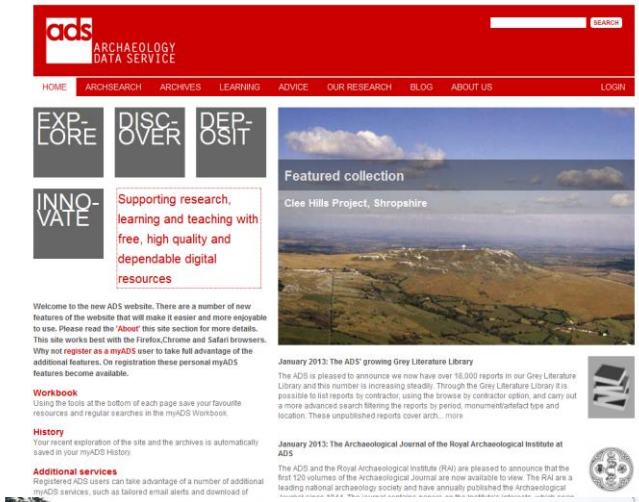
Making sense of a collection:
what happens when things don't go to plan?

ARCHAEOLOGY DATA SERVICE

Getting Started in Digital Preservation
The University of Leeds
31 March 2015

Ray Moore
Digital Archivist

Archaeology Data Service: Who we are and what we do



- Founded 1996
- originally part of Arts and Humanities Data Service (AHDS)
- based within the University of York
- a not for profit organisation
- Collections
  - 1,100,000 metadata records - ArchSearch
  - 30,000+ unpublished fieldwork reports
  - 700+ rich archives
  - 19 journals and series
  - PhDs
- providing advice and guidance
- supporting research, learning and teaching

What's in an archaeological archive?

© Buch Edition

Each requiring specific metadata

- Documents and text
- Raster Images
- Spreadsheets
- Databases
- Video and Audio
- CAD and Vector Images
- Geographical Information Systems
- Geophysics
- Aerial survey
- Virtual reality
- Photogrammetry
- Laser scanning

…etc.

