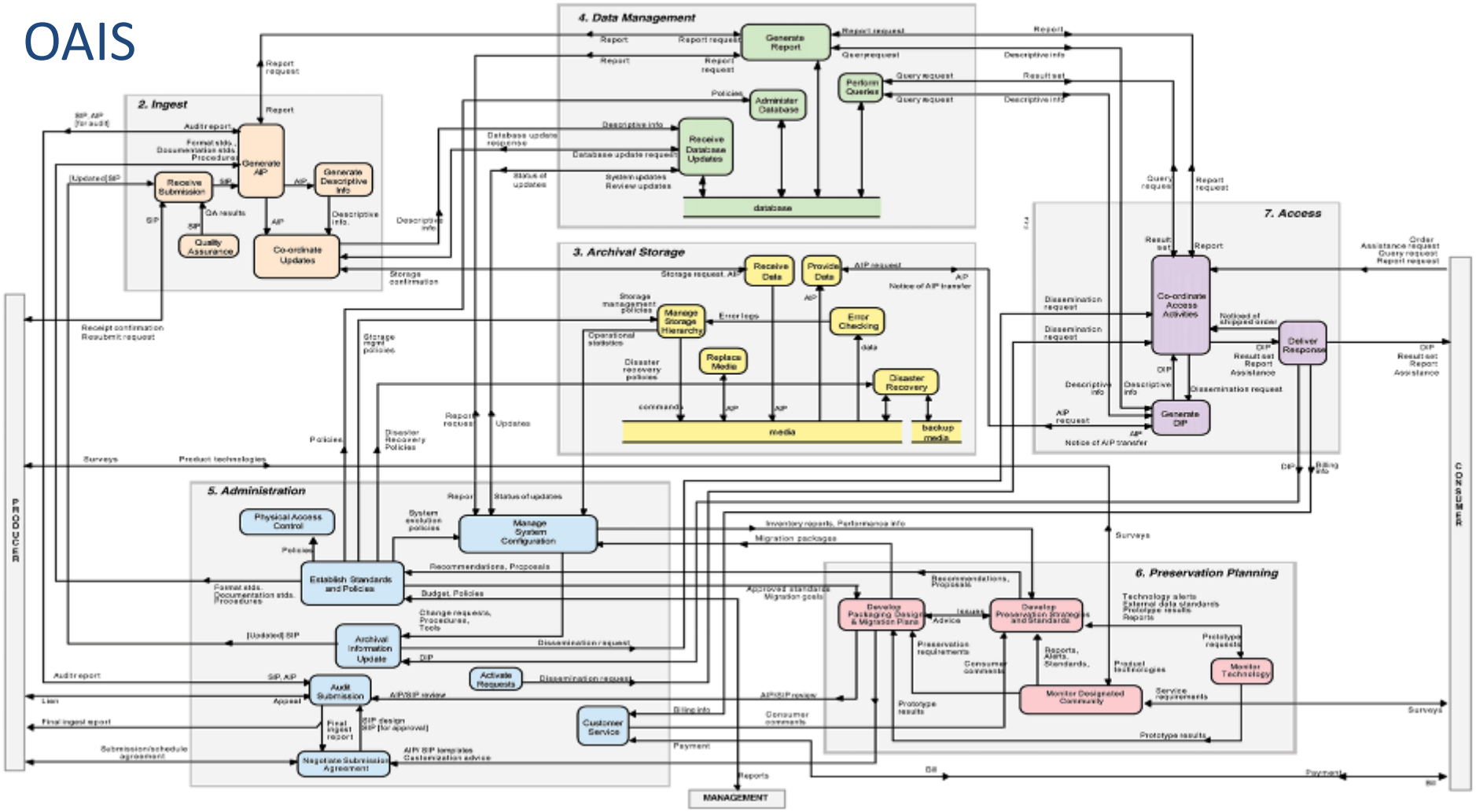
Behind the Scenes at the ADS

- Use data **migration** strategies
- 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
  - 30+ Virtual Servers
  - Tape backup at University of York & Hull
  - Deep Store



ads
ARCHAEOLOGY
DATA SERVICE

OAIS



Open Archival Information System (OAIS) reference model

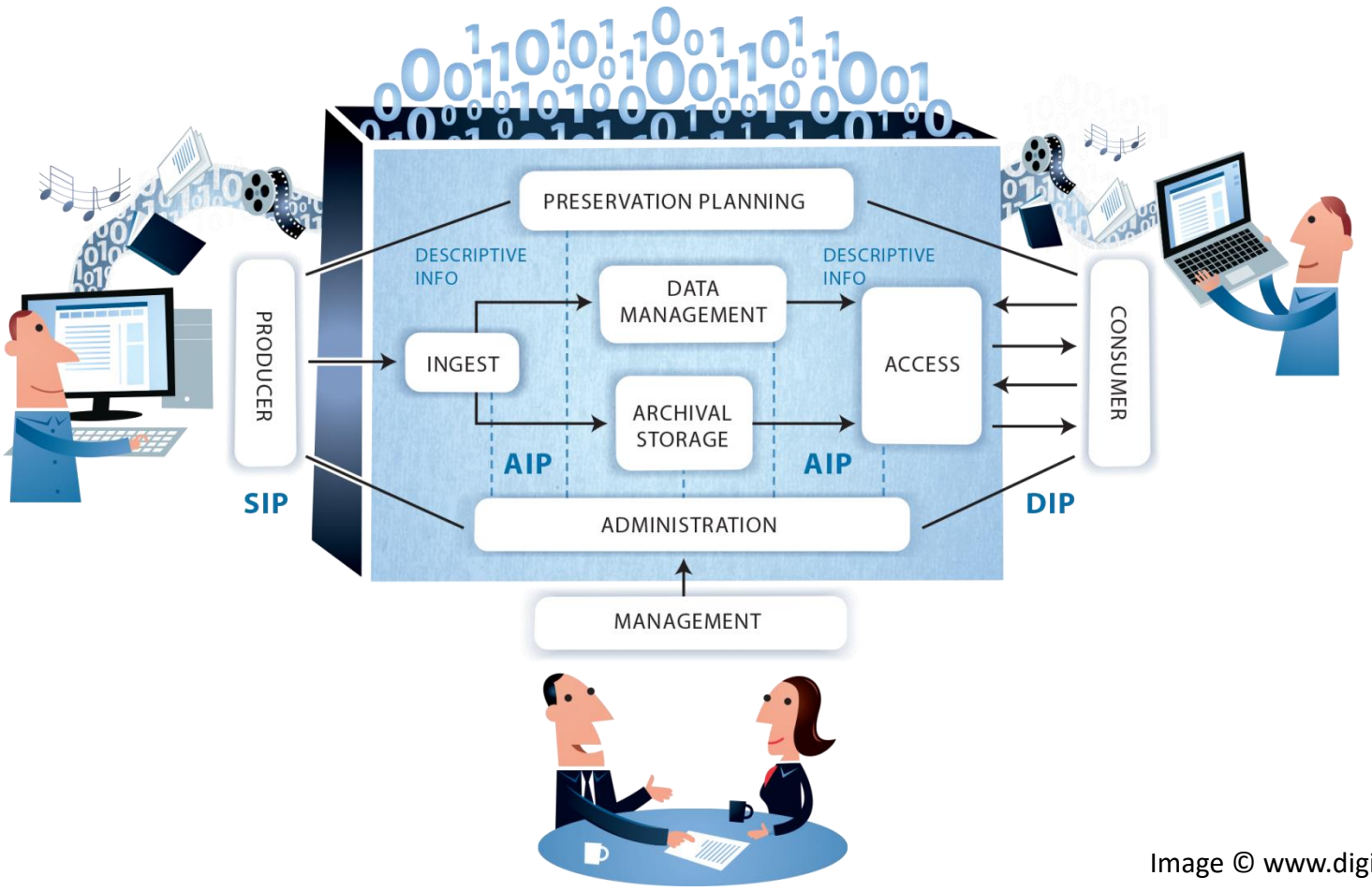


Image © www.digitalbevaring.dk

How do ADS disseminate data?

