



ccs

## File obsolescence at the ADS?

Tim Evans 23-06-2016

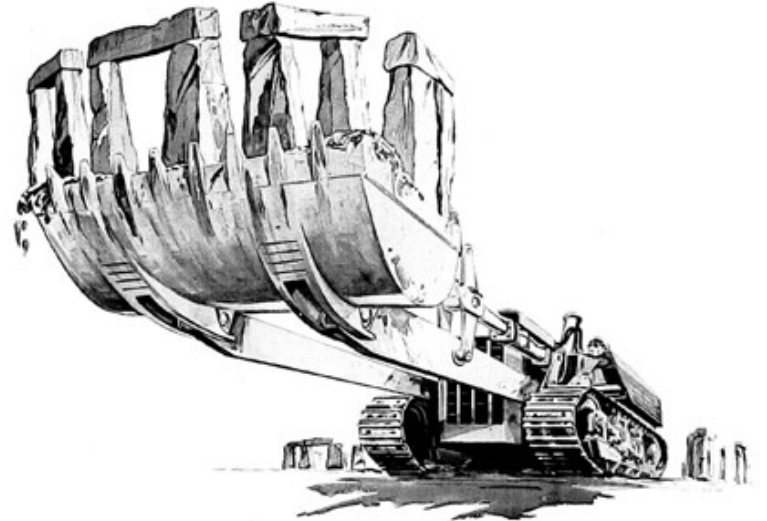
# Introduction

## **The Archaeology Data Service:**

- Established in 1996
- Based within the Department of Archaeology, University of York
- Digital archive for UK-based fieldwork and research in archaeology
  - Academic 'research' archives produced by Higher Education
  - The results of development-led fieldwork

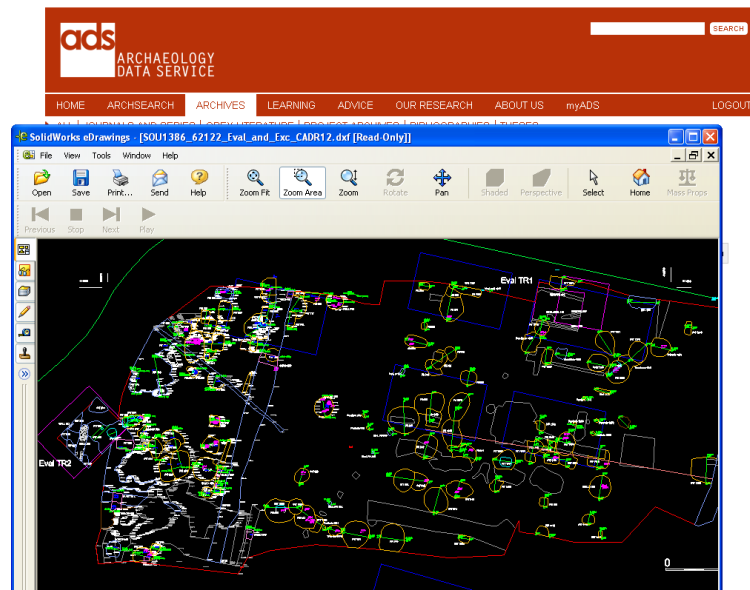
# What was the impetus in 1996?

- Is destructive!
- Resources are unique
- The preserved record becomes the main resource for future interpretation



# Progress to-date

- Collections
  - 1,100,000 metadata records
  - 700+ rich archives
    - Raster + vector images
    - Databases
    - Geospatial data



# The traditional problem

- Software upgrades fail to support legacy files.
- Format is superseded by another or evolves in complexity.
- The format "take up" is low or industry fails to create compatible software.
- The format fails, stagnates, or is no longer compatible with the current environment.
- Software supporting the format fails in the marketplace or is bought by a competitor and withdrawn

# Preservation via migration

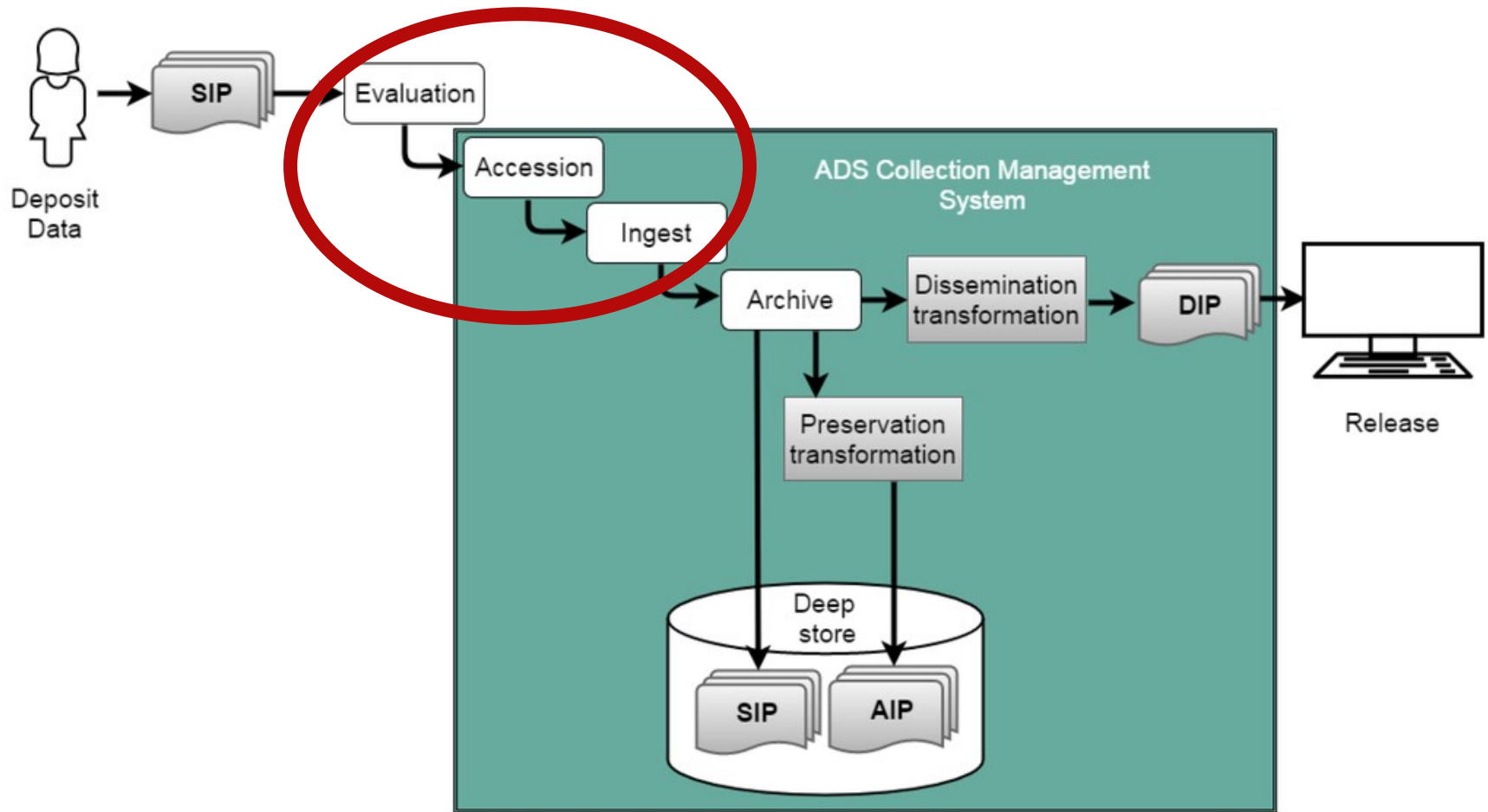
- Every file accessioned preserved as a standardised format e.g. all raster images as uncompressed TIFF, databases as delimited text...
- Formats used based on a mixture of technical consideration (e.g. compression = data loss), but also judgements on longevity of format and ease of establishing future migrations.