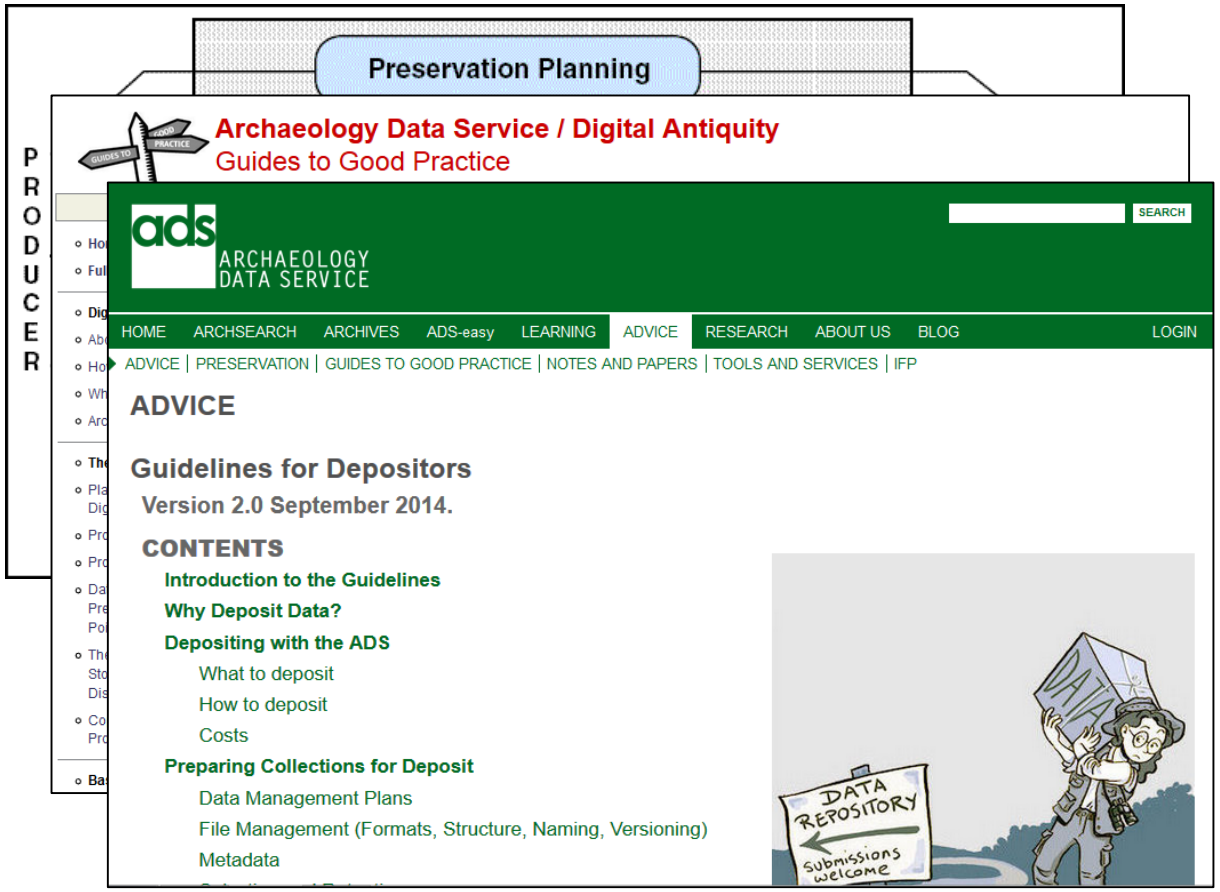
Increasing complexity

Planning for Preservation

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 of ADS

Standards and Guidelines

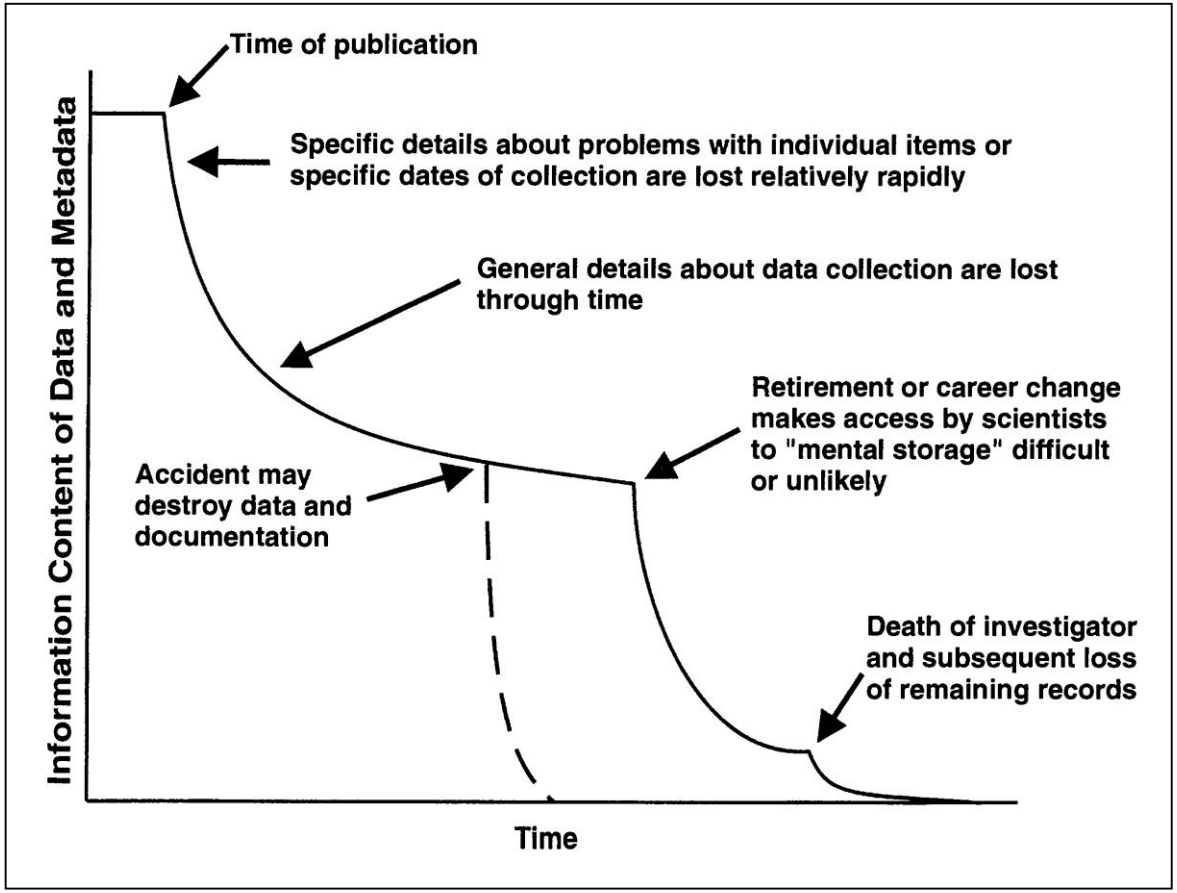


- We follow the **Open Archival Information System** (OAIS) reference model
- We co-publish the **Guides to Good Practice**, which contribute to professional standards within archaeology
- We produce our own **Guidelines for Depositors** which provides help with collection and file-level metadata and formats

The best laid plans

