ads ARCHAEOLOGY DATA SERVICE

Challenges in Digital Preservation – an ADS perspective

Dr Tim Evans 21-07-2016

ads ARCHAEOLOGY DATA SERVICE

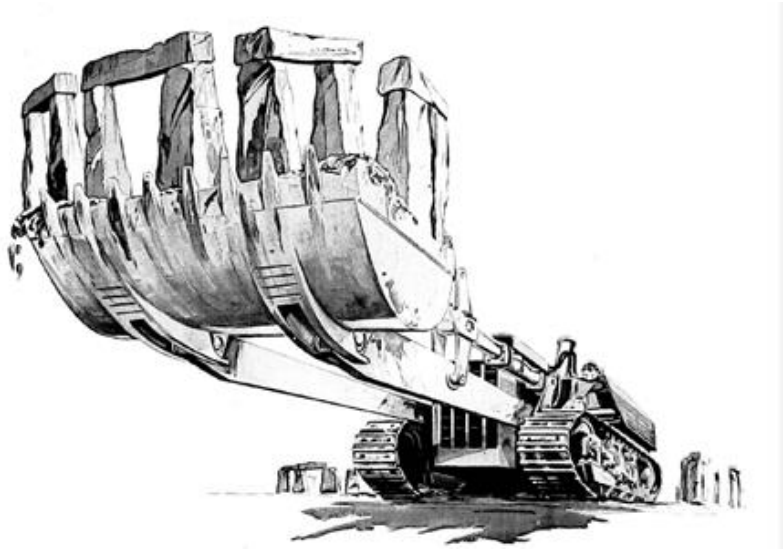
Introduction

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?

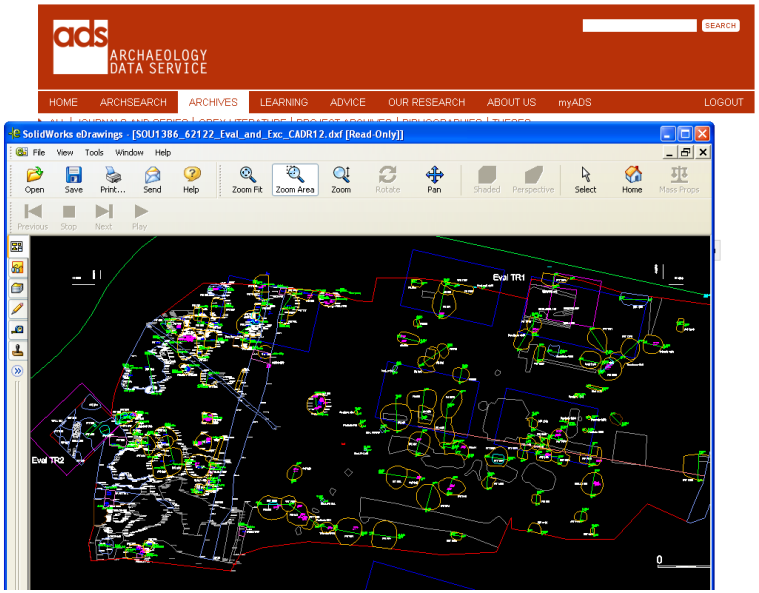
- Is destructive!



- 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



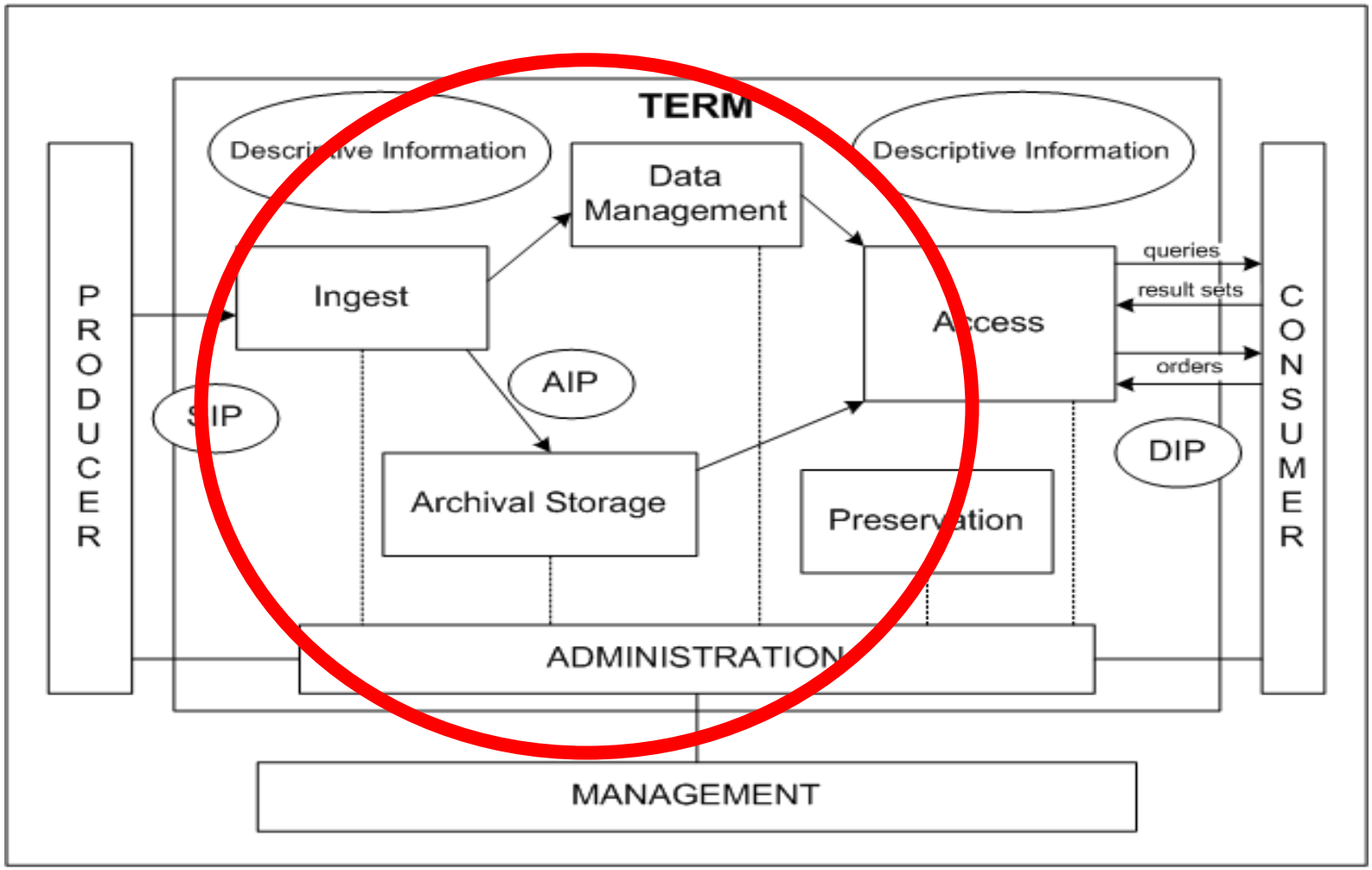


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

http://archaeologydataservice.ac.uk/advice/preservation


DISASTER!

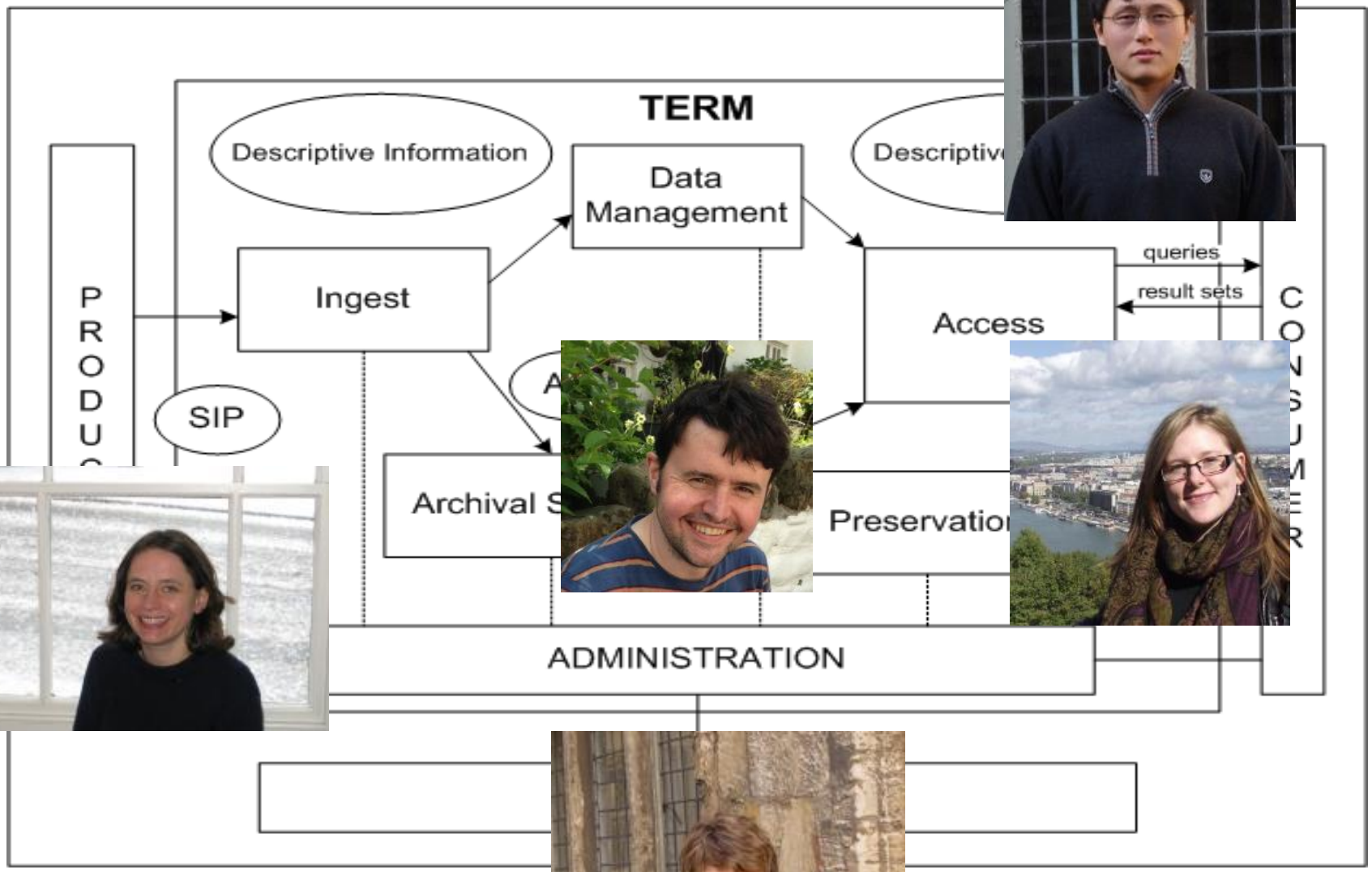


Fig. 1. Major functions of the OAIS Re tative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W- e Methodology Abstract Standard. (OAIS). White Book. Issue 1 r Space Data System Standards.

