



Ingest at the Archaeology Data Service: a case study

Jen Mitcham

Archaeology Data Service

jl10@york.ac.uk



Ingest at the ADS: Overview

- What is the ADS?
- Functions of an OAIS?
- The Digital Pipeline
 - Submission (SIP)
 - Archive (AIP)
 - Dissemination (DIP)



What is the ADS?



‘... to support research learning and teaching with high quality and dependable digital resources’



Archaeology wing of AHDS (AHDS Archaeology)

Funded by AHRC, JISC and others



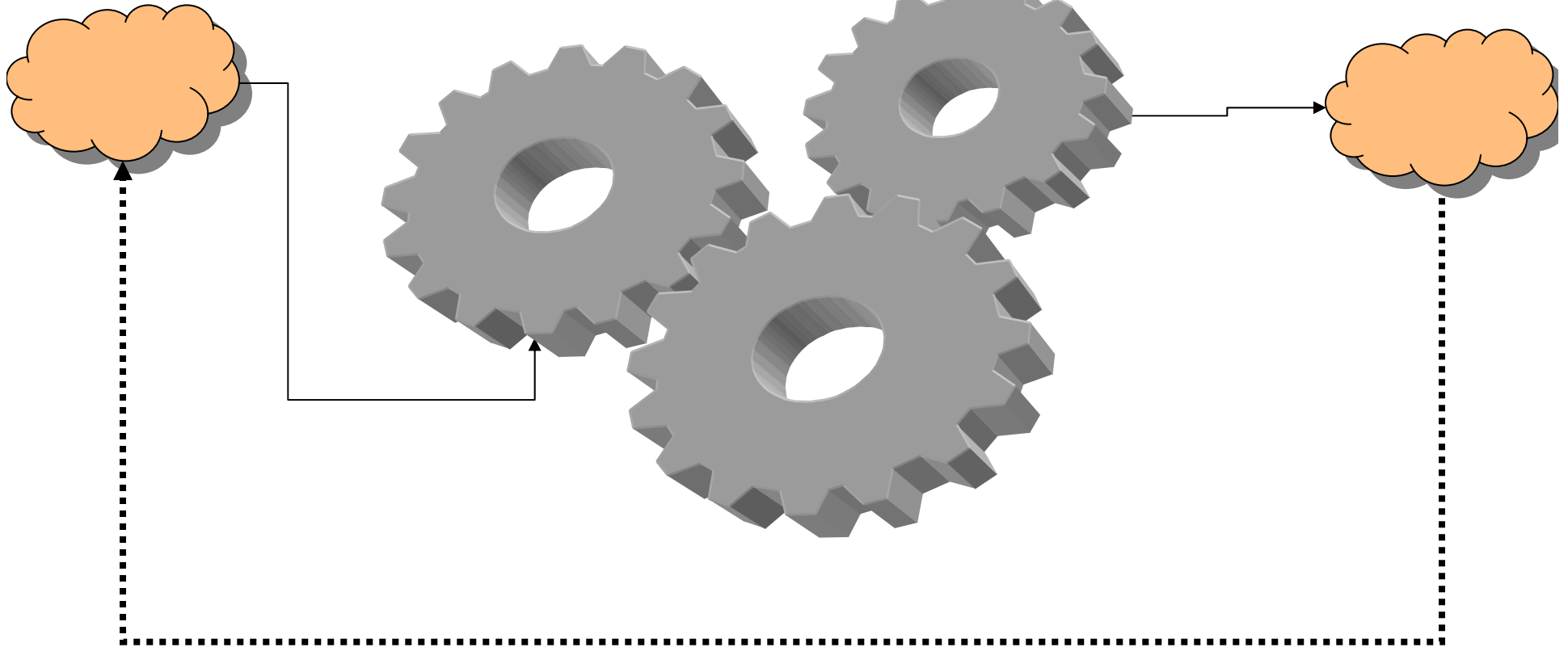
The 4 things we do all day

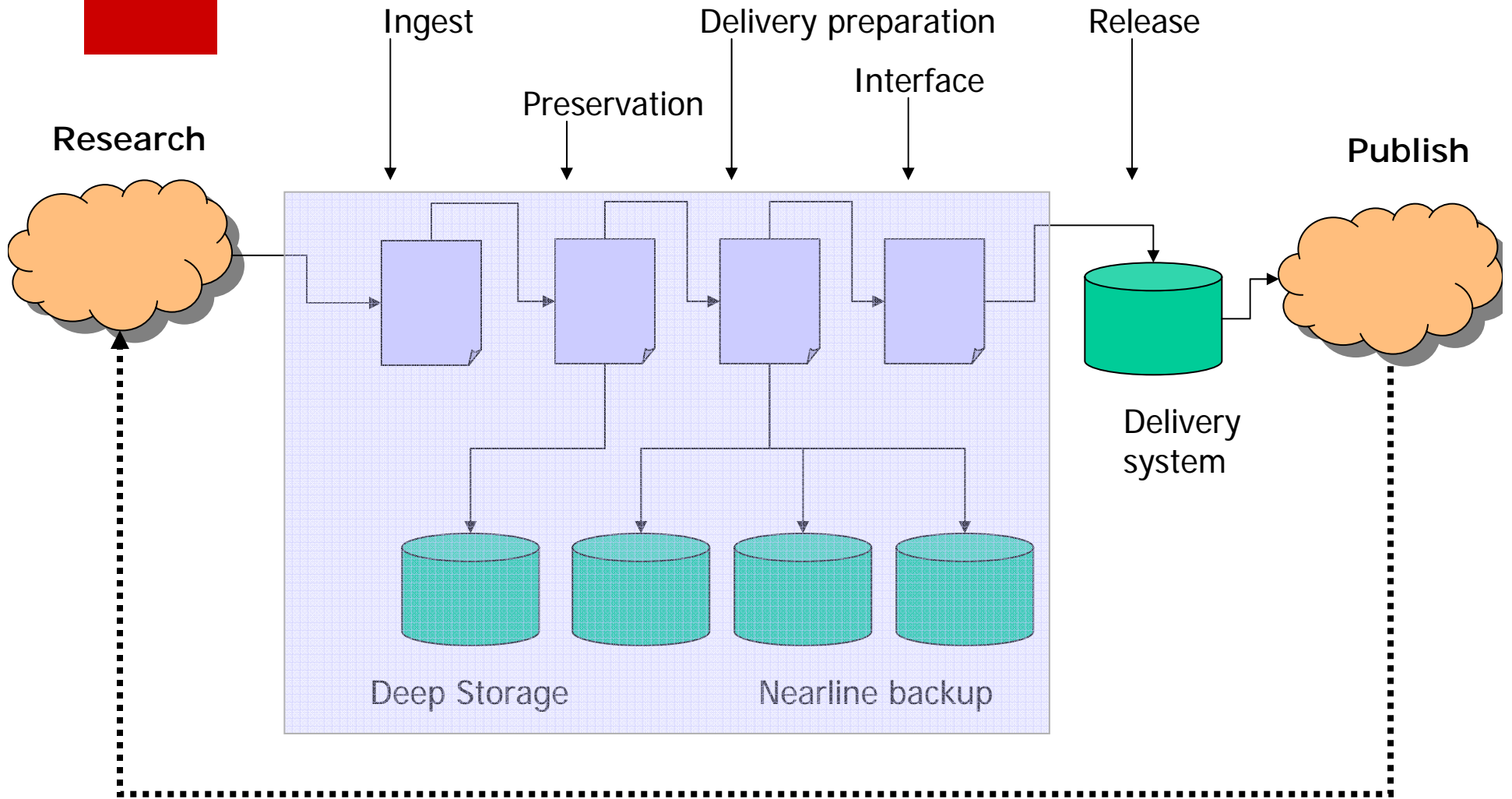
- Digital preservation
- Guidance for data creators
- Access to data
- Guidance for funders