# Preservation Policy

<http://archaeologydataservice.ac.uk/advice/PreservationPolicyRev>

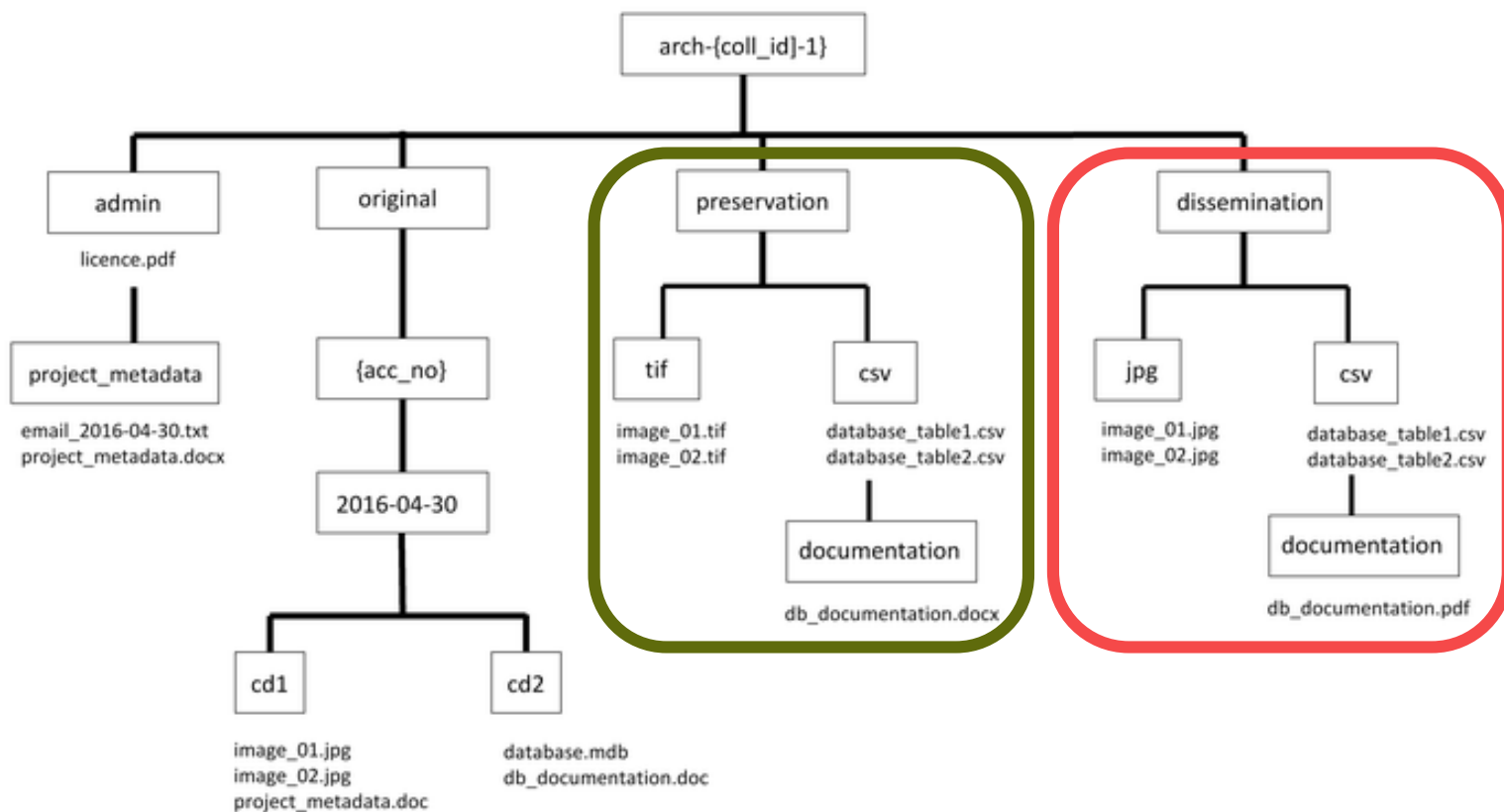
<http://archaeologydataservice.ac.uk/advice/RepositoryOperations>