Everything we archive is **freely** available through the web interface.

1.3 million metadata records

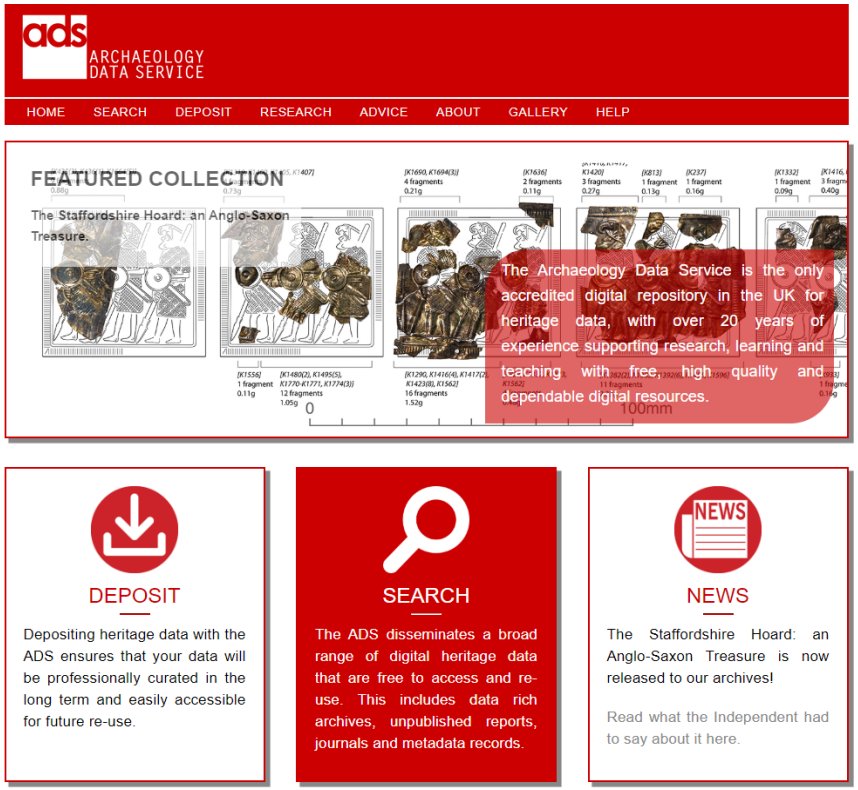
250,000+ bibliographic records

- 37 full text Journals
- 100+ full text monographs
- 56,000+ reports

3000+ data rich archives

- 22+TB data

https://archaeologydataservice.ac.uk/about/annualReports.xhtml



Outline

Part Two

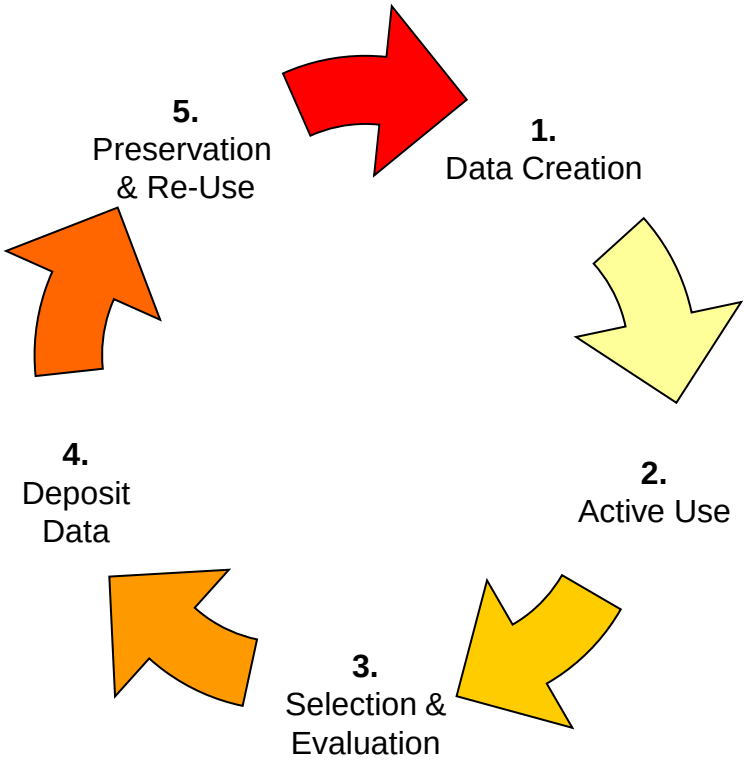
- **Data Management Planning**
- Data Exploration Practical

Data Management Plan

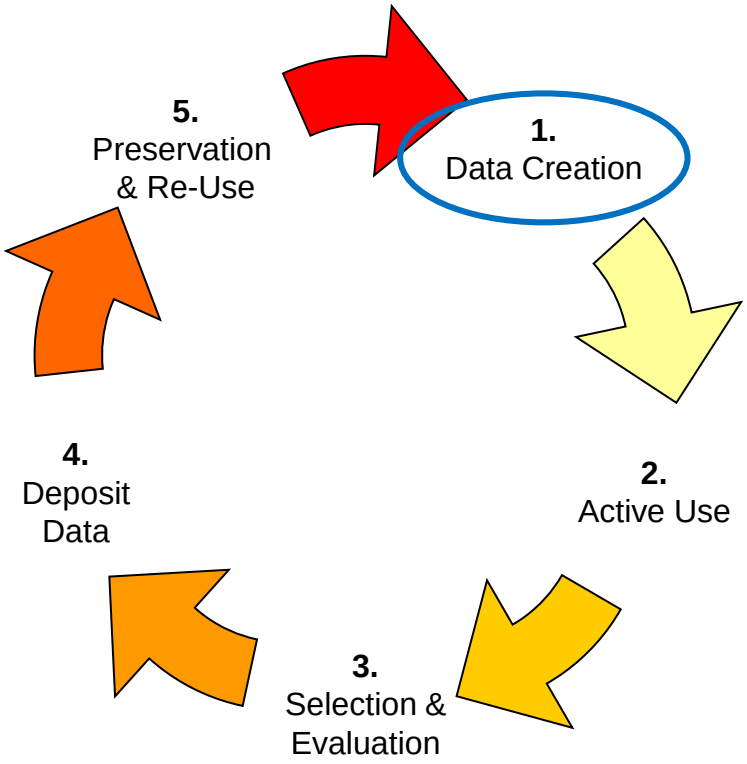
Data Management Plan

Informs all other stages

Data Cycles & Management Plans



Data Cycles & Management Plans

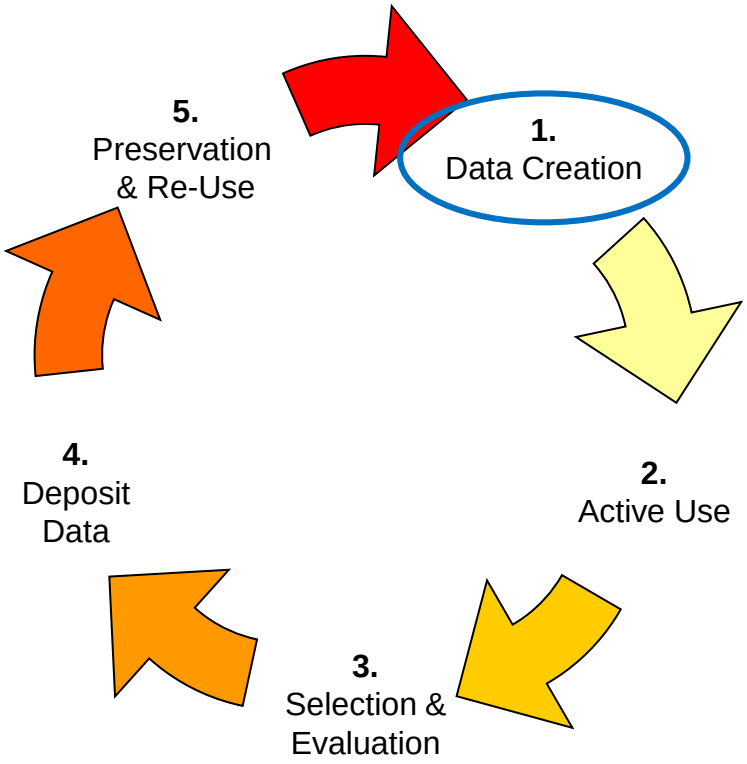


- What data will I produce?
  - **Text documents**
  - **Artefact analyses**
  - **Sample analyses**
  - **Survey data**
  - **Drawings**
  - **Photographs**
  - **Recorded interviews**
  - **Etc..**