Funding

Arts & Humanities Research Council

JISC

Historic England

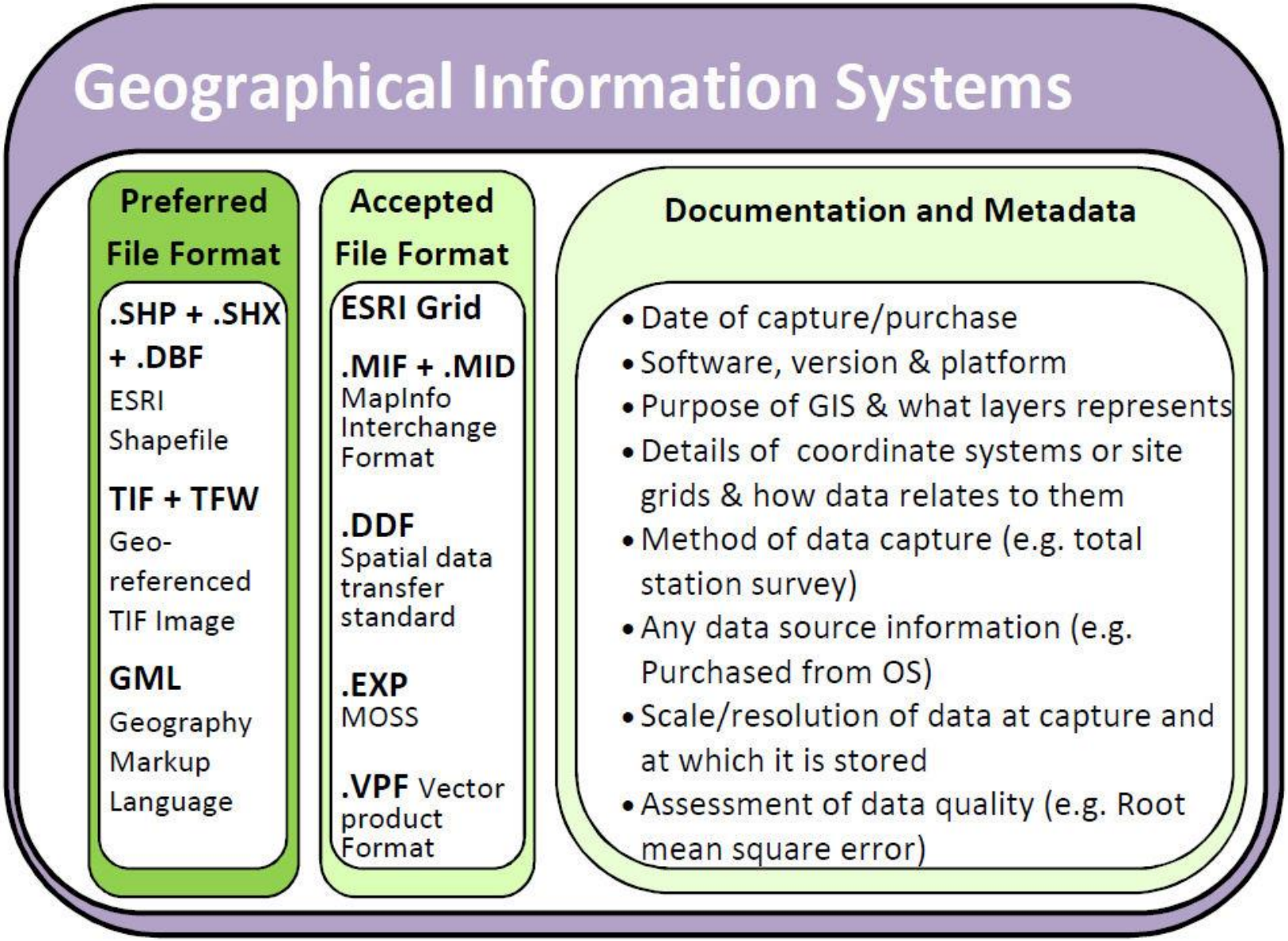
Culture

NATURAL ENVIRONMENT RESEARCH COUNCIL

MEDIN

marine environmental data & information network

Problem 1: People



http://archaeologydataservice.ac.uk/advice

ARCHAEOLOGY DATA SERVICE

M

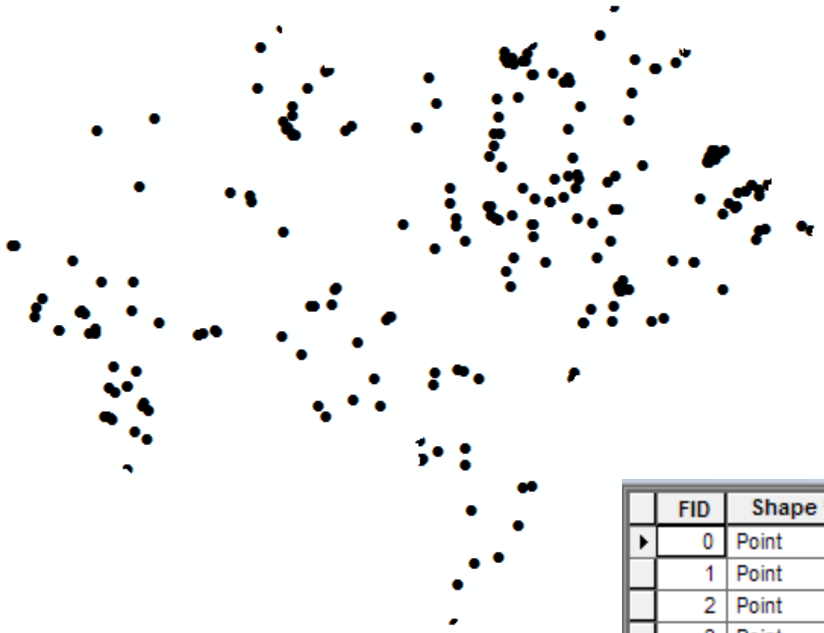
dia

Mul

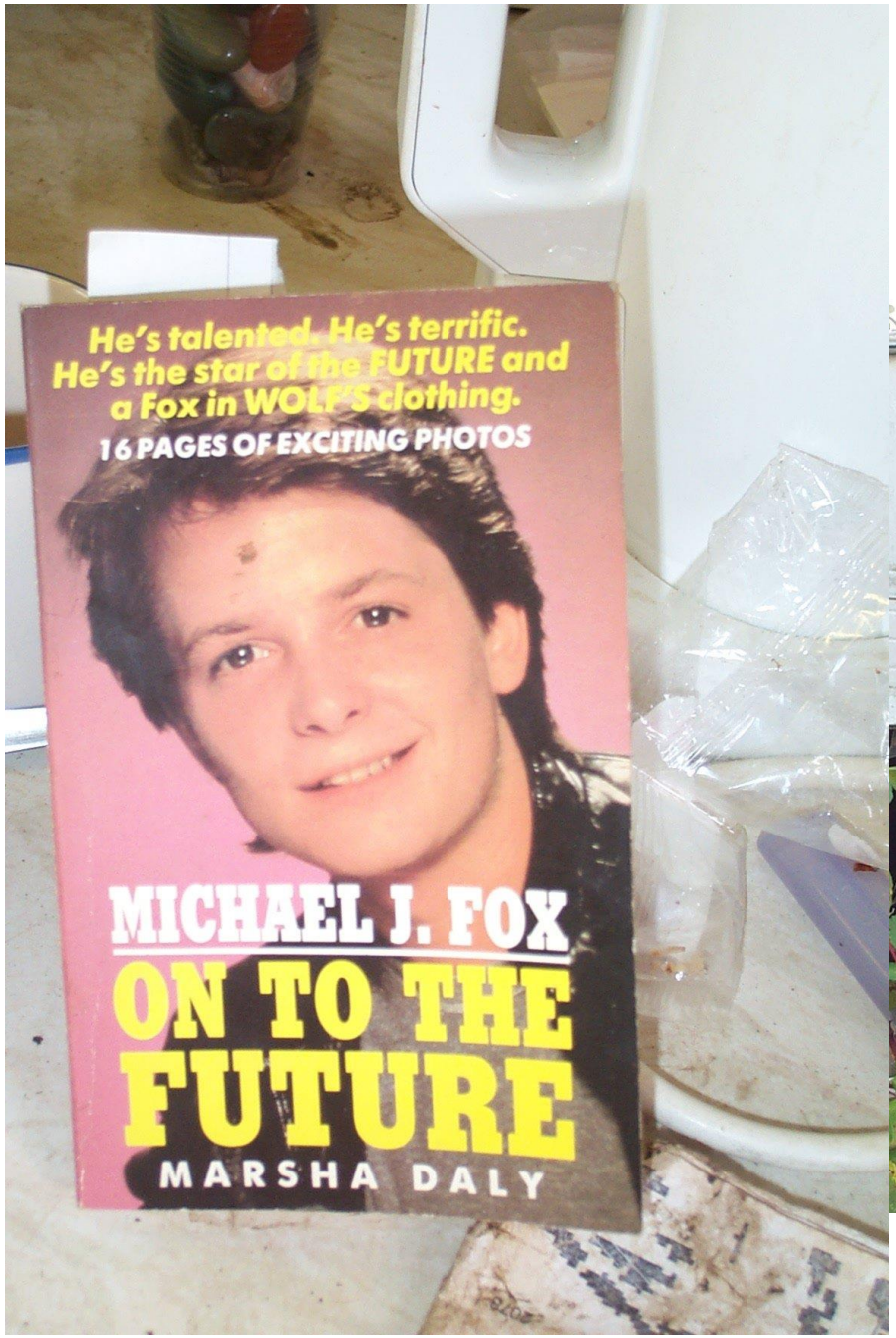
CompactFlash

ads ARCHAEOLOGY DATA SERVICE

Museum of Archaeological Computing



	FID	Shape *	nyexcid	nmr_id	her_event_
▶	0	Point	ny0041	1023794	EYD6391
	1	Point	ny0043	1023854	
	2	Point	ny0044	1023984	
	3	Point	ny0049	1028653	EYD6415
	4	Point	ny0051	1030482	
	5	Point	ny0052	1030840	ENY5461
	6	Point	ny0053	1030859	ENY1032
	7	Point	ny0058	1032430	