# Operational workflow





# Example structure



# "Cannot Open a database created with a previous version of your application" error in Access 2013 and Access 2016

Email

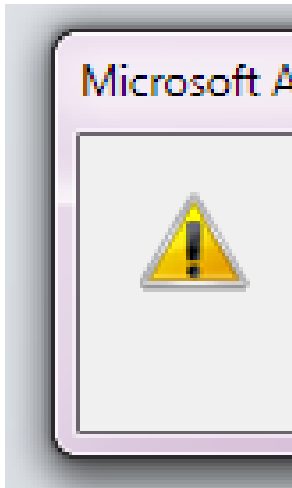
Print

This issue occurs when you try to use Access 2013 or later version of Microsoft Access to open an Access 97 database. To work around this issue, use a pre-Access 2013 version of Access to save the Access 97 database as an .accdb file:

1. Open the Access 97 database in Access 2003.
2. On the Tools menu, click Database Utilities, click Convert Database, and then click to Access 2002-2003 file format.
3. Enter a name for the database, and then click Save.
4. Exit Access 2003.
5. Open the database in Access 2013 or later version of Access.
6. On the File tab, click Save As, select Access Database (\*.accdb), and then click Save As.
7. In the Save As dialog box, click Save.

## Cause

This issue occurs because recent versions of Access cannot convert Access 97 files.



## In more detail...

- Query of the ADS Object Management System returned:
  - 2,118,052 files
  - 860,190 of which are in SIP
    - 253 unique formats
  - By contrast only 38 unique formats used in the AIP

# Most popular?

- 320,000+ .tif
- 180,000+ .jpg (1.01)
- 105,000+ .jpg (1.02)
- 20,000+ .jpg (raw)
- 17,000+ .pdf (1.4)
- 16,000+ .pdf (1.6)
- 13,000+ .doc (97-2003)

# Proteins in dinosaur bones?

19



*Molecular Signatures from the Past*

The Ancient Biomolecules Initiative is a five-year programme to understand the fate of biological molecules in archaeological and fossil materials, and to explore the applications of this new knowledge. The Initiative is funded by the Natural Environment Research Council.

Michael  
the re  
in am  
were  
surviv

DNA do  
bones, b  
building  
damage  
over mil

Dinosaur  
have be  
unusual  
suspect

Document Properties

Description Security Fonts Initial View Custom Advanced

Description

File: 19REV.PDF

Title: 19 (revised)

Author: Ewan Lauder

Subject:

Keywords:

Created: 03/02/1999 11:35:58

Modified:

Application: QuarkXPress® 4.04; LaserWriter 8 B1-8.6

Additional Metadata...

Advanced

PDF Producer: Acrobat Distiller 3.02 for Power Macintosh

PDF Version: 1.1 (Acrobat 2.x)

Location: E:\preservation\_policy\\_examples\

File Size: 443.16 KB (453,797 Bytes)