http://www.jiscdigitalmedia.ac.uk/infokit/file_formats/digital-file-formats

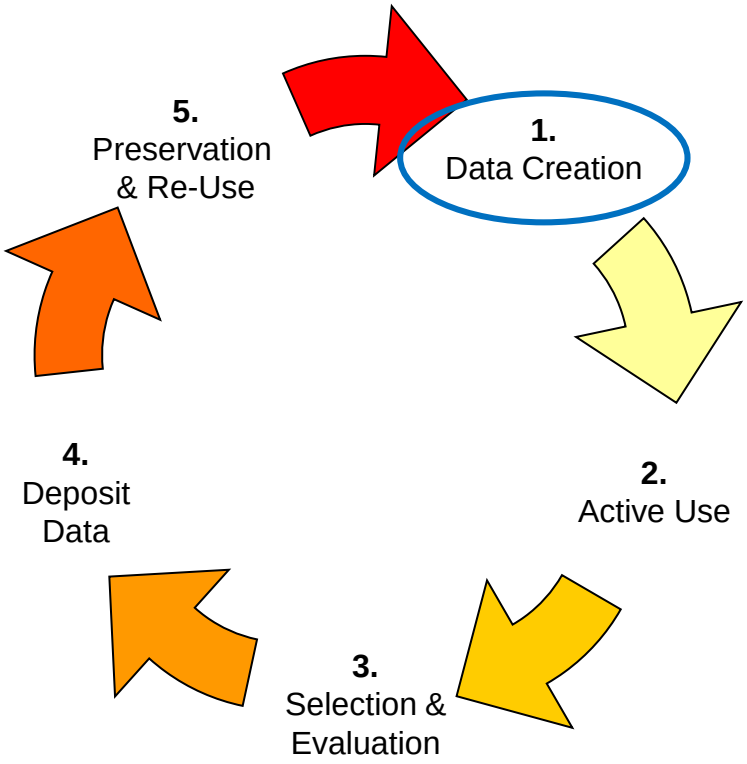
- **Who will owns data? Do I need permissions?**

Data Cycles & Management Plans



- What data will I produce?
- How will I organise the data?

Data Cycles & Management Plans

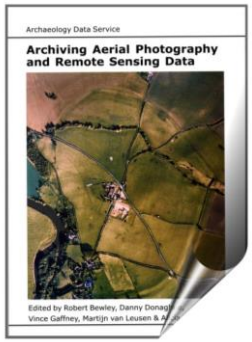


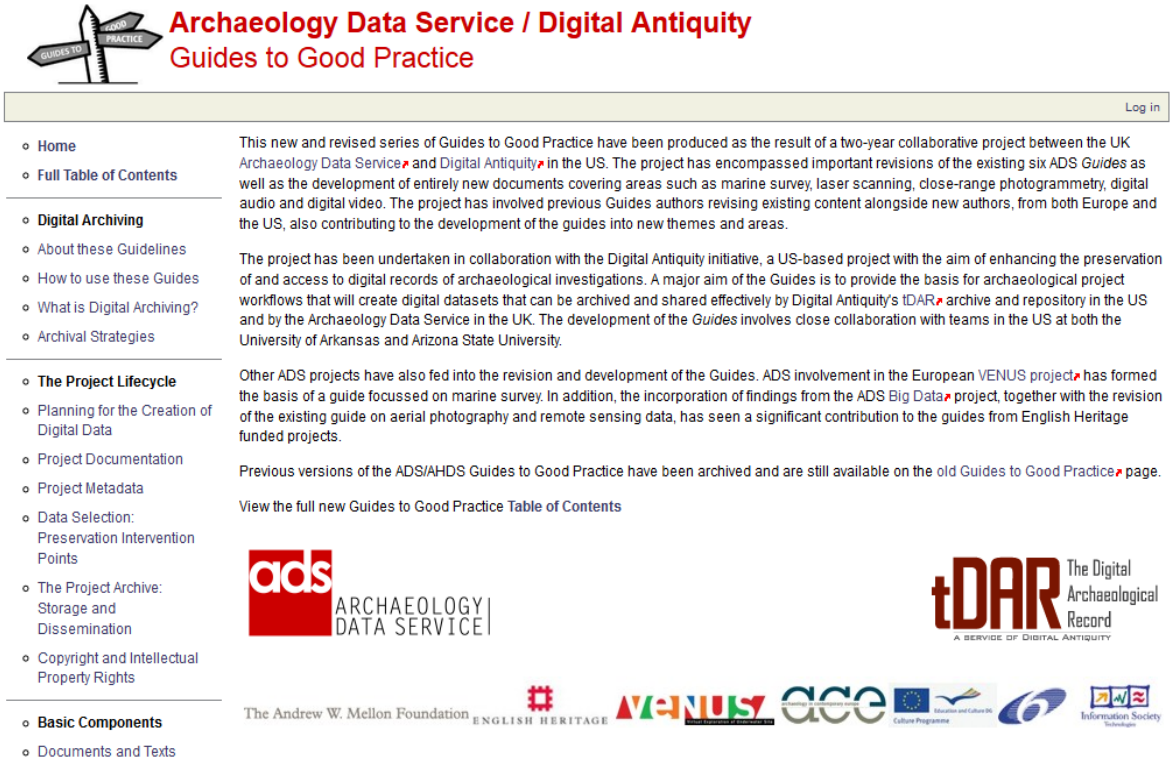
- What data will I produce?
- How will I organise the data?
  - **File structure**
  - **File naming and versioning**
  - **What file formats will I use?**
  - **Which software will I use?**
  - **Roughly how many files?**
  - **How will I describe and document my data? – METADATA**

ARCHAEOLOGY DATA SERVICE

METADATA - Guides to Good Practice

- Digital Data (general)
- GIS
- CAD
- Geophysics
- AP & Remote Sensing
- Excavation & Fieldwork
- Virtual Reality
- UAV