	8	Point	ny0059	1032435	
	9	Point	ny0060	1032437	
	10	Point	ny0074	1033498	
	11	Point	ny0078	1034566	ENY1793
	12	Point	ny0083	1034965	
	13	Point	ny0085	1034976	
	14	Point	ny0086	1034980	
	15	Point	ny0087	1034981	
	16	Point	ny0089	1035020	
	17	Point	ny0092	1035087	
	18	Point	ny0098	1035182	EYD6521
	19	Point	ny0099	1035195	ENYM587
	20	Point	ny0104	1035291	EYD6527
	21	Point	ny0116	1035938	EYD6549



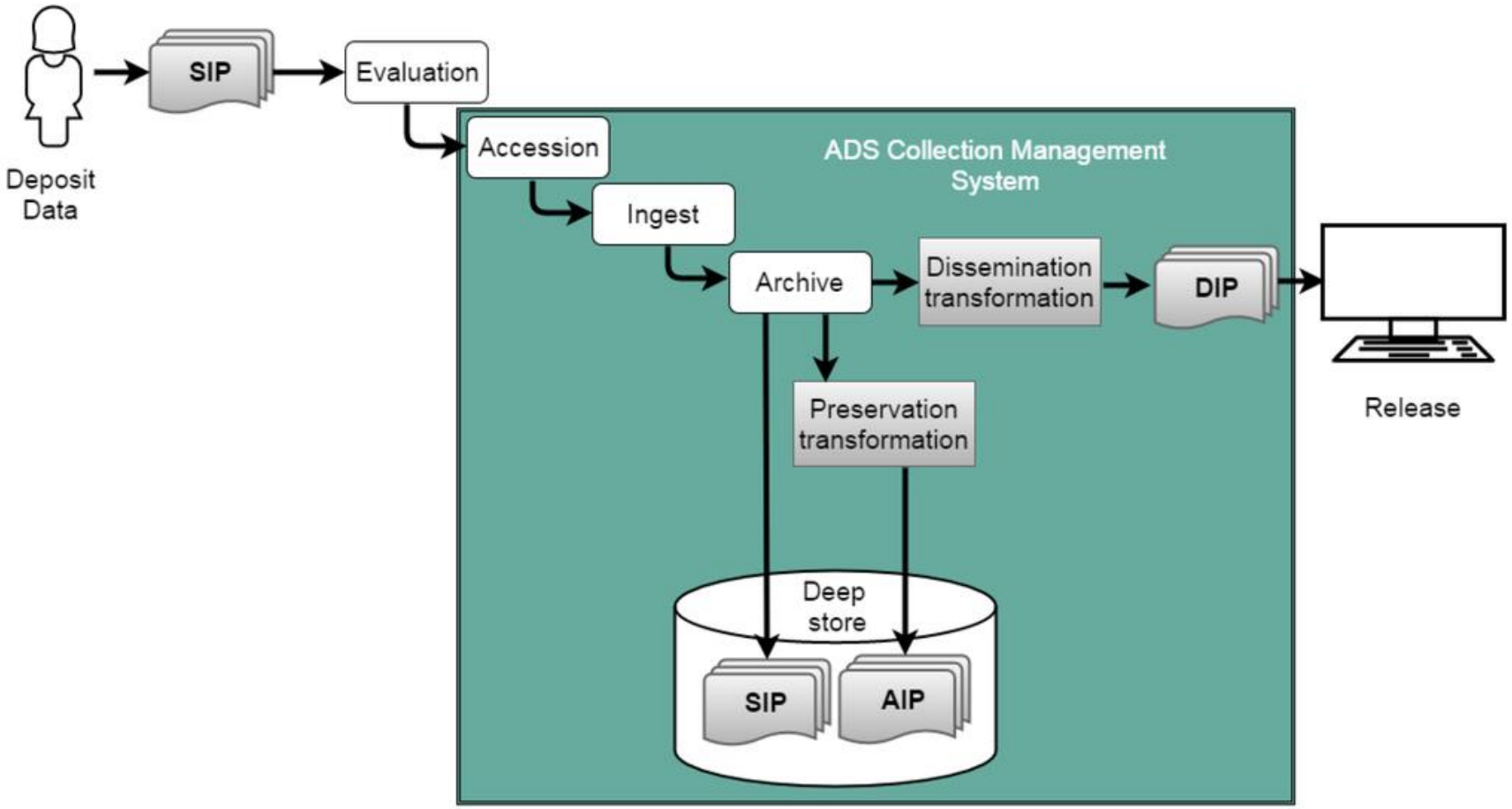
doi:10.5284/1011888

Problem 2: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.

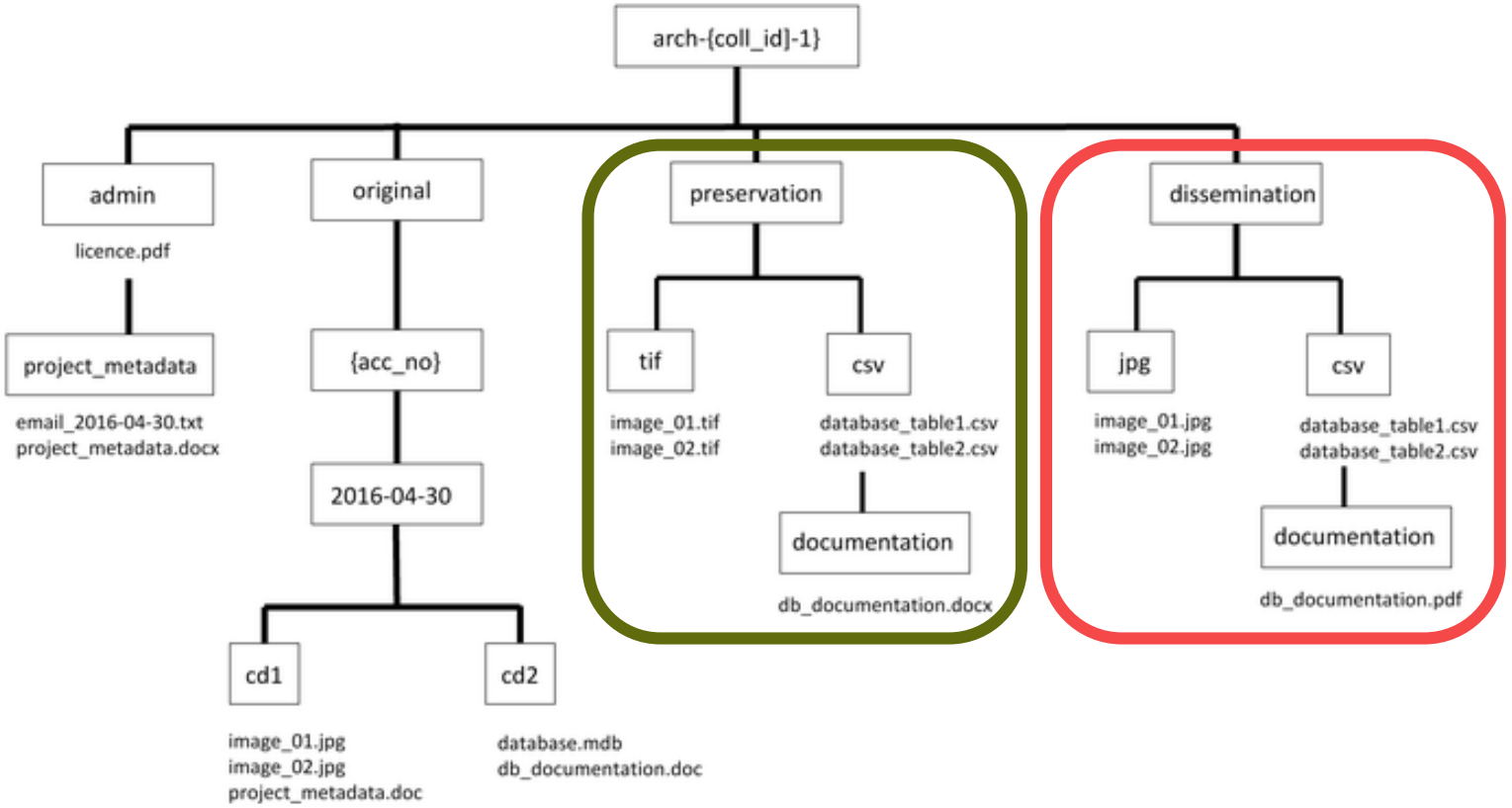
ads ARCHAEOLOGY DATA SERVICE

Operational workflow



ads ARCHAEOLOGY DATA SERVICE

Example structure



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 28 unique formats used in the AIP