…but… what happens when things don't go to plan

A degrading resource



Michener, W.K., Brunt, J.W., *et al.* 1997. 'Nongeospatial Metadata for the Ecological Sciences'. *Ecological Applications* **7**: 330-342.

Why is long-term preservation difficult?

"Digital information lasts forever - or five years, whichever comes first."
(Jeff Rothenberg, RAND Corp., 1997)

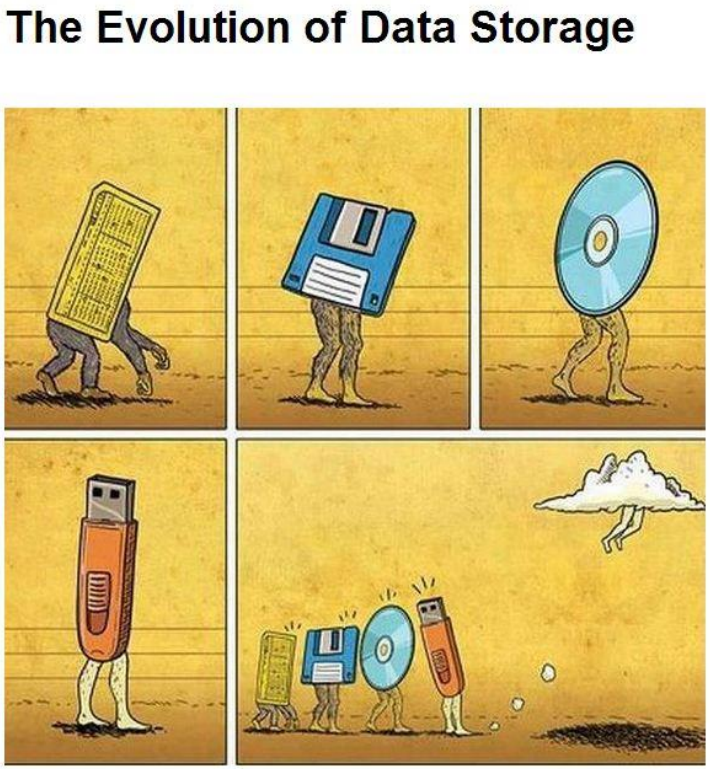
Preservation influenced by tangible…

- Deterioration of the storage medium (CD)
- Obsolescence of the storage medium (5¼ " disk)
- Obsolescence of the hardware (to run software)