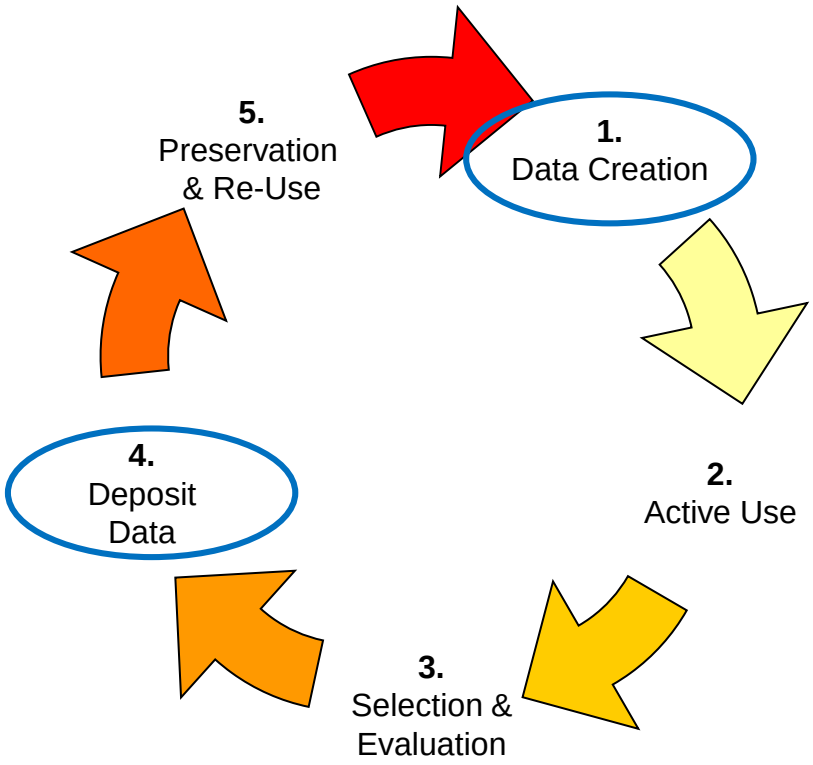
Silbury Hill, archive metadata

Location	World region	British Isles and Ireland
	British Isles country	England
	County	Wiltshire
	District	Kennet
	Parish	Avebury
	Place	Silbury Hill
	TGN	World, Europe, United Kingdom, England, Wiltshire, Silbury Hill (prehistoric site) [7032611]
Grid reference	OSGB	410011 168532
Grid reference	Latitude longitude bounding box	51.4174300461 -1.8605723152 -1.8541350603 51.4138434794
Subject	Archaeological Sciences (England)	SPECIALIST SAMPLING
	Archaeological Sciences (England)	ANTLER
	Event Type (England)	BOREHOLE SURVEY
	Event Type (England)	CORE SAMPLING
	Event Type (England)	ARCHAEOLOGICAL INTERVENTION

Silbury Hill, image metadata

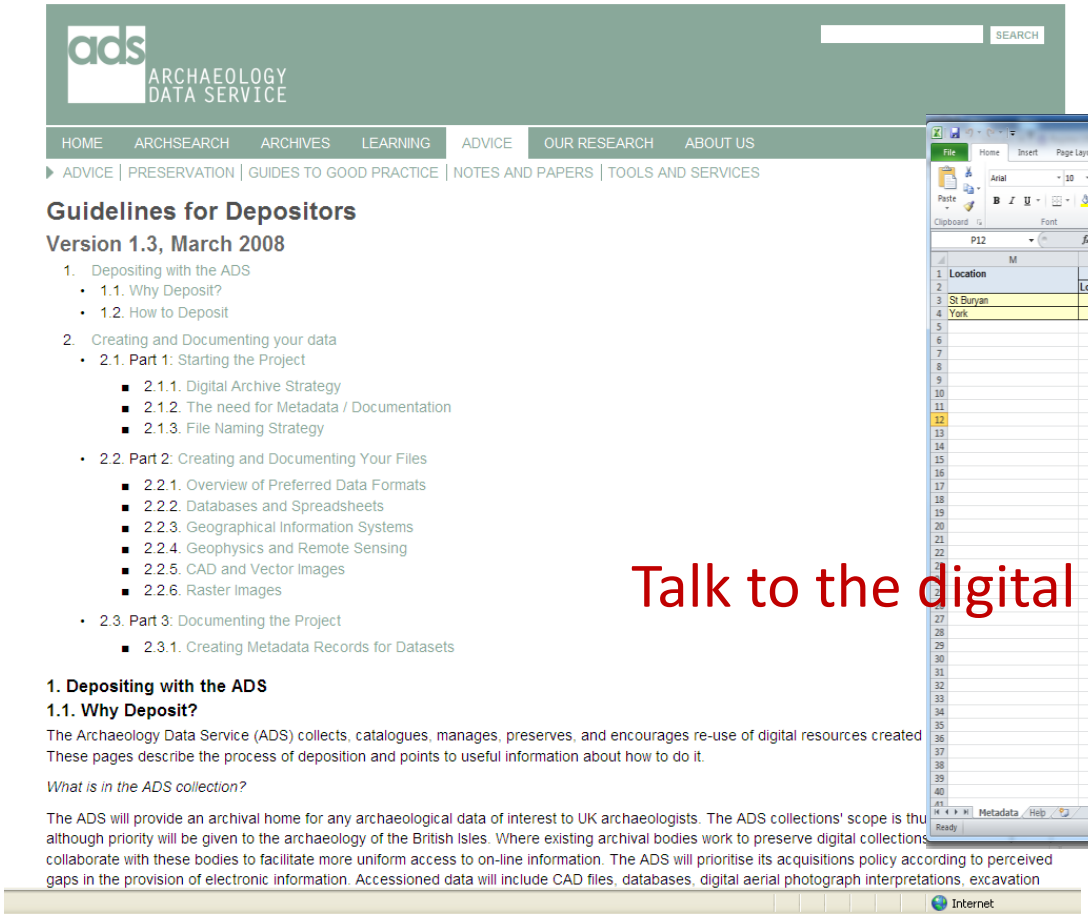
Name of table	**BulkFindsList**	
Purpose of table	**Records the bulk finds retrieved from site and from environmental samples.**	
Number of rows	**237**	
Primary Key	**UniqueBulkID**	
Foreign Key	**ContextNo (links back to the ContextIndex table); Material (links to GlossBulkFindMaterial look-up table); EstQuantity (links to Gloss_Finds_EstimatedQuantities look-up table)**	
Name of field	**Full description of field and codes or terminology used**	**Data type and field length**
(PK) UniqueBulkID	A unique identifier for each instance of bulk finds entry.	AutoNumber-Long Integer
(FK) ContextNo	Unique identifier for the Context the finds are from. Field 01	Number-Long Integer
(FK) Material	Material. (What the artefact is made from:- controlled vocabulary, text to be retrieved from look-up table: GlossBulkFindMaterial) Field 75	Text-50
Weight	Weight in grams. No field number on form.	Number-Long Integer
Quantity	Number of artifacts/fragments associated with the record number. No field number on form.	Number-Long Integer
(FK) EstQuantity	None-numeric field for estimate quantities. No field number on form. (Controlled symbols used to display the estimated quantities, can be retrieved and defined from look-up table: Gloss_Finds_EstimatedQuantities).	Text-10
Box nos	Records which box the artefact was stored in. No field number on form.	Text-50

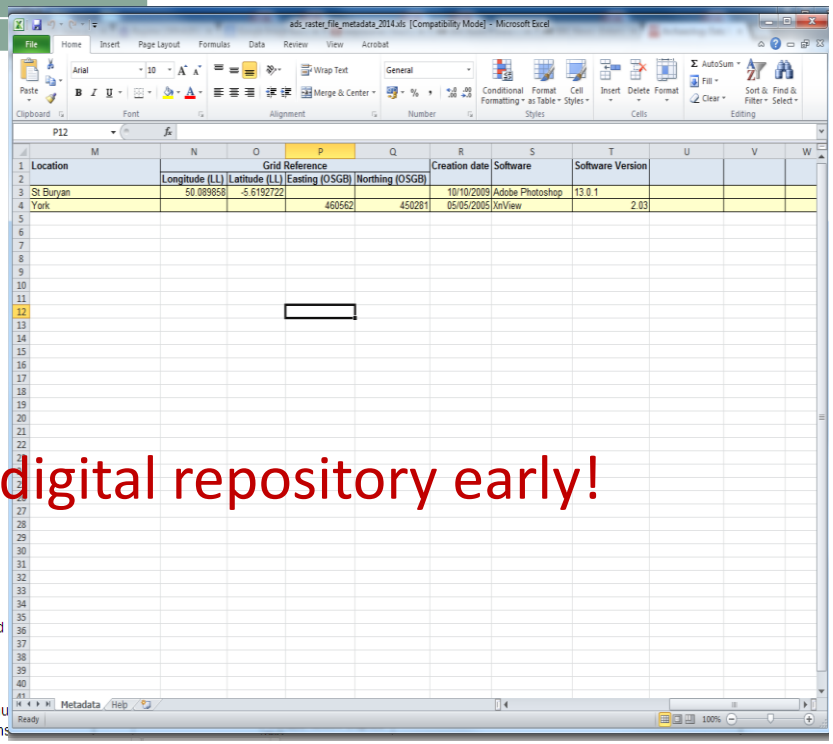
Data Cycles & Management Plans



- What data will I produce?
- How will I organise the data?
  - **File structure**
  - **File naming and versioning**
  - **What file formats will I use?**
  - **Which software will I use?**
  - **Roughly how many files?**
  - **How will I describe and document my data? – METADATA**
  - **Do I have to follow any requirements?**