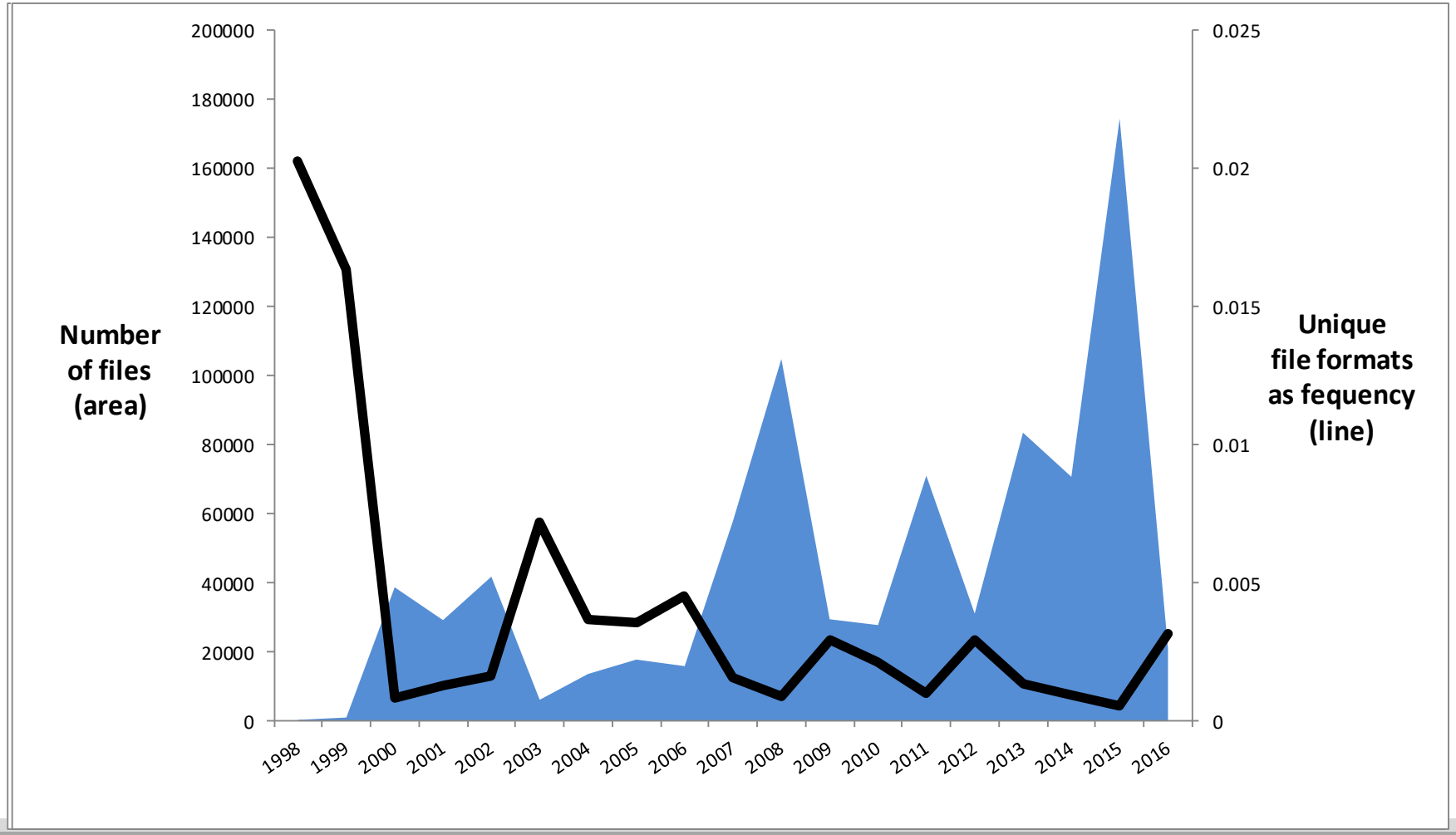
ARCHAEOLOGY DATA SERVICE

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)

ads ARCHAEOLOGY DATA SERVICE

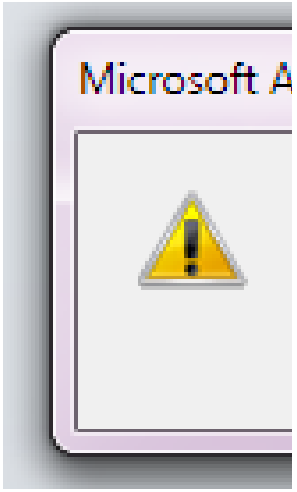
More data – less variation



ads ARCHAEOLOGY DATA SERVICE

"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:

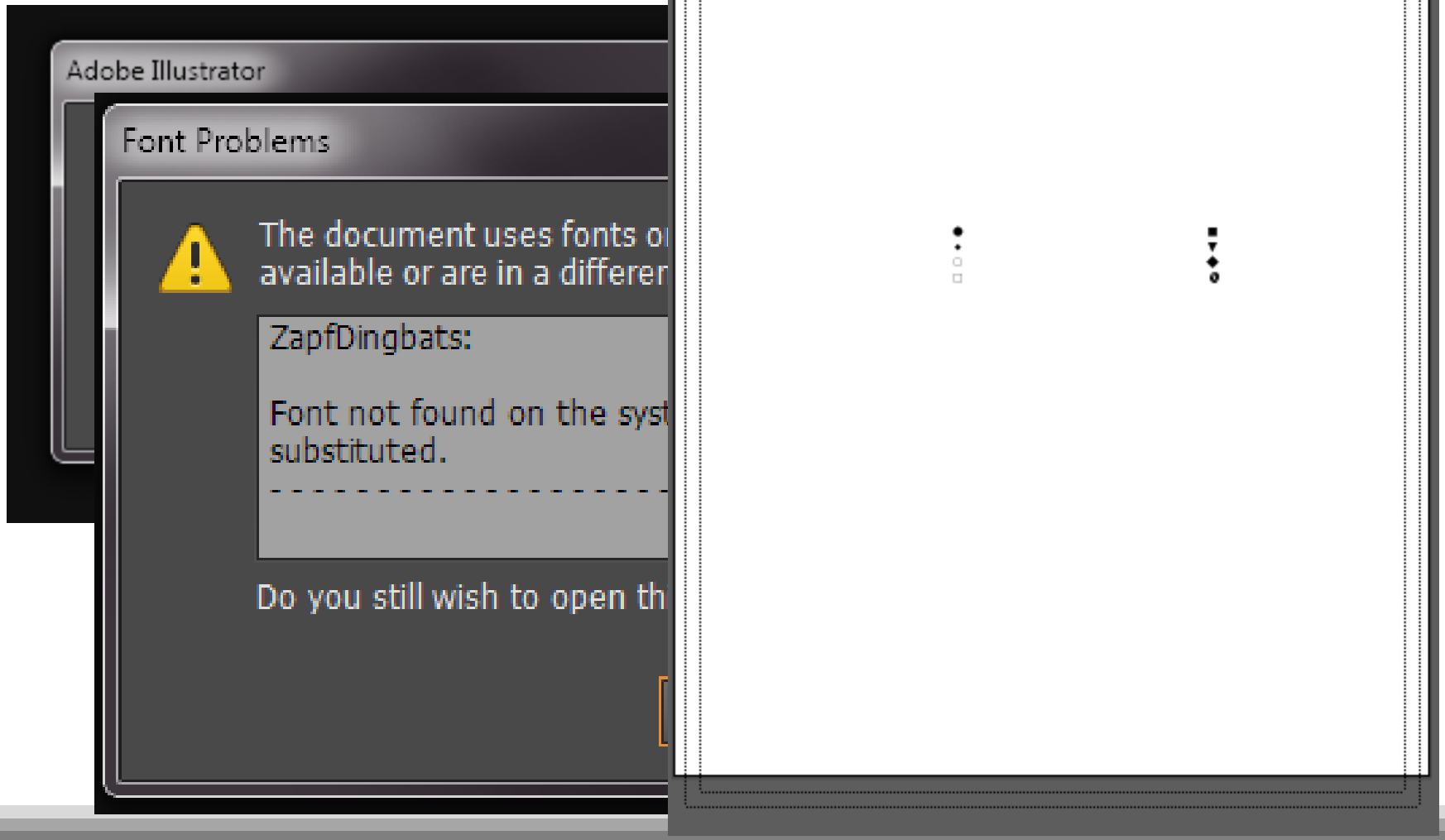
1. Open the Access 97 database in Access 2003.
2. On the **Tools** menu, click **Database Utilities**, click **Convert Database**, and then click **to Access 2002-2003 file format**.
3. Enter a name for the database, and then click **Save**.
4. Exit Access 2003.
5. Open the database in Access 2013 or 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.

Unusual formats

- 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


Adobe Illustrator
Font Problems
The document uses fonts o
available or are in a differer
ZapfDingbats:
Font not found on the syst
substituted.
Do you still wish to open th

User Reviews

Current Version		All Versions	
out of 14 votes		out of 14 votes	
5 star	4	5 star	
4 star	0	4 star	
3 star	0	3 star	
2 star	0	2 star	
1 star	10	1 star	