Page Size: 8.26 x 11.69 in

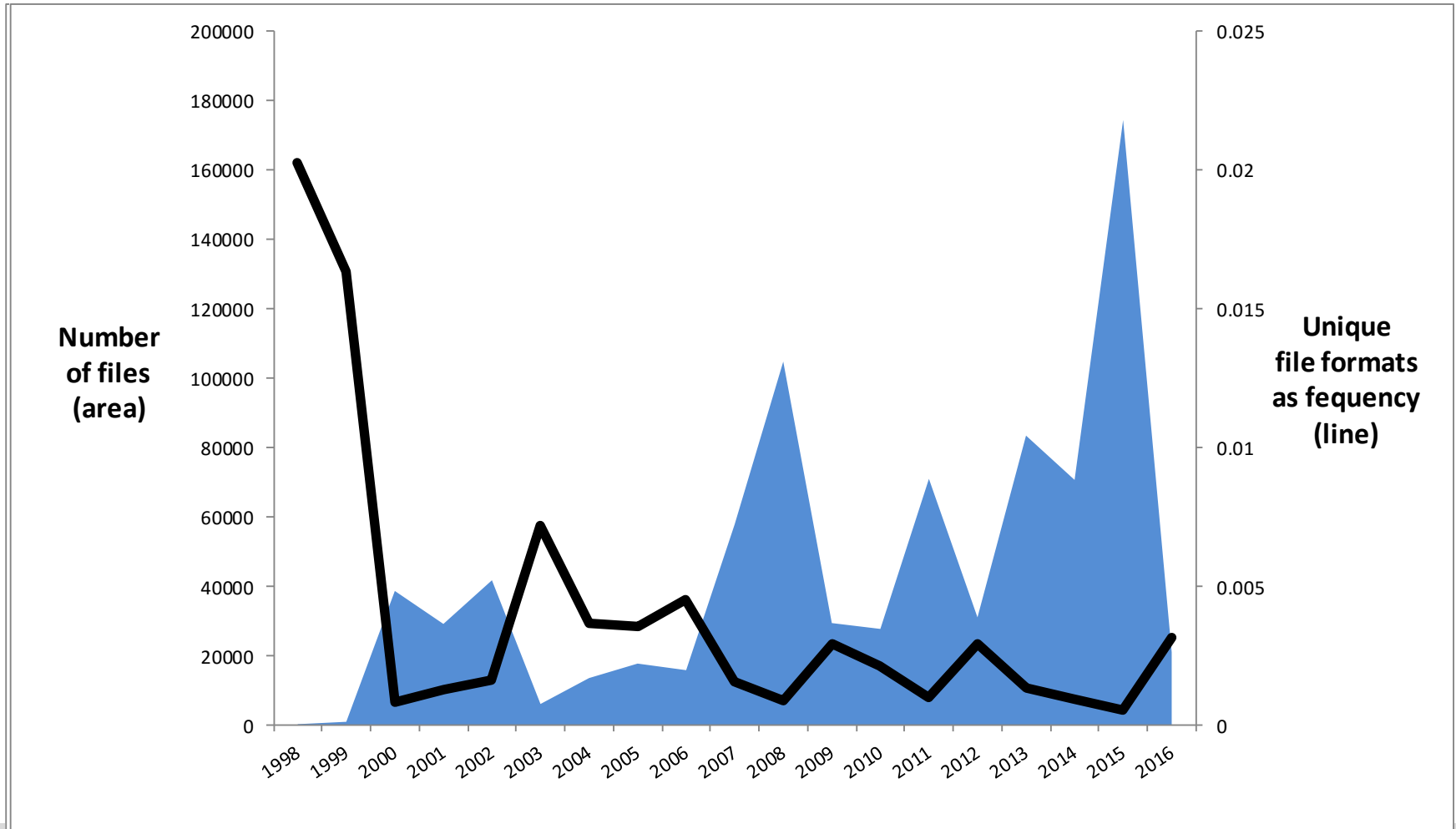
Number of Pages: 2

Tagged PDF: No

Fast Web View: No

Help OK Cancel

# More data – less variation



- Homogenisation of formats has meant that support has continued.

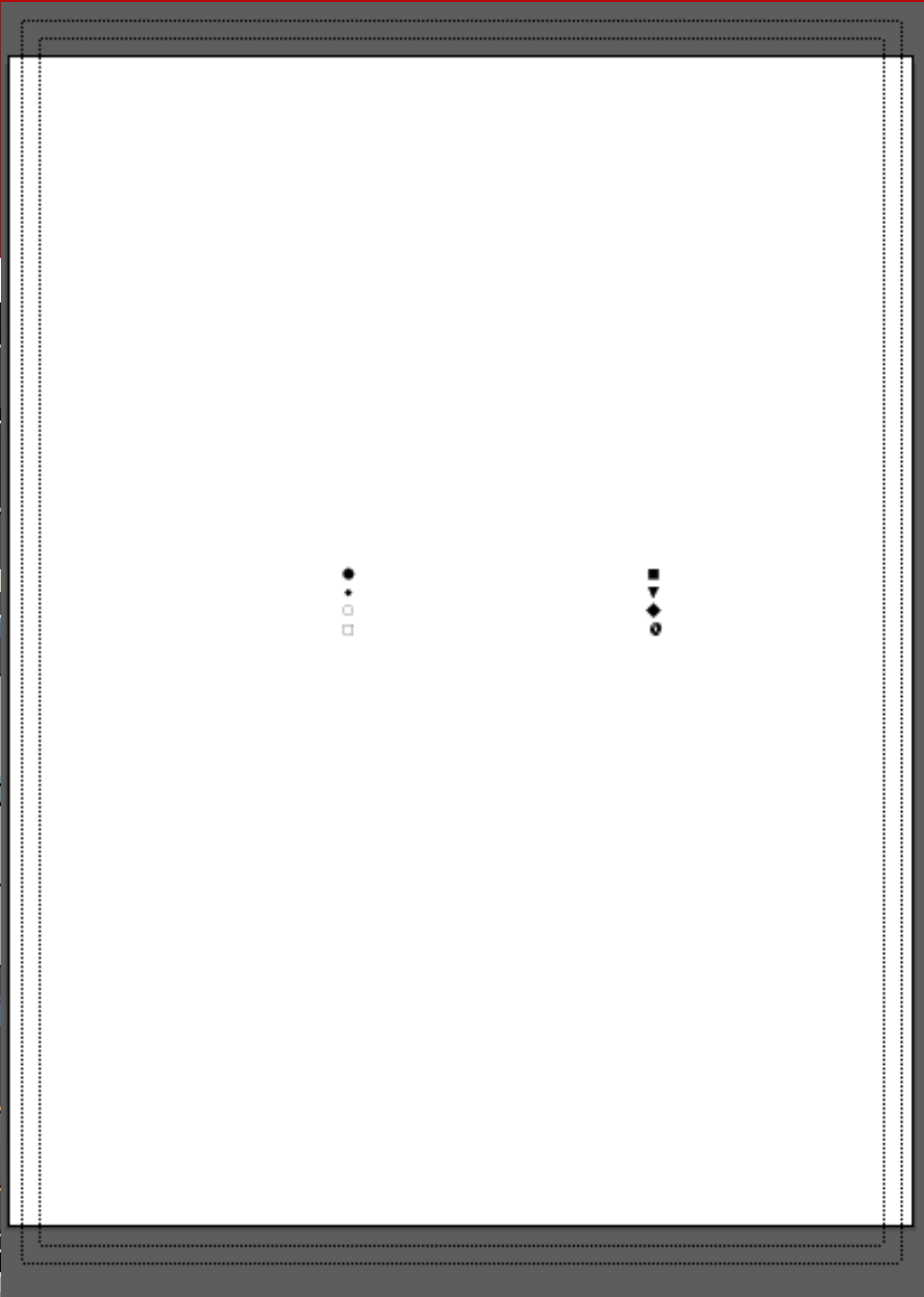
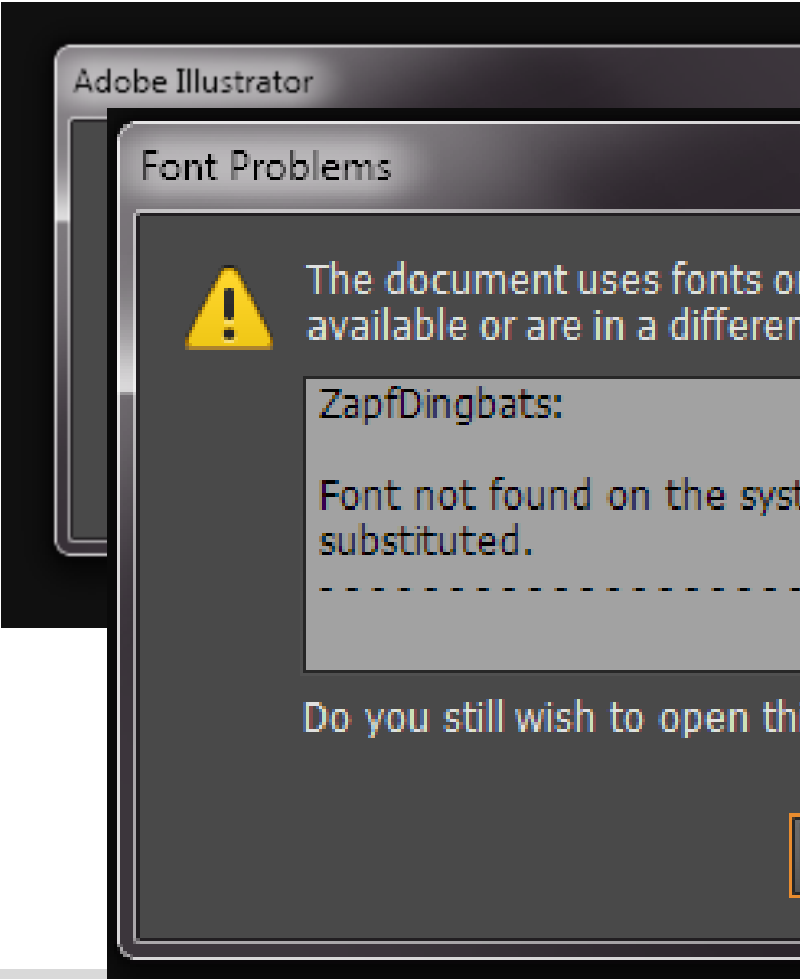
# Uncommon