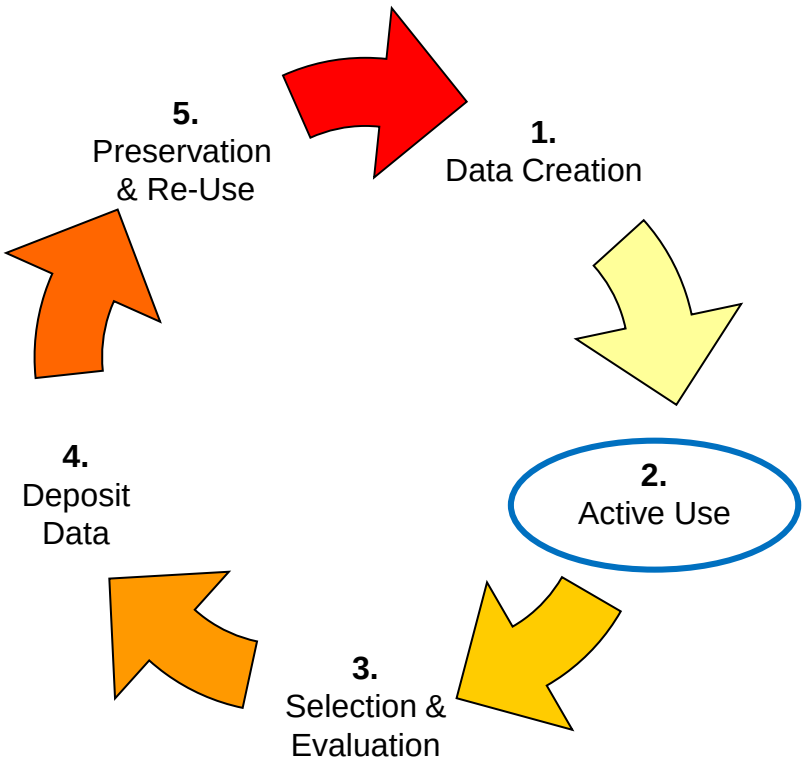
Data Cycles & Management Plans





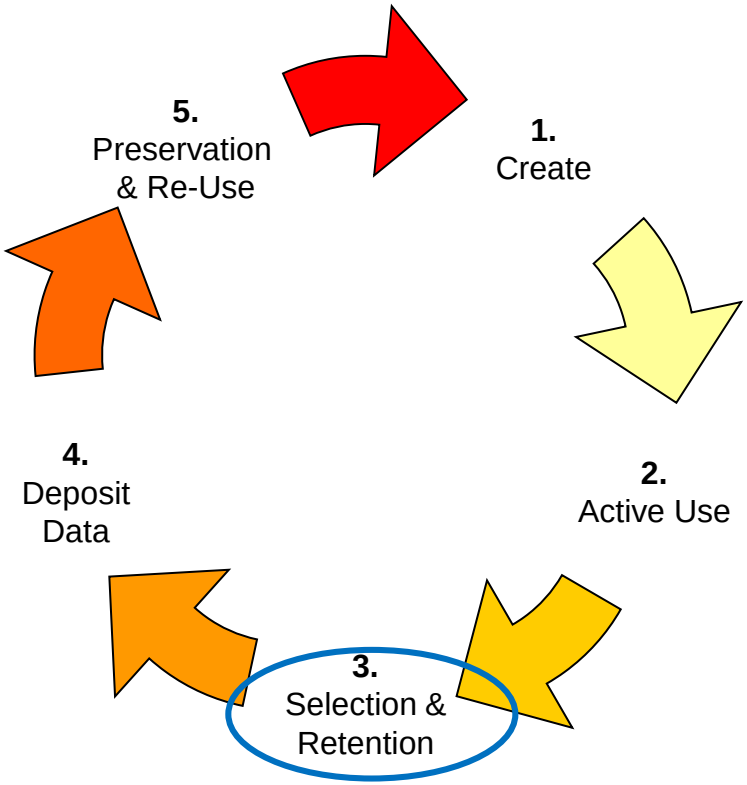
Talk to the digital repository early!

Data Cycles & Management Plans



- What standards and quality assurance might I use?
- How will I share data?
- How will I backup data?
- When will I evaluate if my data management is working?
  - Is the file structure / naming understandable to others?
  - Are further data required?
  - Are new data types required**?**

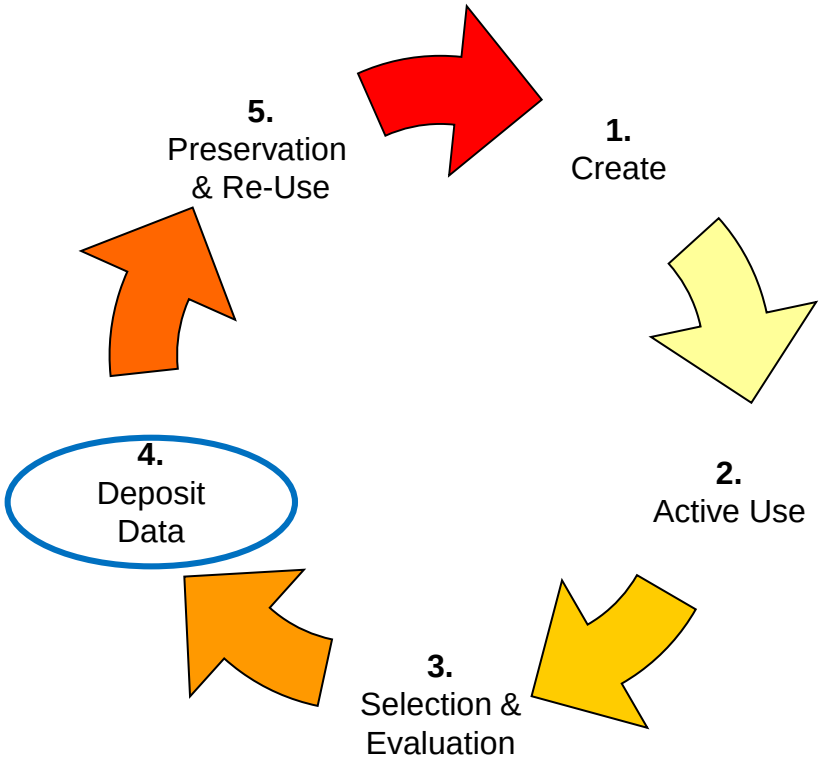
Data Cycles & Management Plans



- What data will I keep?

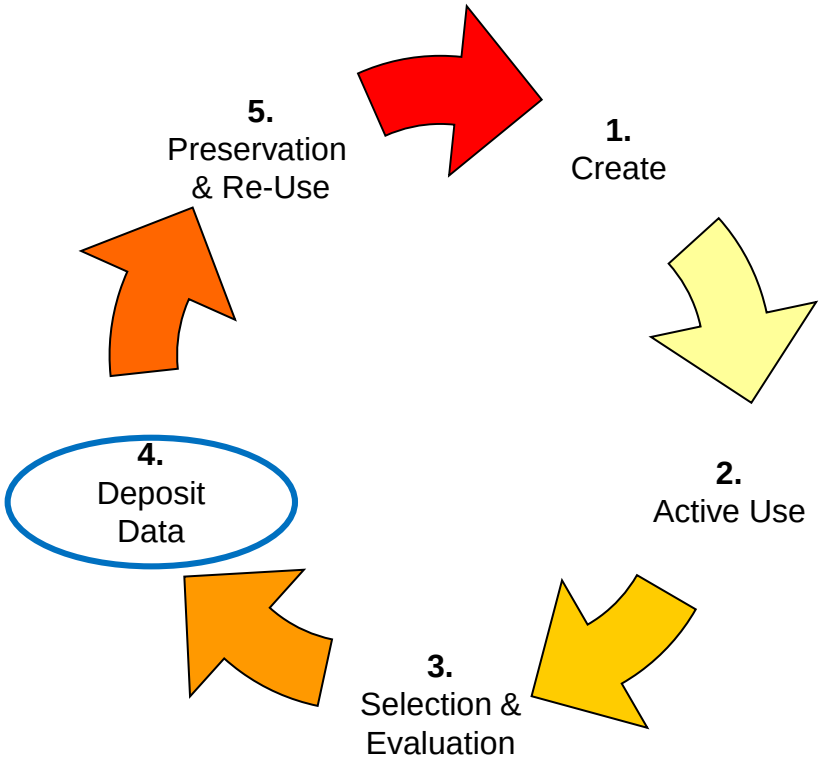
  Selection and Retention strategy

Data Cycles & Management Plans



- What data will be deposited and where?