Sort: Date

★★★★★ "If you want to waste 10

June 12, 2013 | By uk-guy567 | Version: Ai Viewer 3.2

Pros

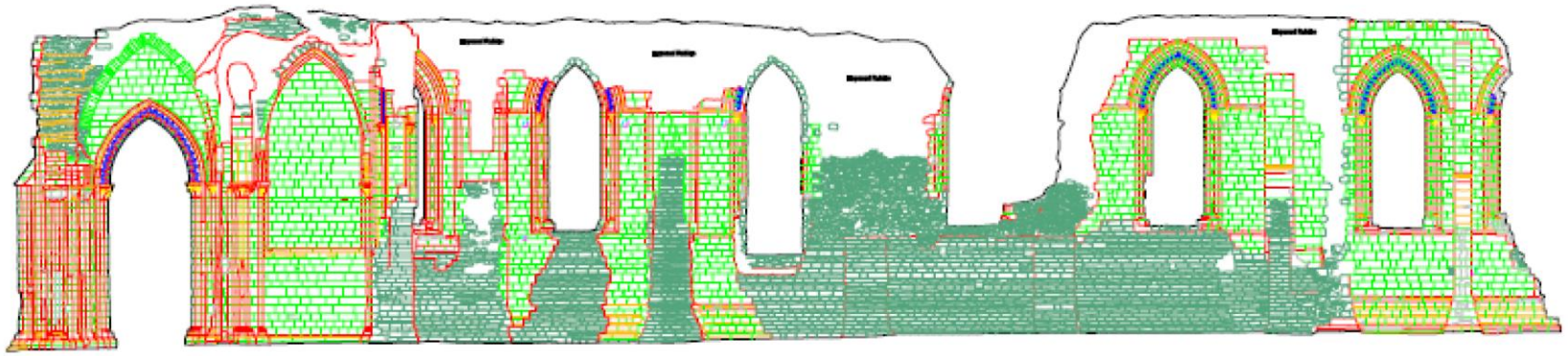
None found

Cons

Crashes on opening under Win7

CAD 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

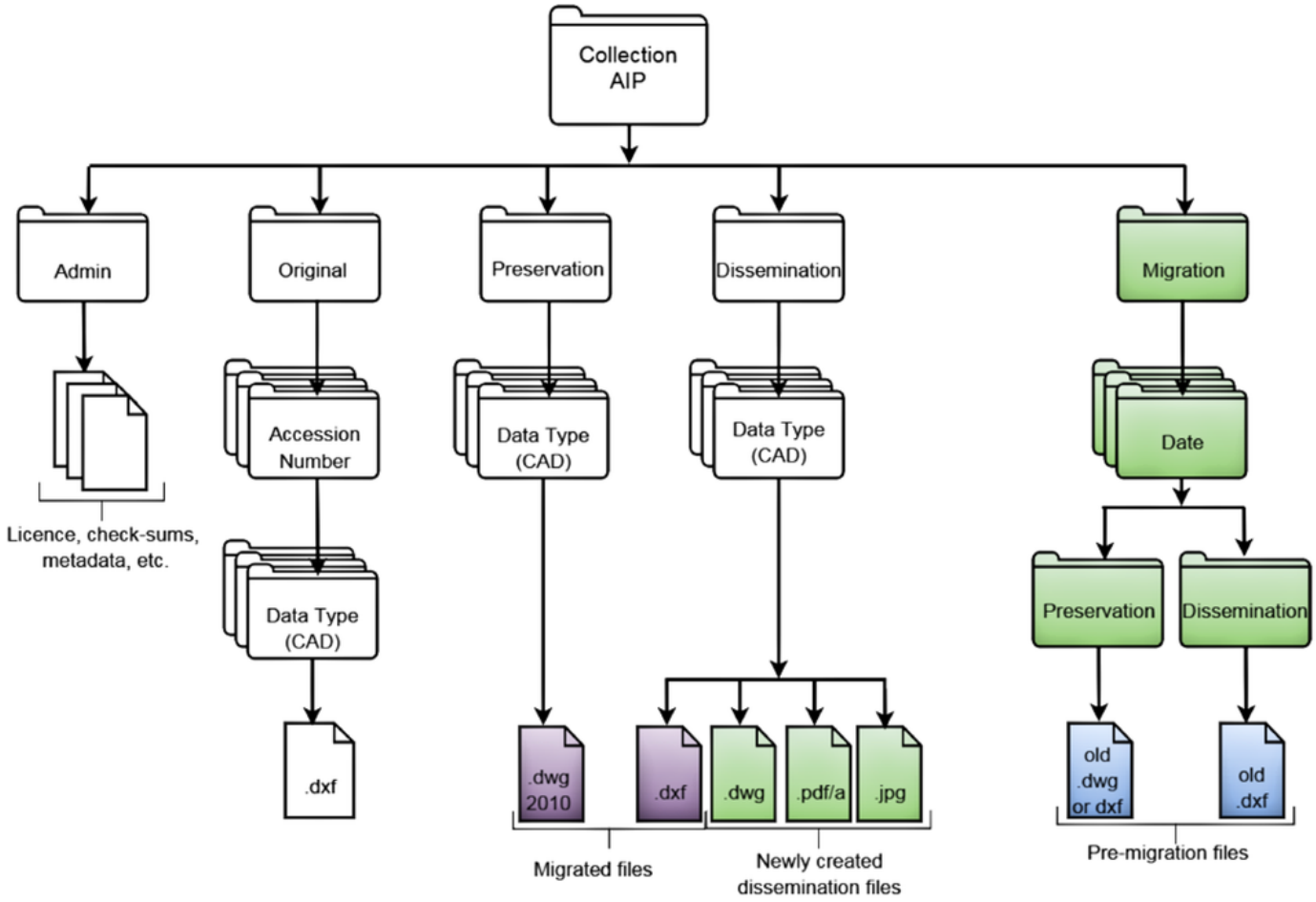


CAD migration

- 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)

ARCHAEOLOGY DATA SERVICE

CAD migration



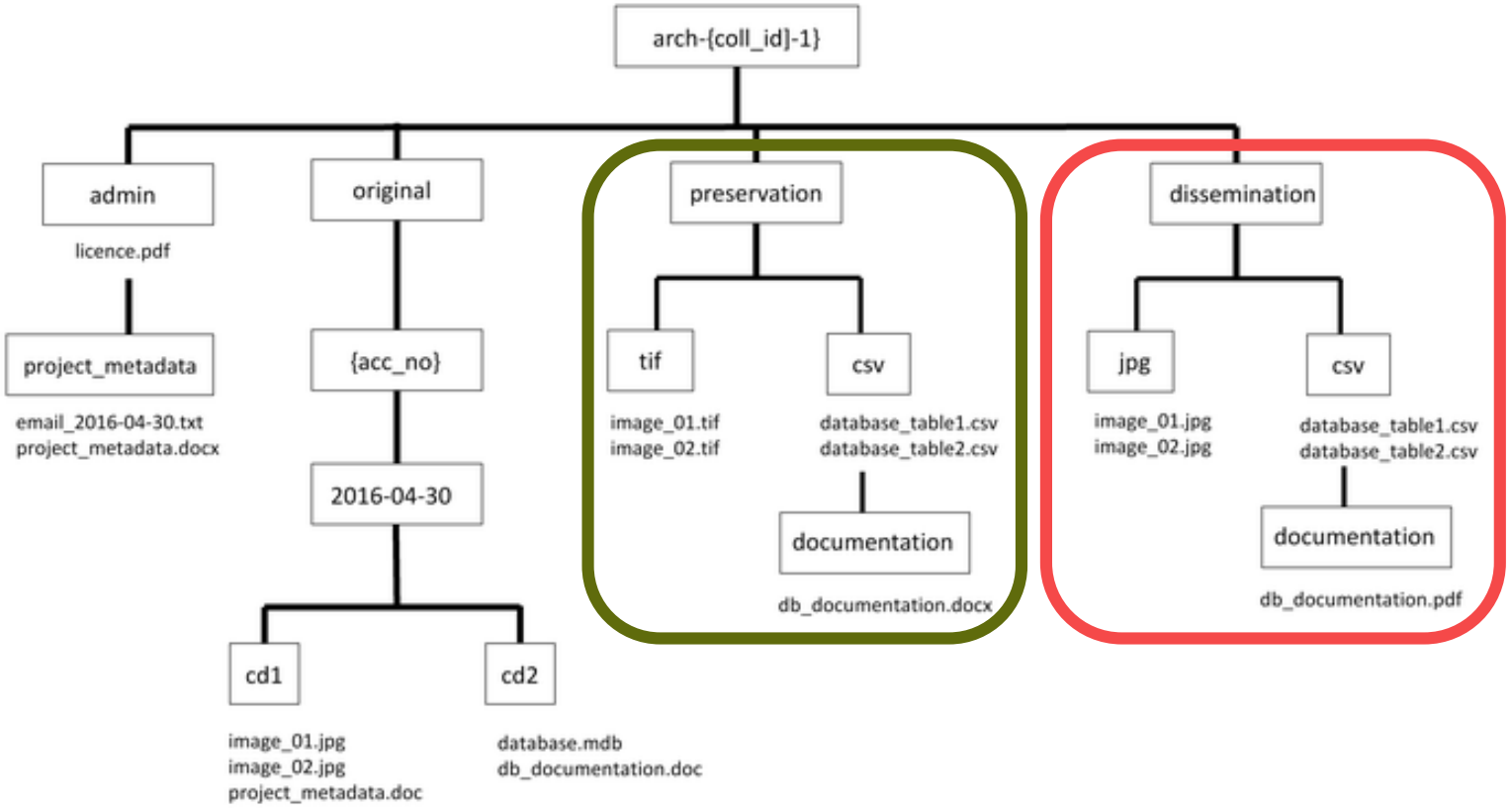
Lessons

- Did have a cost impact
- Mitigated by file management
- Perhaps most importantly: file format not as important as knowing what it was we were actually meant to be preserving = importance of metadata

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

ads ARCHAEOLOGY DATA SERVICE

Will this ever change?