- 609 Macintosh PICT Image
  - Originally created as part of Apple Macintosh
  - Superseded by more modern (Quartz) systems
  - Still successfully open with CorelDraw X6
- 27 Lotus 1-2-3 Worksheets
  - Produced by IBM until 2002 – now considered (by some) obsolete
  - Open successfully in OpenOffice



# Unusual/homebrews

- 594 GSSI RADAN data file .dzt
  - A proprietary format created by a US-based geophysical company
  - At the moment *can only* be exported using RADAN software
- 21 Harris Matrix .hm
  - A stand-alone product created by a (now ended) research project

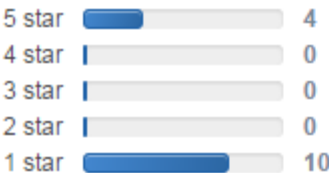




### User Reviews

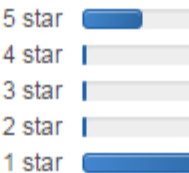
#### Current Version

★★★★★  
out of 14 votes



#### All Versions

★★★★★  
out of 14 votes



Sort: Date

★★★★★ "If you want to waste 10

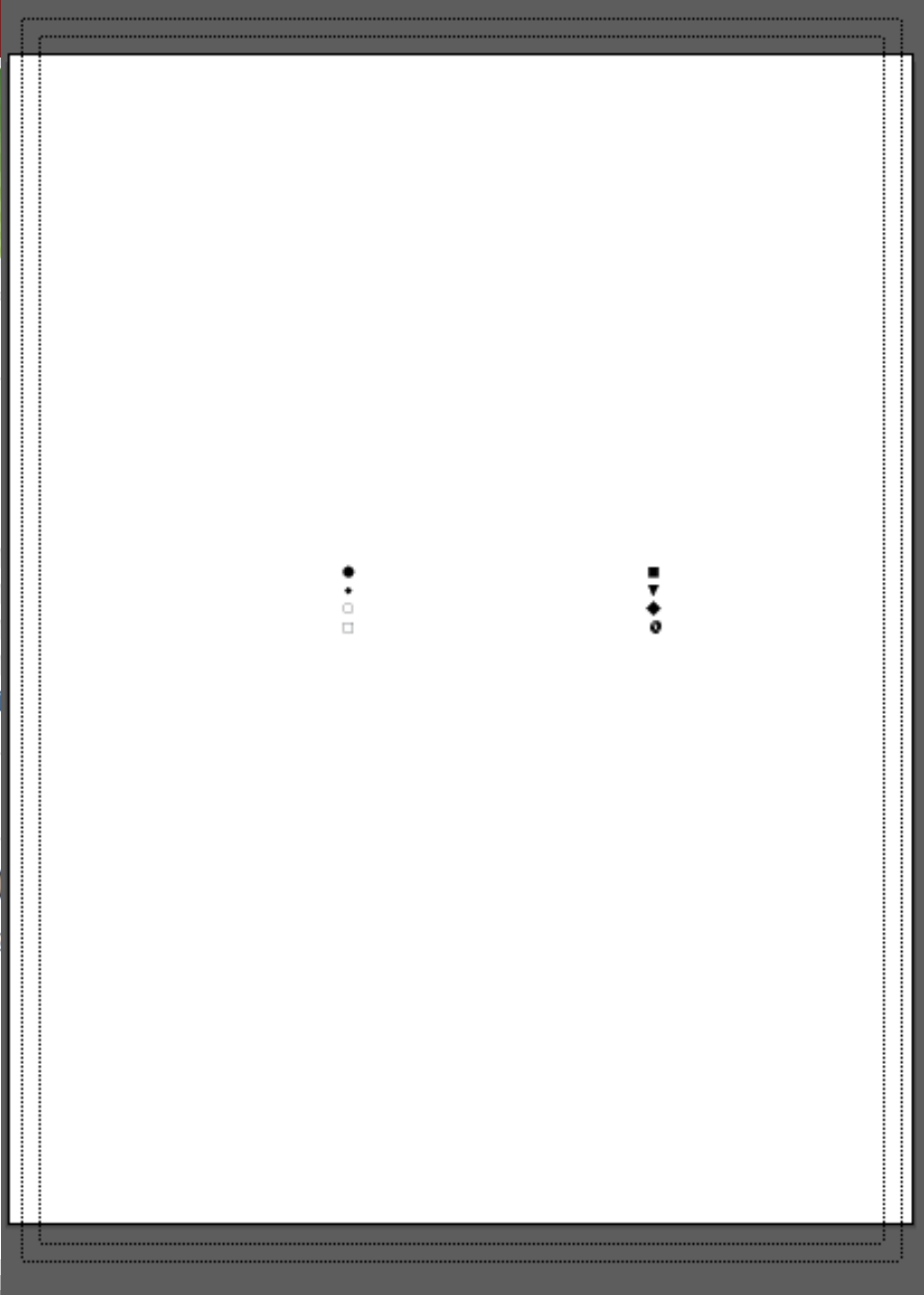
June 12, 2013 | By [uk-guy567](#) | Version: Ai Viewer 3.2