… and intangible problem's

- Obsolescence of the software (Quattro Pro)
- Obsolescence of formats (WordStar/ws)
- Lack of documentation/metadata

All highlight a need for long-term management



©Zoolz, 2013

Making sense of a collection

Case Study: Newham Museum Archaeology Project Archives

- Newham Museum Archaeological Service closed down in 1998
- digital archive passed to ADS by the London Borough of Redbridge
- over ten years work, encompassing about 180 sites, and 200 separate projects
- as delivered archive consists of about 230 5.25" and 3.5” floppy disks, totalling over 130Mb of data, and some 6432 individual files
  - 1500 excavation reports
  - 700 database files
  - 1200 geophysics files
- selection and retention showed “range and quality of the material varies widely”

Making sense of a collection

Case Study: Newham Museum Archaeology Project Archives

Problems:

- physical media not the main problem – **ADS computer museum**
    - data loss corruption (25 files) unrecoverable (12 files)
- software obsolescence e.g Word Star/TurboCAD
    - **ADS computer/software museum** = migration
- little/no documentation, poor file structure and inconsistent file naming
  = almost impossible to deduce the precise nature of each file e.g. database
- issues of copyright – material published elsewhere
- included ephemera, administrative matters and correspondence
- mix of working and complete documents
- linking to digital data not included in archive

"The loss of data from the Newham Archive is as much to do with poor project planning as it is do to with preservation. The archaeologists were not prepared for a sudden closure; there was no exit-strategy providing details of what to do with the data in such a situation" (Dunning 2001. *Excavating Data: Retrieving the Newham Archive*).

Despite these problems, we were able to save a significant amount of the data included in the archive, BUT things could have been easier and less costly if there had been a Data Management Plan

Making sense of a collection

Case Study: Alan Vince Archive

"Archaeologist who transformed the study of Anglo-Saxon and medieval ceramics."

Lohman, J 2009. 'Alan Vince' *The Guardian*, 29 April 2009.

His work was "a cornerstone of medieval ceramic studies".