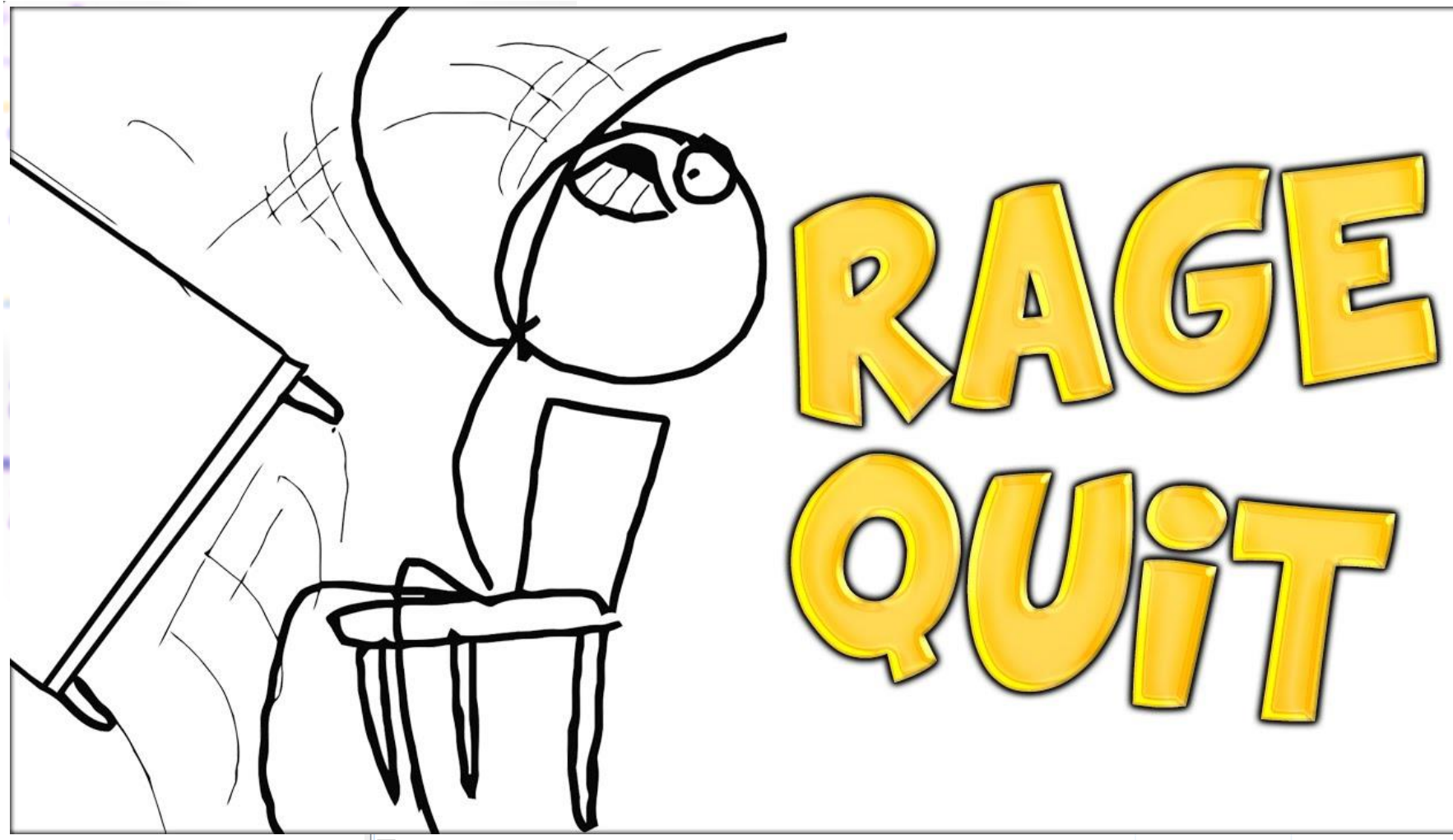


Future migrations

- 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?

Problem 3: File management

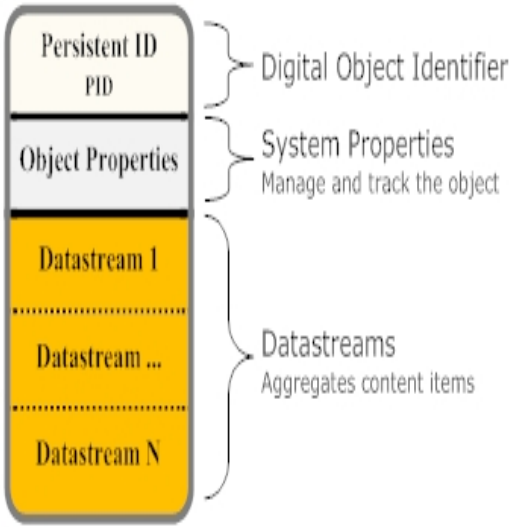
- Query of the ADS Object Management System returned:
  - 2,118,052 files
  - 20+ Tb
- How do we know what we have?
- How do we manage 'objects' comprising different files/data streams?



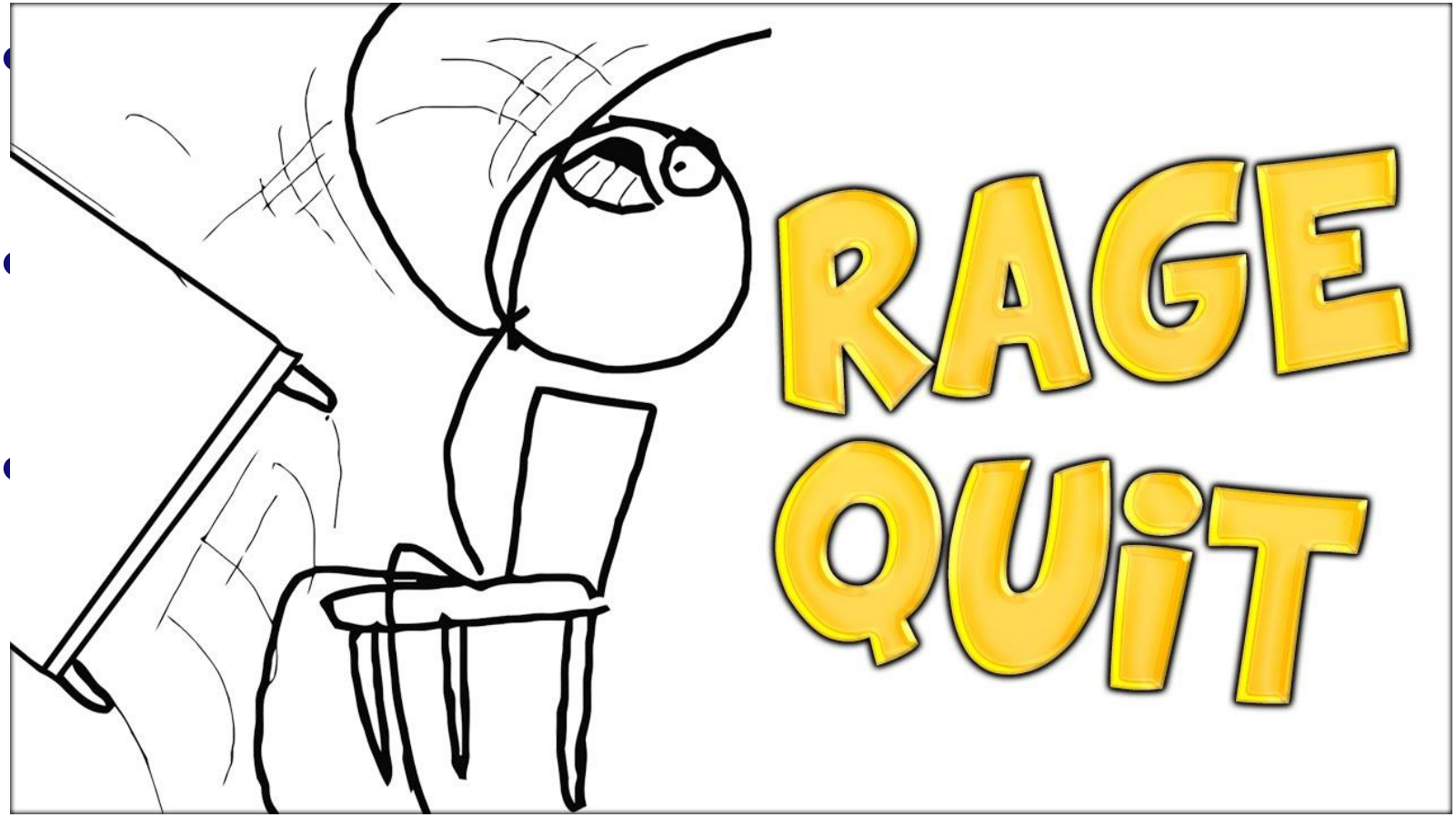
E_Facing_Section_of_1810_Ditch.tif	59,084 KiB	14/06/2016 16:10:46
E_Facing_Section_of_ditch_1810.tif	59,084 KiB	14/06/2016 16:10:48
General_Shot_of_Trench_13.tif	59,084 KiB	14/06/2016 16:10:38
General_Shot_of_Trench-16.tif	59,084 KiB	14/06/2016 16:10:49
IMG_0024.tif	59,084 KiB	14/06/2016 16:10:39
IMG_0039.tif	59,084 KiB	14/06/2016 16:10:41
IMG_0041.tif	59,084 KiB	14/06/2016 16:10:42
IMG_0044.tif	59,084 KiB	14/06/2016 16:10:43
IMG_0048.tif	59,084 KiB	14/06/2016 16:10:45