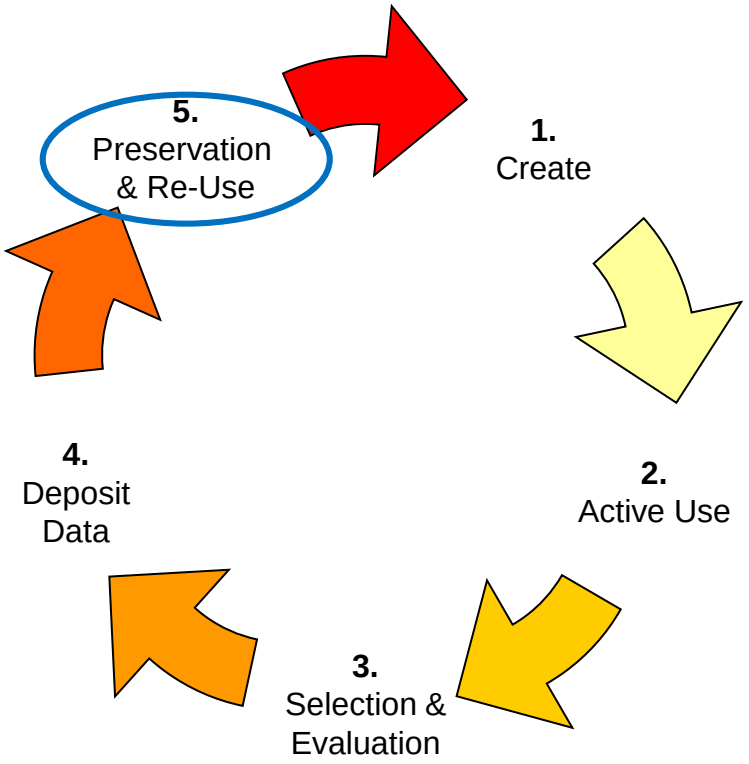
Data Cycles & Management Plans



- What data will be deposited and where?
  - Define the core data set of the project
  - Which data are supplementary?
  - Licences
  - Metadata
  - Where? Trusted Repository!

Talk to the digital repository early!

Data Cycles & Management Plans



- Who will be interested in re-using the data?
  - Who will be interested in re-using the data?
  - Is there sufficient information to allow easy re-use of the data?

Data Cycles & Management Plans

"The single most useful thing you can do to ensure the long-term preservation of your data is to plan for it to be re-used. Imagining it being reused by someone else who has never met you and who never will meet you, will cause you to approach the creation and design of your data in a new light.

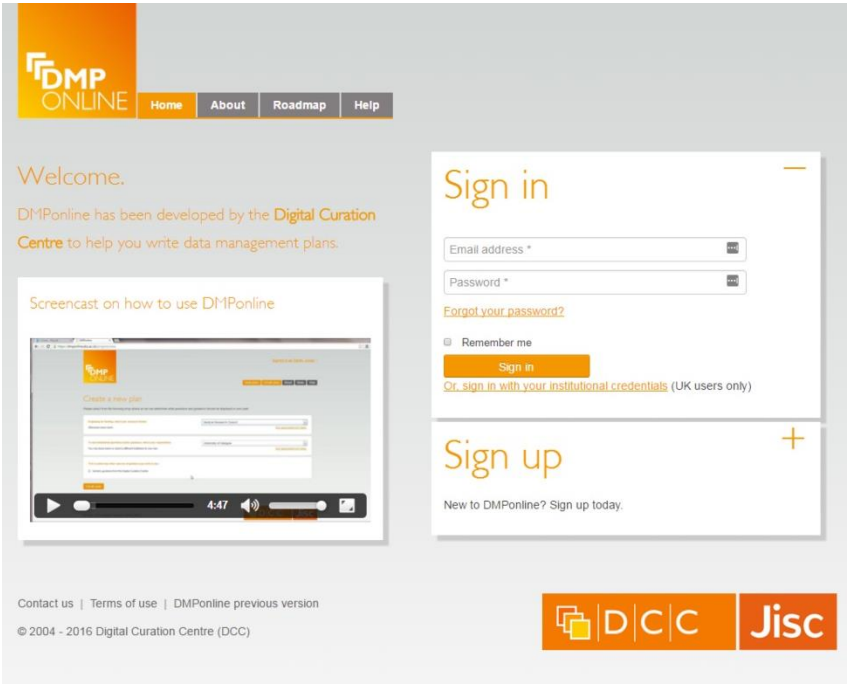
Moreover, studies show that re-use of data is the single surest way of maintaining the integrity of data and tracking errors and problems with it. In short, always plan for re-use"

Prof. Julian Richards, Director ADS.

Creating a DMP: Some useful sources

General guidance on data management and the creation of plans :-
http://www.dcc.ac.uk/resources/data-management-plans

DMP ONLINE
https://dmponline.dcc.ac.uk/



Why bother?

- Provides a **practical starting point** to help structure thoughts on your research/project
- Improves efficiency
- Help others understand the research process and how it developed
- Helps plan for data **reuse** by others, so the full potential of a research can be realised. Its lifecycle doesn't end here!
- Shows we take research integrity seriously and therefore increases trust in the archaeological community
- It is good practice
- Funding bodies require it!

What could go wrong?

- Cats-in-a-tent photos….and more

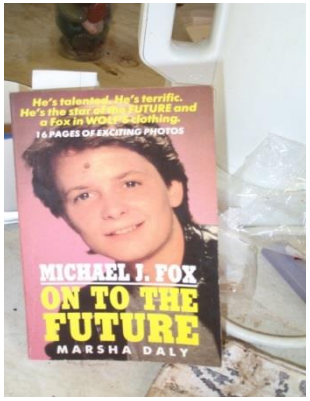
- Oral history project with no consent forms and therefore no audio files!

Downloads

We regret that we are unable to offer the archived audio files for this project at present due to copyright restrictions.

- Project in the red sea with videos of the rock of Gibraltar
- **Recently** an archive with the wrong archaeological site name in all the metadata

Odd one out – which image was included in a deposit but is not on the ADS web site?



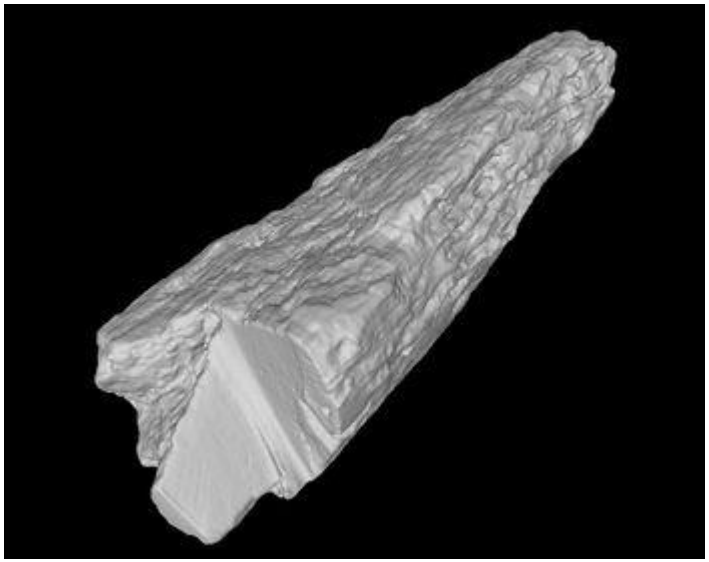
For an example of good data management:

Denisova 11 Human Bone Fragment ForSEAdiscovery

For an example of poor data management:

https://archaeologydataservice.ac.uk/learning/uniworkshop2020.xhtml

Denisova 11 Human Bone Fragment



```
msdata:
  id: 20150618_DC5
  version: 1.1.0
  cvList:
    cv:
      id: MS
      fullName: Proteomics Standards Initiative Mass Spectrometry
      version: 4.1.1
      URI: https://raw.githubusercontent.com/HUPO-PSI/psi-ms-CV/m
    cv:
      id: UO
      fullName: Unit Ontology
      version: 09:04:2014
      URI: https://raw.githubusercontent.com/bio-ontology-researc
  fileDescription:
    fileContent:
      cvParam: MS1 spectrum
      cvParam: profile spectrum
    sourceFileList:
      sourceFile:
        id: fid
        name: fid
        location: e:/20150618_DC5/0_I22/1/1SRef
        cvParam: SHA-1, b305b66c1045710c0ea9fda0b6629cfb893175f7
        cvParam: Bruker FID format
        cvParam: scan number only nativeID format
      sourceFile:
        id: I22_20150618_DC5_1.mzXML
```

ForSEAdiscovery

ForSEAdiscovery

RIB

This model is intended to show the overall site of the Ribadeo Shipwreck, located in Galicia, Northern Spain. The site was investigated by ForSEAdisovery during a fieldwork campaign in June 2015. (Brandon Mason Maritime Archaeology Trust)

Metadata

RIB01 EXIF	TXT	408 Kb
RIB01 Photogrammetry Report	XLSX	54 Kb

OBJ files

RIB obj	ZIP	39 Mb

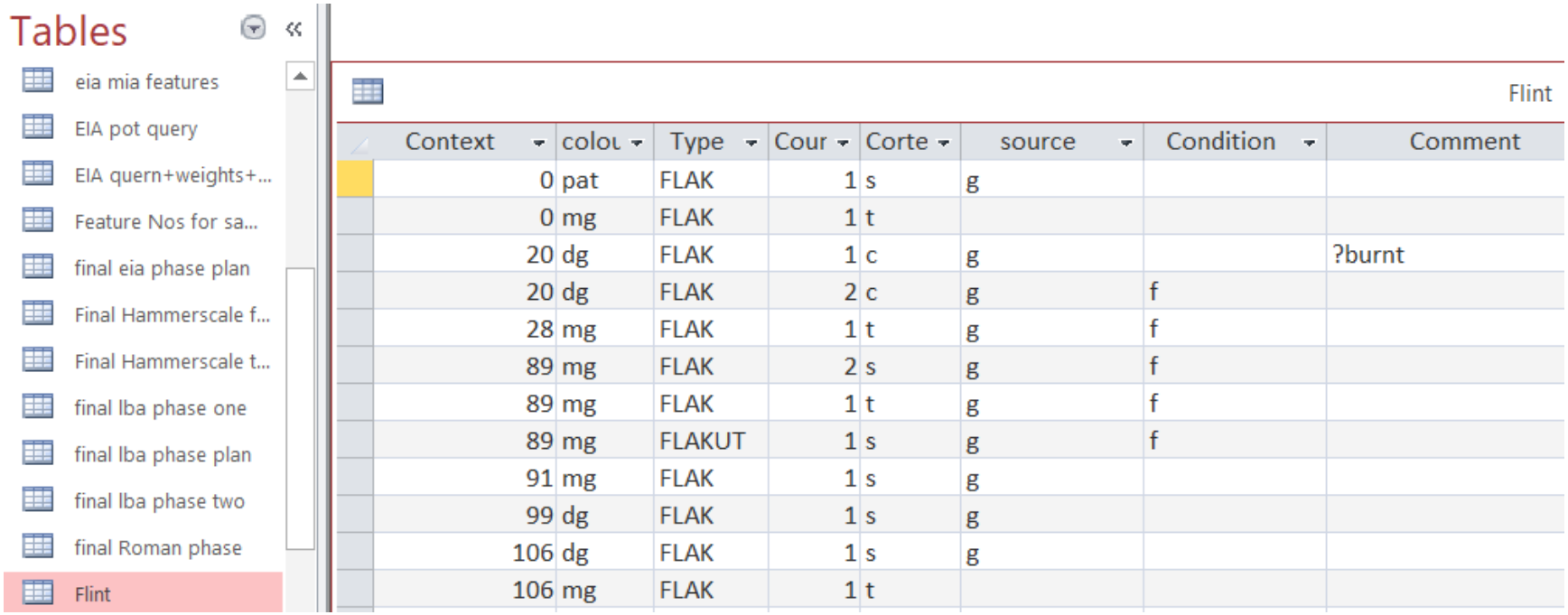
Source Images

A-RIB01-28-06-2015-1-BM-P-0001-0018	ZIP	97 Mb
A-RIB01-28-06-2015-1-BM-P-0019-0037	ZIP	98 Mb
A-RIB01-28-06-2015-1-BM-P-0038-0055	ZIP	94 Mb

For an example of poor data management:

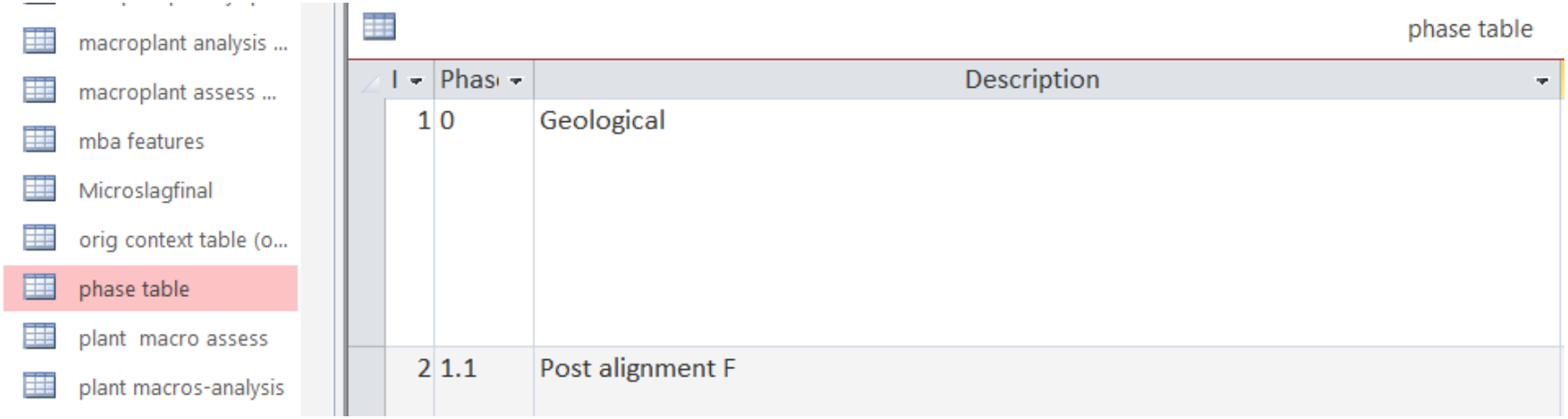
https://archaeologydataservice.ac.uk/learning/uniworkshop2020.xhtml