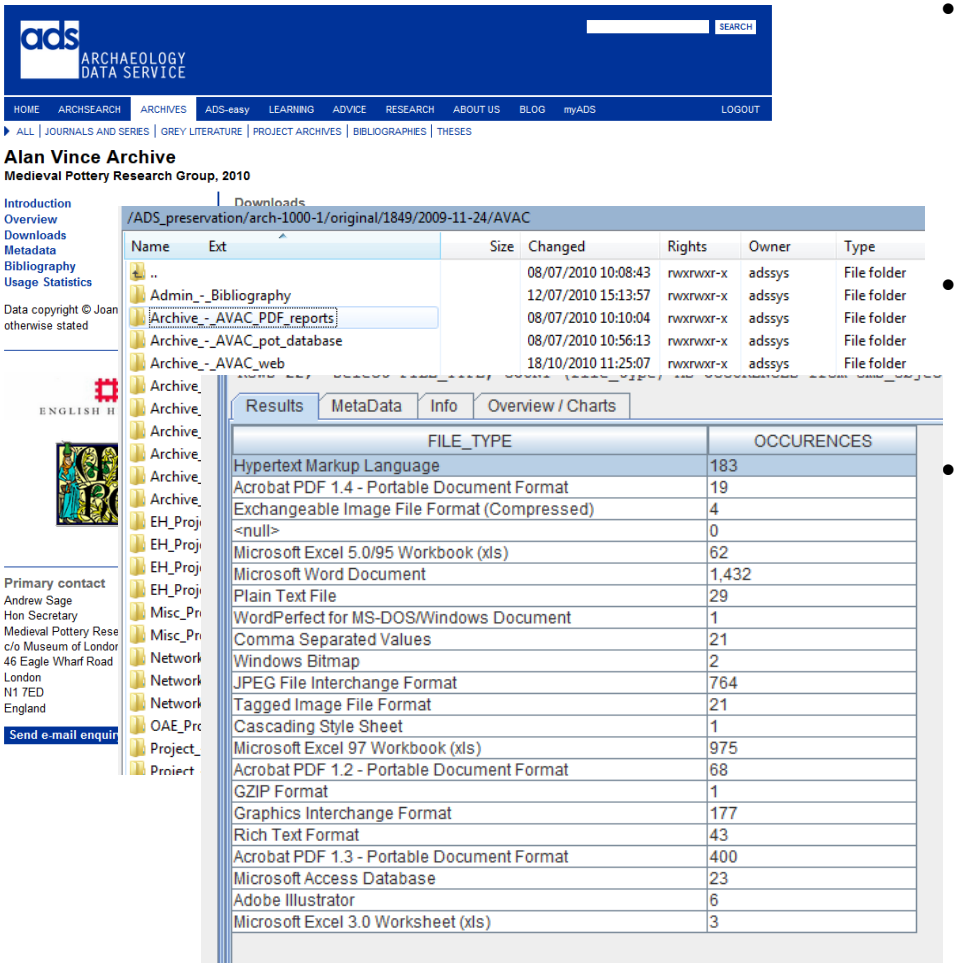
ADS website

- with the good grace of Joanna Vince (his widow)
- the Medieval Pottery Research Group, funded by English Heritage, undertook a comprehensive audit of his work
- physical archives returned, or deposited (English Heritage)
- thin section slides and associated paperwork (British Museum)
- his ceramic databases, reports and website, including the AVAC ceramic catalogue, Ceramic Chemical Composition Database and his unpublished grey literature reports (ADS)

30 Mar 1952 –
23 Feb 2009

Making sense of a collection

Case Study: Alan Vince Archive



FILE_TYPE	OCCURENCES
Hypertext Markup Language	183
Acrobat PDF 1.4 - Portable Document Format	19
Exchangeable Image File Format (Compressed)	4
<null>	0
Microsoft Excel 5.0/95 Workbook (xls)	62
Microsoft Word Document	1,432
Plain Text File	29
WordPerfect for MS-DOS/Windows Document	1
Comma Separated Values	21
Windows Bitmap	2
JPEG File Interchange Format	764
Tagged Image File Format	21
Cascading Style Sheet	1
Microsoft Excel 97 Workbook (xls)	975
Acrobat PDF 1.2 - Portable Document Format	68
GZIP Format	1
Graphics Interchange Format	177
Rich Text Format	43
Acrobat PDF 1.3 - Portable Document Format	400
Microsoft Access Database	23
Adobe Illustrator	6
Microsoft Excel 3.0 Worksheet (xls)	3

- Selection and retention
  - USB drive – **media stability**
  - unsupported formats - **few**
  - corrupt data streams - **little**
  - personal correspondence - **limited**
  - some already archived – Grey Literature Library
- Metadata
  - collection level metadata - **completed**
  - file-level – some documentation, and helped by the fact much of the archive was textual
- File management
  - logical file structure
  - contextual file naming
  - file formats

Problems:

- were limited really
- small-ish archive
- not a huge amount of data
- Alan technologically aware - close association with the ADS (first editor of Internet Archaeology) so aware of the issues

Making sense of a collection

Some lessons

- give individuals and organisations the tools and advice to plan for preservation at the outset

BUT don't expect that this will always be used, or followed

- create a preservation plan

BUT accept that things won't always go to plan

- selection and retention of data is always necessary

BUT be pragmatic about what is possible

- decide what you want to preserve

BUT accept that for whatever reason everything cannot be preserved

- documentation and metadata are essential

WITHOUT it data is much less valuable and preservation may not be worthwhile

Follow us on Twitter:
@ADS_Update

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

ads ARCHAEOLOGY DATA SERVICE

http://archaeologydataservice.ac.uk

ray.moore@york.ac.uk

Thank You