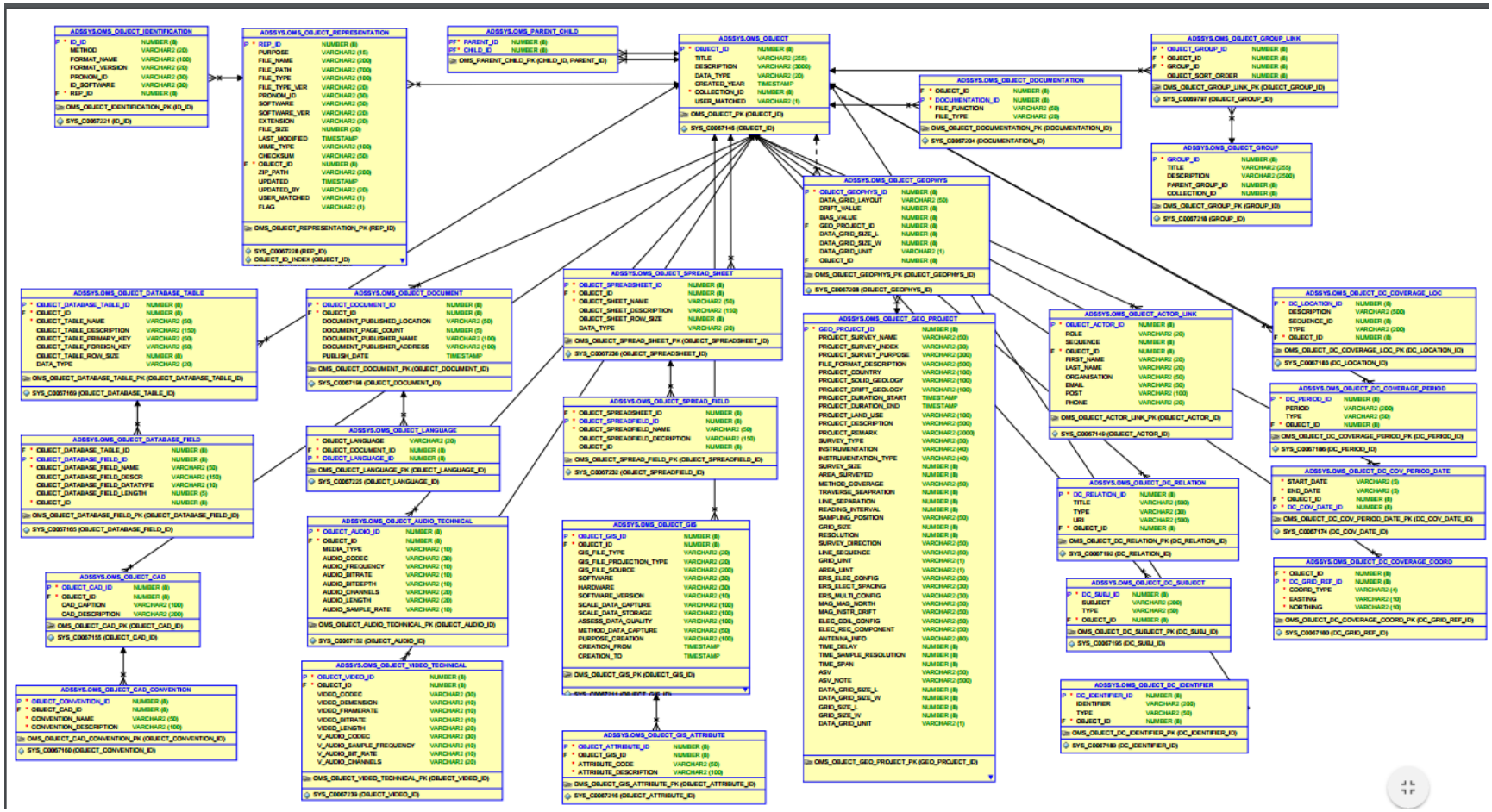
Fedora

- Implementing Fedora (Flexible Extensible Digital Object Repository Architecture).



Fedora



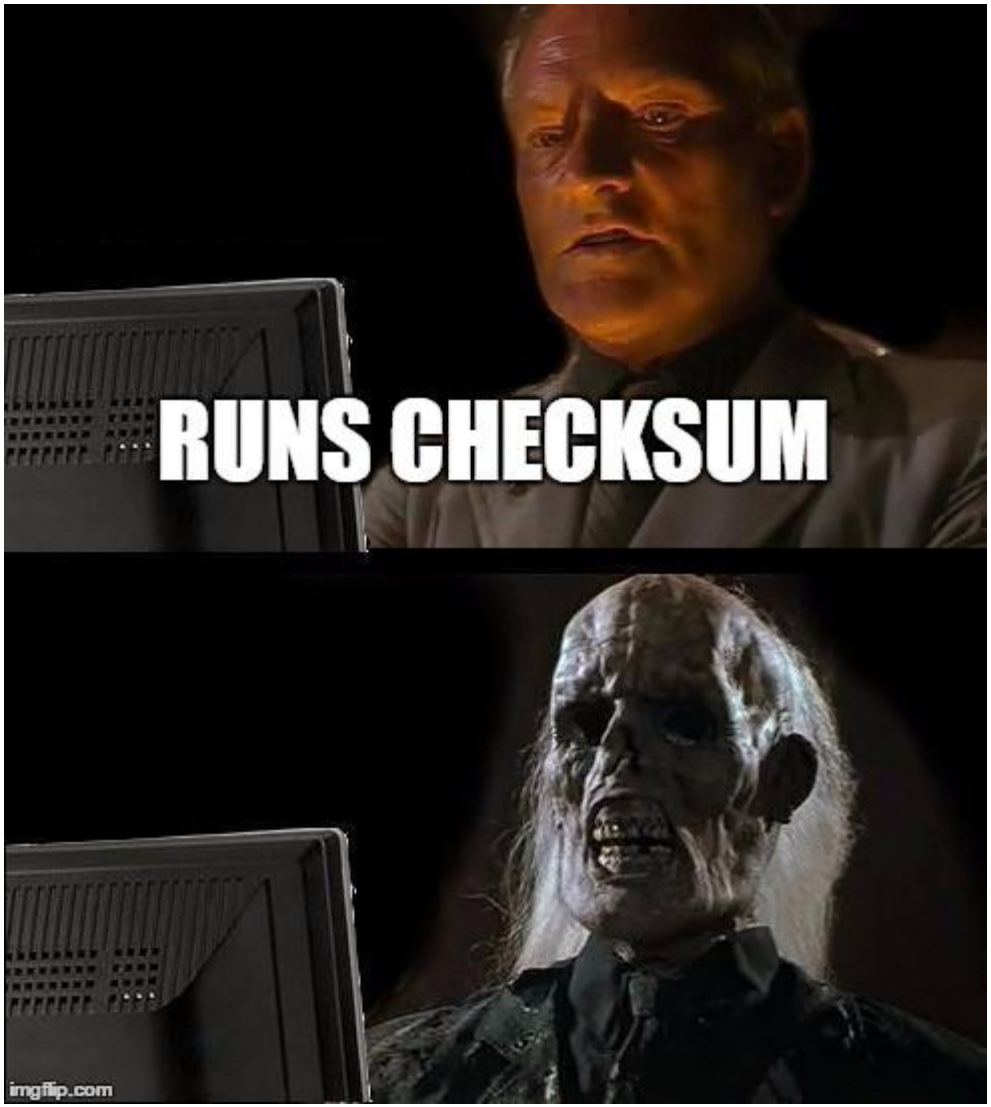

ADSSYS.OMS_OBJECT_IDENTIFICATION
ADSSYS.OMS_OBJECT_REPRESENTATION
ADSSYS.OMS_PARENT_CHILD
ADSSYS.OMS_OBJECT
ADSSYS.OMS_OBJECT_DOCUMENTATION
ADSSYS.OMS_OBJECT_GROUP_LINK
ADSSYS.OMS_OBJECT_GROUP
ADSSYS.OMS_OBJECT_GEOPHYS
ADSSYS.OMS_OBJECT_GEO_PROJECT
ADSSYS.OMS_OBJECT_ACTOR_LINK
ADSSYS.OMS_OBJECT_DC_COVERAGE_LOC
ADSSYS.OMS_OBJECT_DC_COVERAGE_PERIOD
ADSSYS.OMS_OBJECT_DC_COV_PERIOD_DATE
ADSSYS.OMS_OBJECT_DC_COVERAGE_COORD
ADSSYS.OMS_OBJECT_DC_RELATION
ADSSYS.OMS_OBJECT_DC_SUBJECT
ADSSYS.OMS_OBJECT_DC_IDENTIFIER
ADSSYS.OMS_OBJECT_DATABASE_TABLE
ADSSYS.OMS_OBJECT_DATABASE_FIELD
ADSSYS.OMS_OBJECT_CAD
ADSSYS.OMS_OBJECT_CAD_CONVENTION
ADSSYS.OMS_OBJECT_DOCUMENT
ADSSYS.OMS_OBJECT_LANGUAGE
ADSSYS.OMS_OBJECT_AUDIO_TECHNICAL
ADSSYS.OMS_OBJECT_VIDEO_TECHNICAL
ADSSYS.OMS_OBJECT_SPREAD_SHEET
ADSSYS.OMS_OBJECT_SPREAD_FIELD
ADSSYS.OMS_OBJECT_GIS
ADSSYS.OMS_OBJECT_GIS_ATTRIBUTE