#### Pros

None found

#### Cons

Crashes on opening under Win7



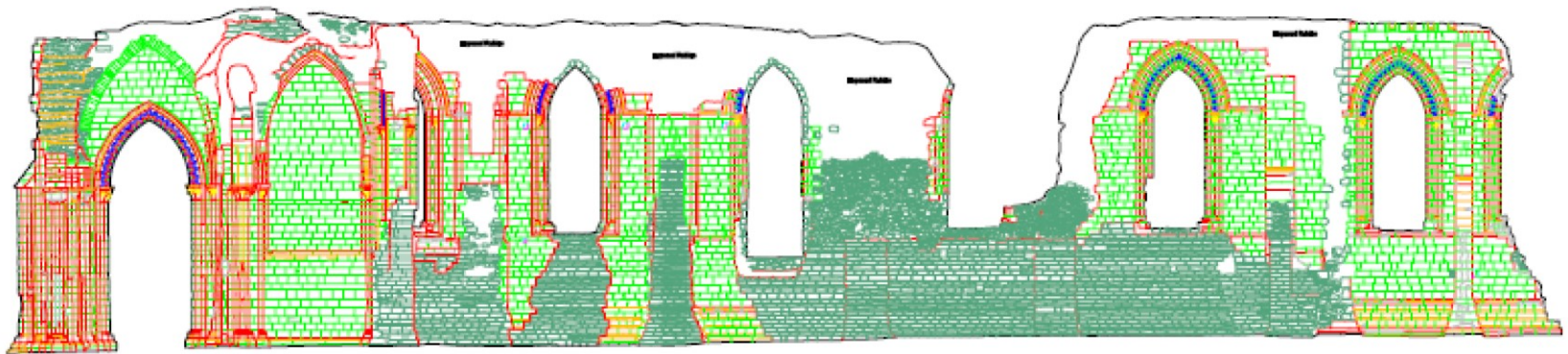
OKS

# So our migration policy is working...



# Cue CAD (vector) migration

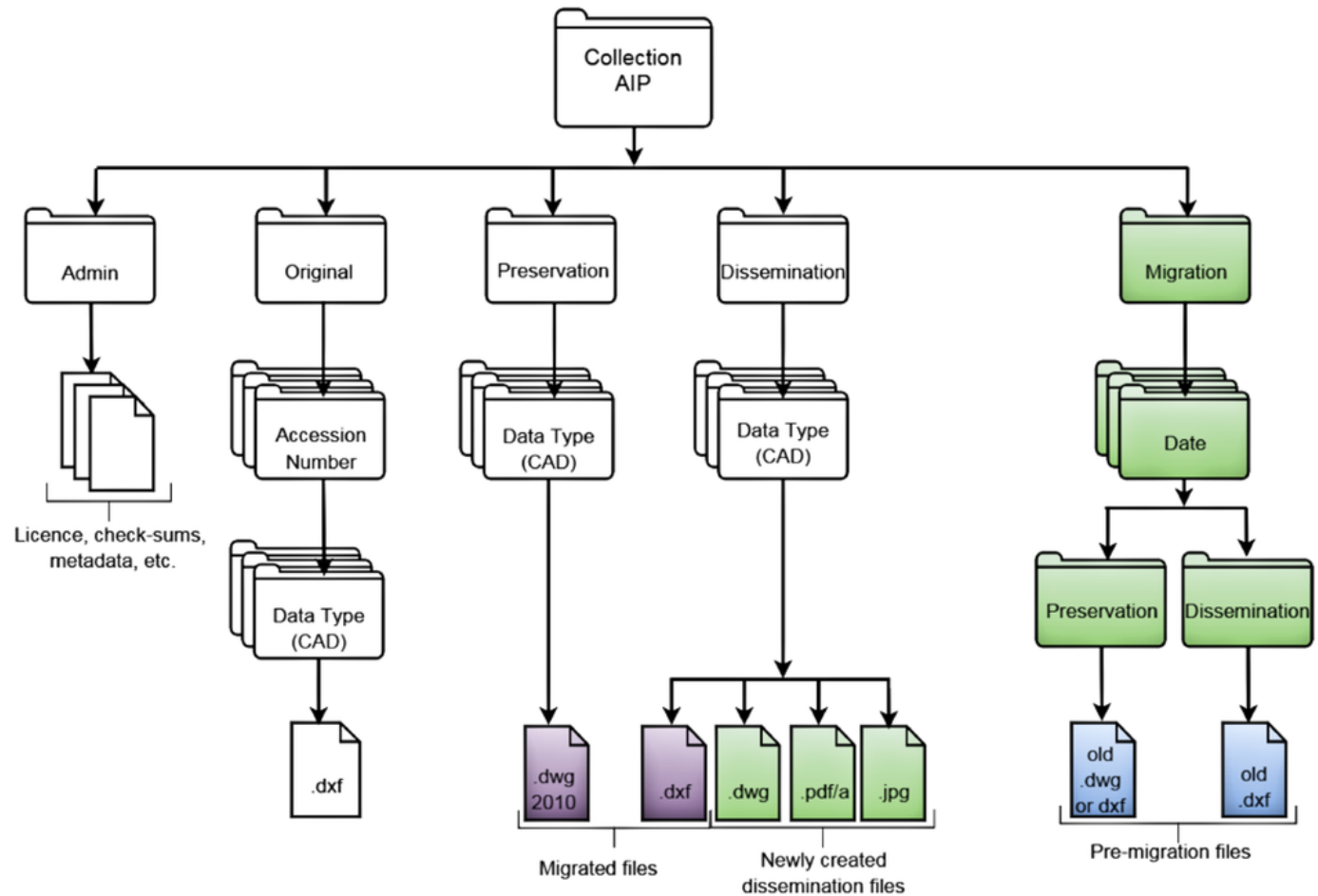
- CAD – defacto Autodesk's AutoCAD software
- ADS had traditionally used DXF (R14) in AIP and DIP support for textual encoding (ASCII) and its primary purpose as an exchange format which could be used beyond Autodesk software