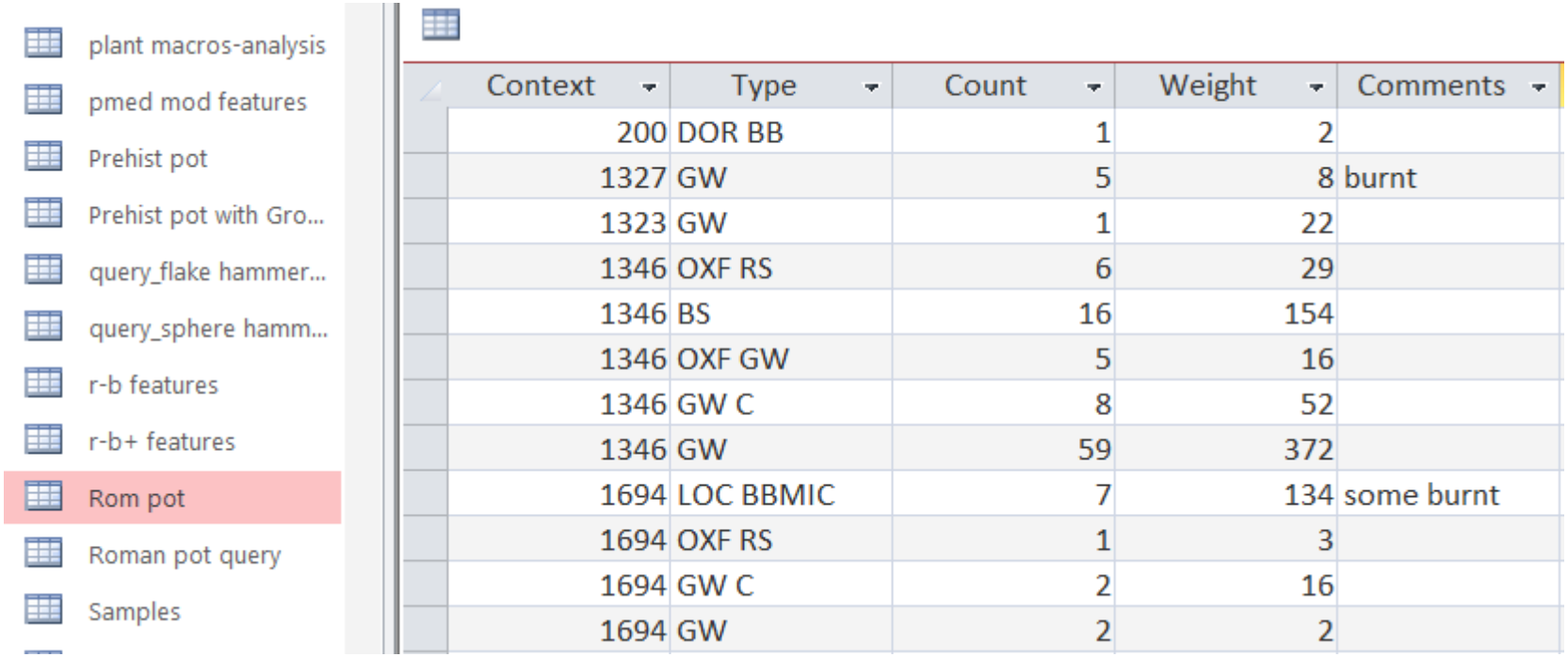
Figs.pdf	08/04/2016 10:09	Adobe Acrobat D...	11,031 KB
Questions.doc	08/04/2016 10:09	Microsoft Word 97...	23 KB
Report.doc	08/04/2016 10:09	Microsoft Word 97...	1,618 KB
TP database.mdb	27/01/2021 09:34	Microsoft Access ...	14,404 KB
TP_2590.JPG	08/04/2016 10:10	JPG File	487 KB
TP_2593.JPG	08/04/2016 10:10	JPG File	698 KB
TP_2594.JPG	08/04/2016 10:10	JPG File	684 KB
TP_2684.JPG	08/04/2016 10:10	JPG File	447 KB
TP_2689.JPG	08/04/2016 10:11	JPG File	531 KB
TP_2707.JPG	08/04/2016 10:11	JPG File	660 KB



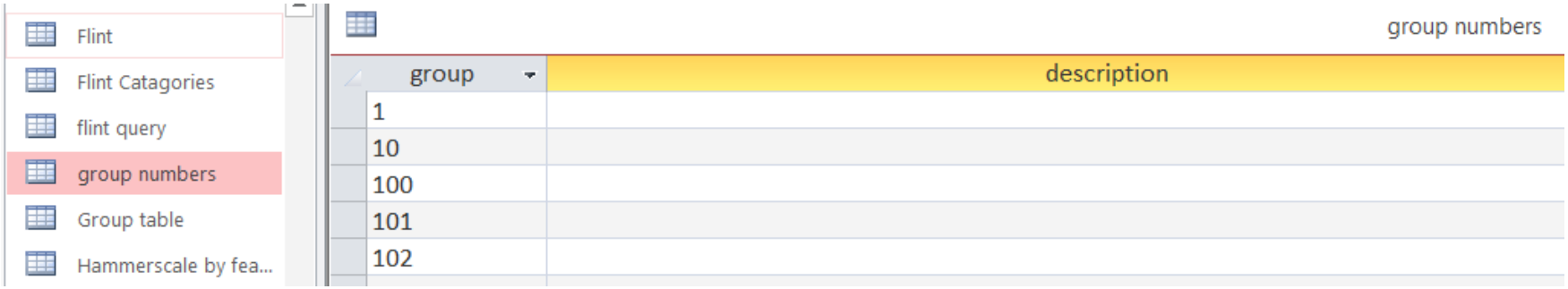
Flint

Context	colou	Type	Cour	Corte	source	Condition	Comment
0	pat	FLAK	1	s	g		
0	mg	FLAK	1	t			
20	dg	FLAK	1	c	g		?burnt
20	dg	FLAK	2	c	g	f	
28	mg	FLAK	1	t	g	f	
89	mg	FLAK	2	s	g	f	
89	mg	FLAK	1	t	g	f	
89	mg	FLAKUT	1	s	g	f	
91	mg	FLAK	1	s	g		
99	dg	FLAK	1	s	g		
106	dg	FLAK	1	s	g		
106	mg	FLAK	1	t			


macroplant analysis ...
macroplant assess ...
mba features
Microslagfinal
orig context table (o...
phase table
plant macro assess
plant macros-analysis
phase table
I
Phas
Description
1 0 Geological
2 1.1 Post alignment F



Context	Type	Count	Weight	Comments
200	DOR BB	1	2	
1327	GW	5	8	burnt
1323	GW	1	22	
1346	OXF RS	6	29	
1346	BS	16	154	
1346	OXF GW	5	16	
1346	GW C	8	52	
1346	GW	59	372	
1694	LOC BBMIC	7	134	some burnt
1694	OXF RS	1	3	
1694	GW C	2	16	
1694	GW	2	2	


Flint
Flint Catagories
flint query
group numbers
Group table
Hammerscale by fea...
group numbers
group
description
1
10
100
101
102

There are no references within the document nor any documentation to tell us the significance of these features or even which way is North.

We only know it's titled TP_2707

Answers

1. What is shown in photograph DCP _2707.jpg? **No idea**
2. On what plan is Context 890 drawn? **80/520**
3. What contexts appear on Section drawing 249? **1173,1174,1175,1176**
4. In which phases /stratigraphic groups are post-holes 1231, 1235 and 1243? **Group 34 Phase 1.6 LBA**
5. List the finds recovered and recorded from Context 931. **Worked stone 1/0g, Worked stone 1/0, Stone tile 1/0, Roman pot 28/303, Daub 81109, Fe object 2/377**
6. In what type of feature was Small Find 6 found? **Pit 1141 context 1105**
7. What is Small Find 6 and what is it made from? **Loomweight, Fired clay**
8. How many fragments of worked flint came from contexts 89 and 1704? **4 and 2.**
9. Describe the cortex condition of the above fragments. **No idea, 89 = f.**
10. How many pottery jar sherds are there in Context 89? **15**
11. What is the meaning of pottery form 'P'? **No idea**
12. What is the meaning of Roman pottery type BS? **No idea**
13. Explain Table J Overall Forms by Period B1 and B2. **Haven't a clue!**

Teagan Zoldoske

teagan.zoldoske@york.ac.uk

Archaeology Data Service

Department of Archaeology

The King's Manor

University of York

York

YO1 7EP

Follow us on Instagram:
@archaeologydataservice

Follow us on Twitter:
@ADS_Update

Like us on Facebook:
http://www.facebook.com/archaeology.data.service