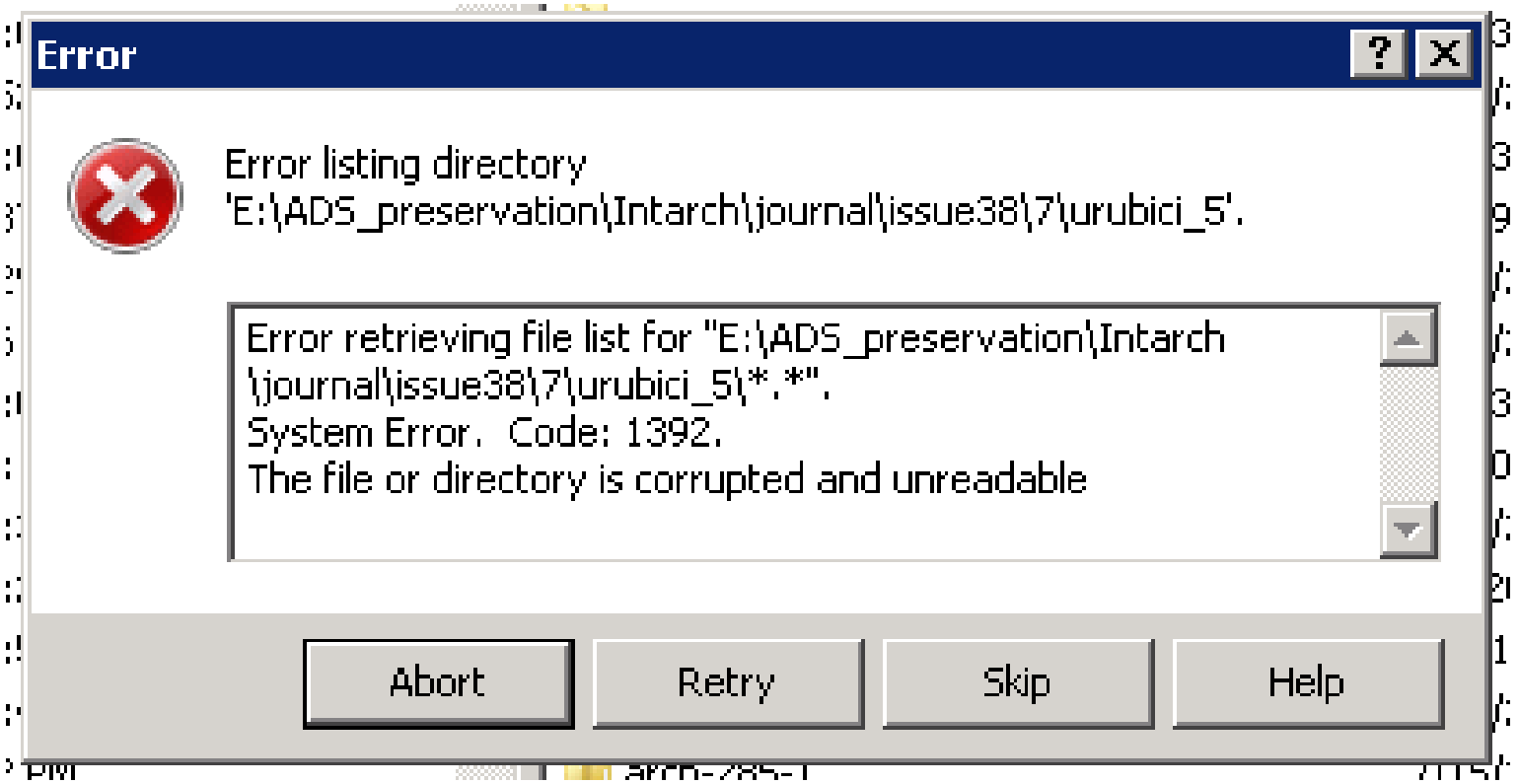
FILE_SIZE	LAST_MODIFI...	MIME_TYPE	CHECKSUM	OBJECT_ID
333,146	2009-09-10 2...	image/jpeg	eb42610ee767aecd6fe0518df4a5ef22	1,000,913
1,217,603	2009-11-21 0...	image/jpeg	f87a9a59175243a559eb5de78756c441	1,001,414
1,217,603	2009-11-21 0...	image/jpeg	f87a9a59175243a559eb5de78756c441	1,001,414
332,348	2009-09-10 2...	image/jpeg	cb1c33f2b613fad68c8f743a6e00003c	1,000,914
15,610,682	2009-09-10 2...	image/tiff	db614cc9ec1ceb1e2066df0d7edfcf4b	1,000,914
14,761,790	2009-11-21 0...	image/tiff	aa233d95956c3e5d08a82e0403beec8c	1,001,414
332,348	2009-09-10 2...	image/jpeg	cb1c33f2b613fad68c8f743a6e00003c	1,000,914
1,228,192	2009-11-21 0...	image/jpeg	512c531d6cb7335c3bd8995cb370c06b	1,001,415
1,228,192	2009-11-21 0...	image/jpeg	512c531d6cb7335c3bd8995cb370c06b	1,001,415
336,111	2009-09-10 2...	image/jpeg	933f3e228143bb06915721021bb934a3	1,000,915
15,610,682	2009-09-10 2...	image/tiff	ec6d354771aee4c941af7bcbbc37ce98	1,000,915
14,761,790	2009-11-21 0...	image/tiff	bd976f4dd2d7343768f2e7e82ef9714a	1,001,415
336,111	2009-09-10 2...	image/jpeg	933f3e228143bb06915721021bb934a3	1,000,915
1,254,455	2009-11-21 0...	image/jpeg	16dd2afe6d27f48496f292b365d6c727	1,001,416
1,254,455	2009-11-21 0...	image/jpeg	16dd2afe6d27f48496f292b365d6c727	1,001,416
493,179	2009-09-10 2...	image/jpeg	a0f96c6e40ef137427a85a219d377f9e	1,000,916
15,610,682	2009-09-10 2...	image/tiff	9c4b0c08b31adc46b7e803b74a3cf85f	1,000,916
14,761,790	2009-11-21 0...	image/tiff	237d20d08e6c1824b3435856d31747b2	1,001,416
493,179	2009-09-10 2...	image/jpeg	a0f96c6e40ef137427a85a219d377f9e	1,000,916
1,225,347	2009-11-21 0...	image/jpeg	7a0dbfbfe03688959febc2a0fcb979c2	1,001,417
1,225,347	2009-11-21 0...	image/jpeg	7a0dbfbfe03688959febc2a0fcb979c2	1,001,417
379,548	2009-09-10 2...	image/jpeg	bed5efb4c92ea5c8d77870913e13a0a6	1,000,917
15,610,682	2009-09-10 2...	image/tiff	e5b0d5c40adc917495971933363e0615	1,000,917
14,761,790	2009-11-21 0...	image/tiff	2e648d0ef43b772595689502988f2f75	1,001,417
379,548	2009-09-10 2...	image/jpeg	bed5efb4c92ea5c8d77870913e13a0a6	1,000,917
1,223,243	2009-11-21 0...	image/jpeg	a9310ac12f2111451b00cdc3af045c43	1,001,418
1,223,243	2009-11-21 0...	image/jpeg	a9310ac12f2111451b00cdc3af045c43	1,001,418
280,368	2009-09-10 2...	image/jpeg	a65f4615e48ca79108e5dc9fe261773f	1,000,918
15,610,682	2009-09-10 2...	image/tiff	0960687e6a7dd8d4591ac37ecdbde040	1,000,918
14,761,790	2009-11-21 0...	image/tiff	1c14156f502dec925b3b84e8b96f0448	1,001,418
280,368	2009-09-10 2...	image/jpeg	a65f4615e48ca79108e5dc9fe261773f	1,000,918
704,225	2009-09-10 2...	image/jpeg	a5babbb8b7ce643f32b8f6e05328b279	1,000,919
704,225	2009-09-10 2...	image/jpeg	a5babbb8b7ce643f32b8f6e05328b279	1,000,919
1,191,687	2009-10-30 0...	image/jpeg	0001015c94683659fed1b70c6c81b4dd	1,001,419
19,633,462	2009-09-10 2...	image/tiff	7aba613922e8f99834e385f8663157d4	1,000,919
18,935,008	2009-10-30 0...	image/tiff	7173e70af02ef985926936e35f0b6c5f	1,001,419
18,935,008	2009-10-30 0...	image/tiff	7173e70af02ef985926936e35f0b6c5f	1,001,419
621,776	2009-09-10 2...	image/jpeg	b32af33660585536127db6006ad1caf5	1,000,920
621,776	2009-09-10 2...	image/jpeg	b32af33660585536127db6006ad1caf5	1,000,920
1,036,506	2009-10-30 0...	image/jpeg	704d25964e545c4e75235f4b2cd2bbc3	1,001,420

1,003,730	Preservation	p0000020.tif	/ADS_preservation/arch-1591-1/preservation/tif/origimages/p0...
1,003,731	Original	p0000020.tif	/ADS_preservation/arch-1591-1/original/3306/2013-08-05/orig...

Object Id	Object Title	Data Type	Created Date	Object Representations
1328034	Map of traditonal farmsteads in Greater Lincolnshire with house located with gable side on to yard	GIS	12-Sep-2015	side_on_point.dbf side_on_point.prj side_on_point.shp side_on_point.shx side_on_point.gml side_on_point.prj side_on_point.xsd side_on_point.zip
1328033	Map of traditonal farmsteads in Greater Lincolnshire with detached farm houses	GIS	12-Sep-2015	detached_point.dbf detached_point.prj detached_point.shp detached_point.shx detached_point.gml detached_point.prj detached_point.xsd detached_point.zip
1328032	Map of traditonal farmsteads in Greater Lincolnshire with zadditonal deatched elements to main plan	GIS	12-Sep-2015	d_point.dbf d_point.prj d_point.shp d_point.shx d_point.gml d_point.prj d_point.xsd d_point.zip

JonathanParkhouse_Warwickshire.pdf
LizPearson_Environmental.pdf
MalcolmAtkin_Worcestershire.pdf
MarkBowden_Earthworks.pdf
MikeHodder_Birmingham.pdf
PaulStamper_Countryside.pdf
AnnWoodward_PotsPitsandMonuments_(2).pdf
JamesGreig_Environmental.pdf
JohnHalsted_BronzeAgeShropshire.pdf


RUNS CHECKSUM
imgflip.com


Error
Error listing directory
'E:\ADS_preservation\Intarch\journal\issue38\7\urubici_5'.
Error retrieving file list for "E:\ADS_preservation\Intarch
\journal\issue38\7\urubici_5*.*".
System Error. Code: 1392.
The file or directory is corrupted and unreadable
Abort
Retry
Skip
Help

What's an object?

1373975	EAA_REPORT_9_VOL_2	Text	09-Nov-2015	EAA_REPORT_9_VOL_2.pdf EAA_REPORT_9_VOL_2.pdf EAA_REPORT_9_VOL_2.pdf EAA_REPORT_9_VOL_2.zip.001 EAA_REPORT_9_VOL_2.zip.002
1373974	EAA_REPORT_9_VOL_1	Text	30-Dec-2015	EAA_REPORT_9_VOL_1.pdf EAA_REPORT_9_VOL_1.pdf EAA_REPORT_9_VOL_1.pdf EAA_REPORT_9_VOL_1.zip.001 EAA_REPORT_9_VOL_1.zip.002

Mis-identification

File Name	**MapInfo-GIS-files-for-the-Historic-Landscape-Character-Areas-and-Zones.TAB**
Rep Id	3356836
Purpose	Original
File Path	/ADS_preservation/arch-2248-1/original/5444/2016-02-04/MapInfo-GIS-files-for-the-Historic-Landscape-Character-Areas-and-Zones.TAB
File Type	
File Type Ver	
Pronom Id	
Extension	tab
File Size	208
Last Modified Date	04-Feb-2016
Mime Type	
Checksum	143d4b27e18936a0a93cdd348bb4ec61
Updated	04-Feb-2016
Updated By	kjn103

File Name	**detached_point.gml**
Rep Id	3324042
Purpose	Preservation
File Path	/ADS_preservation/arch-2128-1/preservation/gml/GIS_Layers_and_Data/Farmstead_data_extractions/House_Position/detached_point.gml
File Type	Extensible Markup Language
File Type Ver	1.0
Pronom Id	fmt/101
Extension	gml
File Size	11672363
Last Modified Date	17-Nov-2015
Mime Type	application/xml, text/xml
Checksum	3ae4[illegible]b75b99b880ae29d
Updated	04-Dec-2015
Updated By	te1

ads ARCHAEOLOGY DATA SERVICE

Thanks!

tim.evans@york.ac.uk

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

https://twitter.com/ADS_Chatter

https://twitter.com/ADS_Update