# Cue CAD (vector) migration

- Due to the fast development of the AutoCAD software, the DXF format has seen almost as many version updates as the proprietary DWG format (which has seen eighteen new versions since 1982)
- Effectively using an old format for new content – at risk of losing data!
- As a result, the decision was made in early 2014 to change the ADS archiving policy and adopt DWG version 2010 (AC1024)

# CAD migration



# Lessons

- Did have a cost impact
- Mitigated by file management
- Sometimes had to return to SIP
  - “the ideal migration activity depicted in the OAIS model is not always achievable or appropriate, and strategic decisions have to be made on a case-by-case basis to achieve the best possible result for the long-term preservation of data
- Perhaps most importantly: file format not as important as knowing what it was we were actually meant to be preserving = importance of metadata



Preservation via migration requires work!

## **Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service**

*ISPRS Int. J. Geo-Inf.* **2016**, 5(4), 44; doi:[10.3390/ijgi5040044](https://doi.org/10.3390/ijgi5040044)

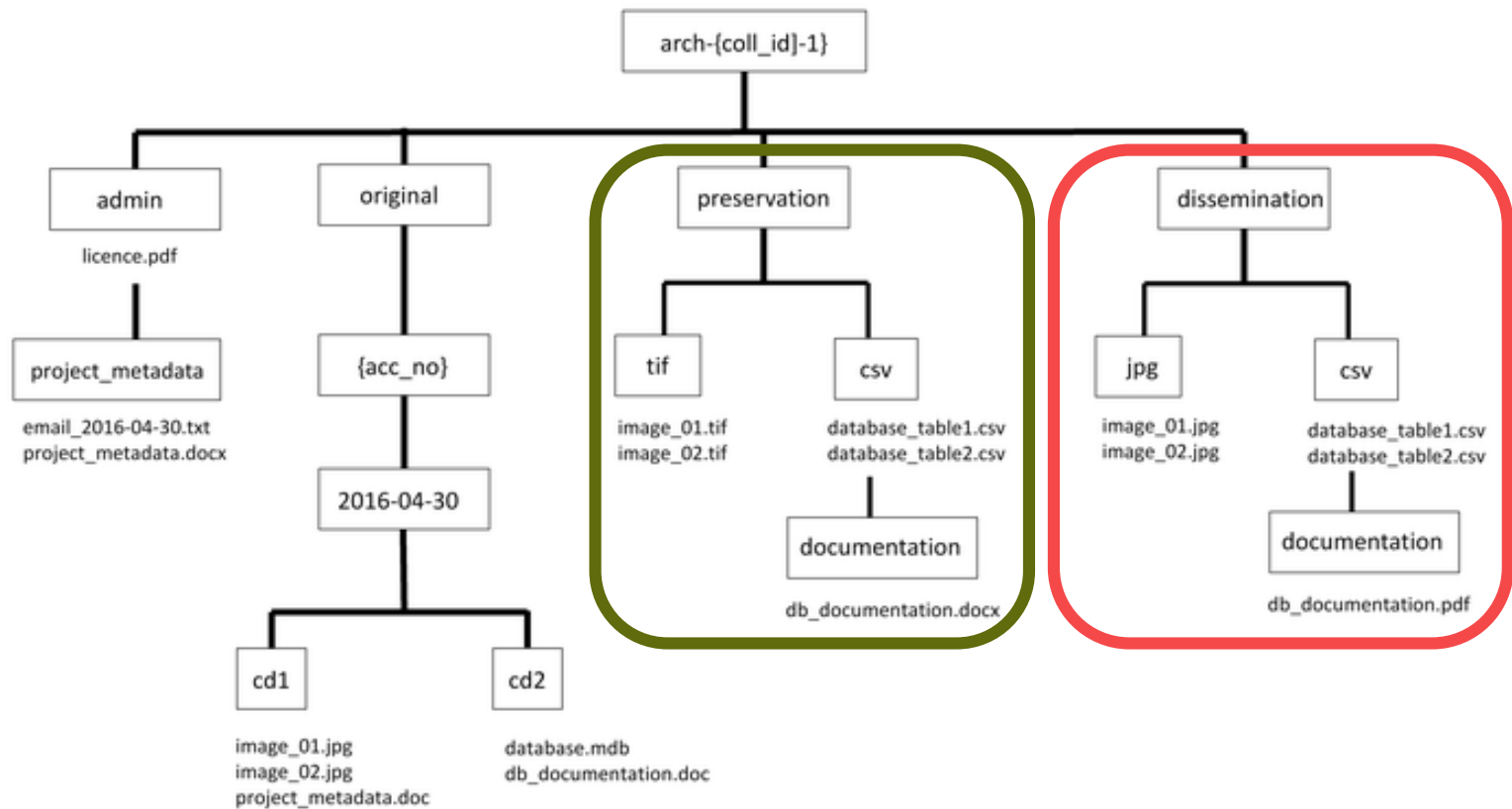
# Future migrations

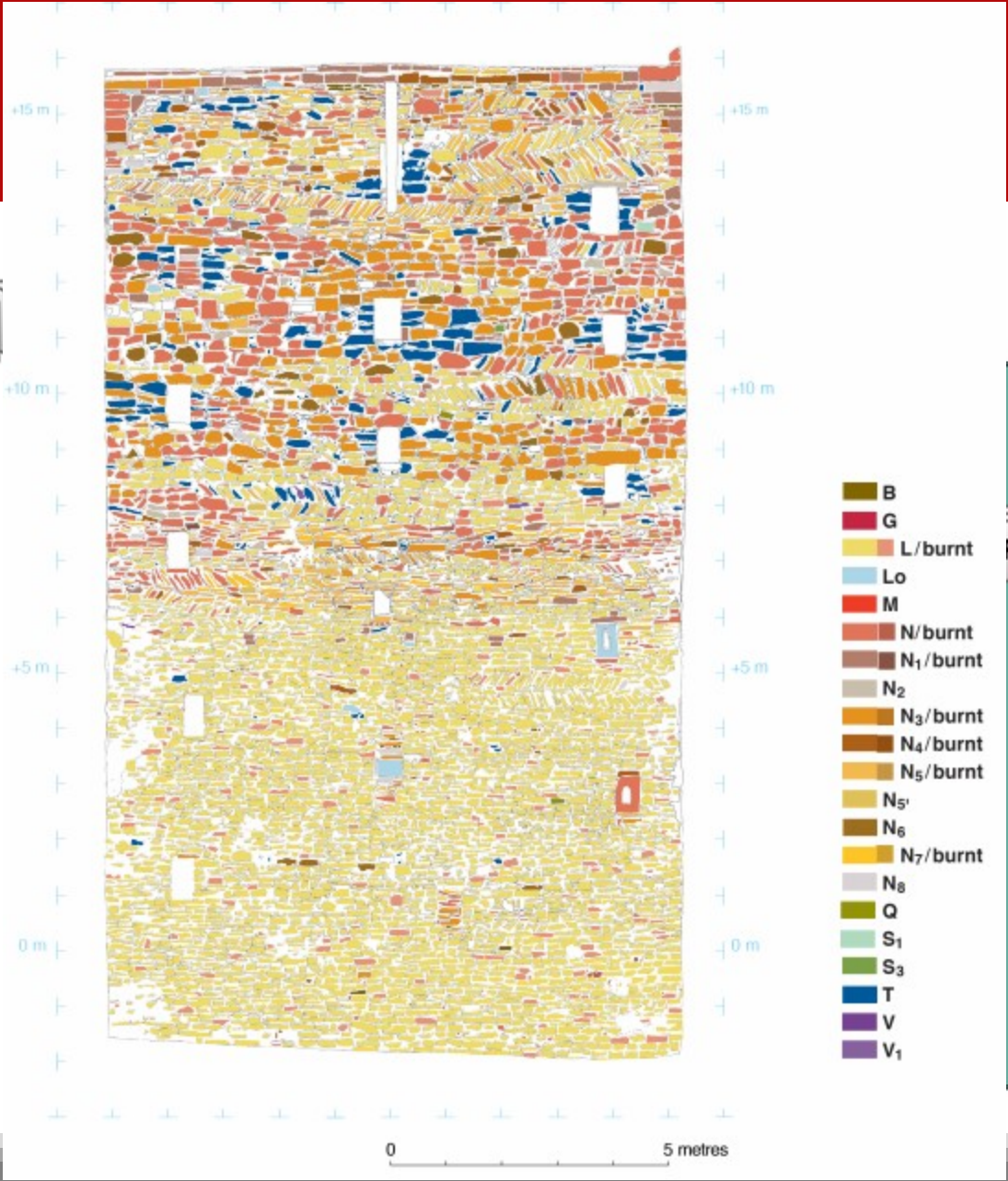
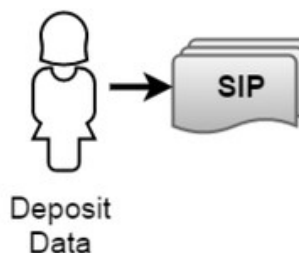
- DOCX
  - When ADS started using OOXML formats were using Office 2010, which only supported the ISO 29500 Transitional standard
  - Although it is anticipated that the transitional format will continue to be supported (including OpenOffice), is it better to move away now?
- GML
  - Slow march to ISO littered with ‘working’ versions

# Conclusions (truisms?)

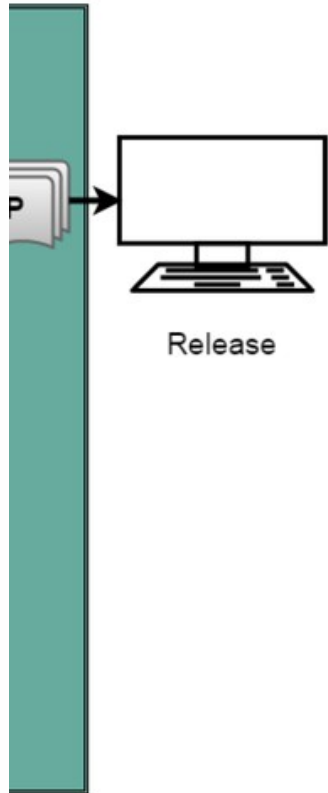
- File obsolescence is *a* factor, but not *the* factor
- File management – what do we have and where is it?
- Perhaps more important is an understanding of normalized formats – program reviews of your formats
- Even more important is your metadata

# Will this ever change?





- B
- G
- L/burnt
- Lo
- M
- N/burnt
- N<sub>1</sub>/burnt
- N<sub>2</sub>
- N<sub>3</sub>/burnt
- N<sub>4</sub>/burnt
- N<sub>5</sub>/burnt
- N<sub>5'</sub>
- N<sub>6</sub>
- N<sub>7</sub>/burnt
- N<sub>8</sub>
- Q
- S<sub>1</sub>
- S<sub>3</sub>
- T
- V
- V<sub>1</sub>



# Thankyou

[tim.evans@york.ac.uk](mailto:tim.evans@york.ac.uk)

<http://archaeologydataservice.ac.uk/blog/>