Research

Publish







SIP: Submitting Information

Community interaction

Agreeing File Formats and Metadata

Negotiation on interface and constituents

Deposit licence

Delivery and receipt processes

Accession and Management



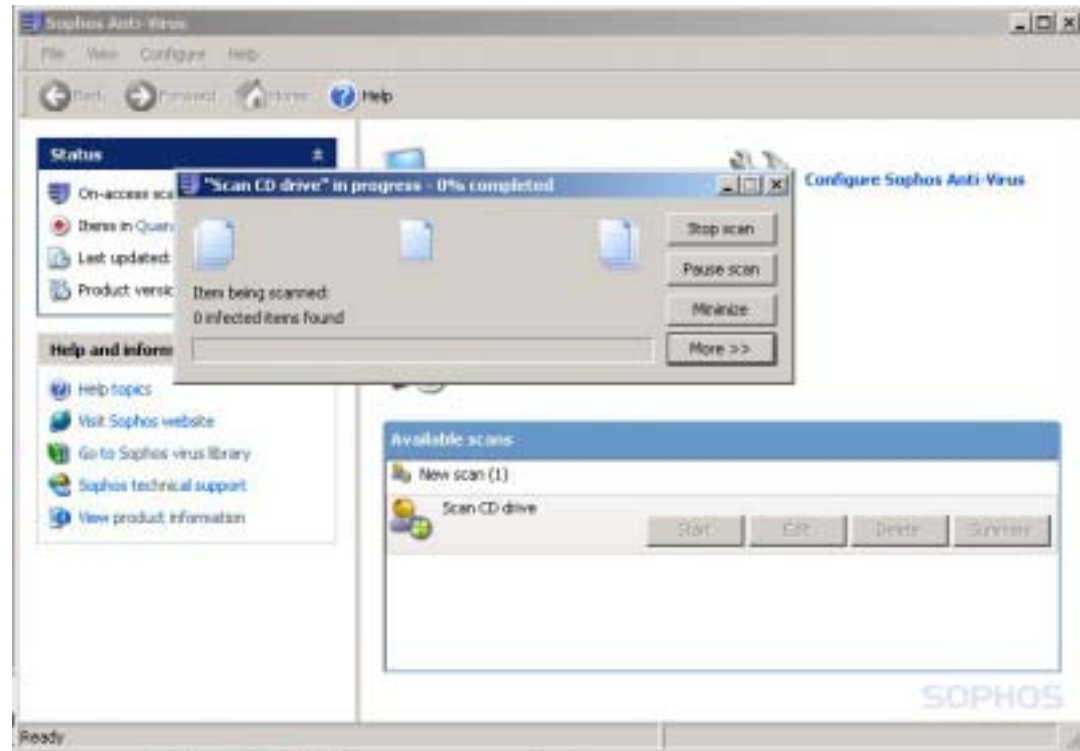
SIP to AIP – Ingest checklist

1. Check data for viruses
2. Can we read the CD/can we open the files?
3. Check file counts against documentation provided
4. Is documentation adequate for data provided?
5. Run checksum on CD
6. Copy data to server
7. Run another checksum and compare
8. Replace spaces in filenames with underscores
9. Scan licence agreement and store with SIP
10. Record receipt of data in management database
11. Create receipt to send to depositor



SIP to AIP – Ingest checklist

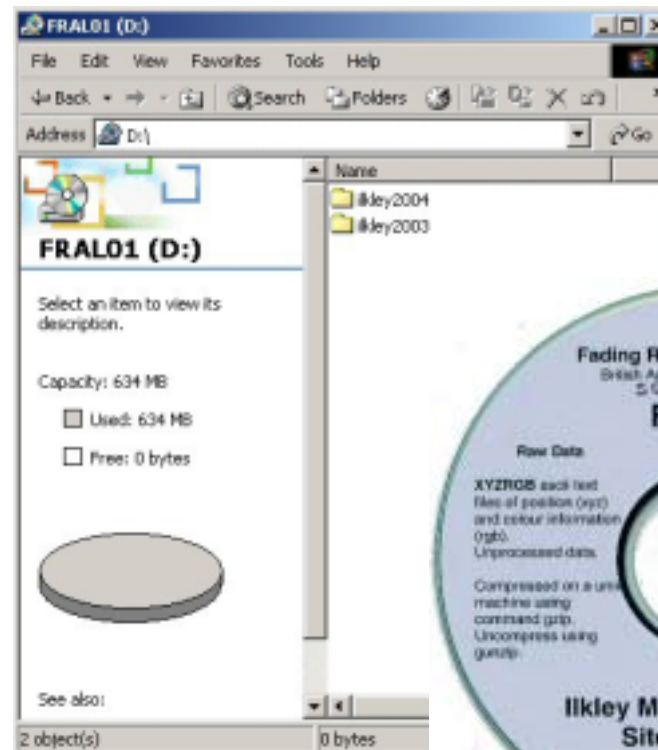
1. Check data for viruses





SIP to AIP – Ingest checklist

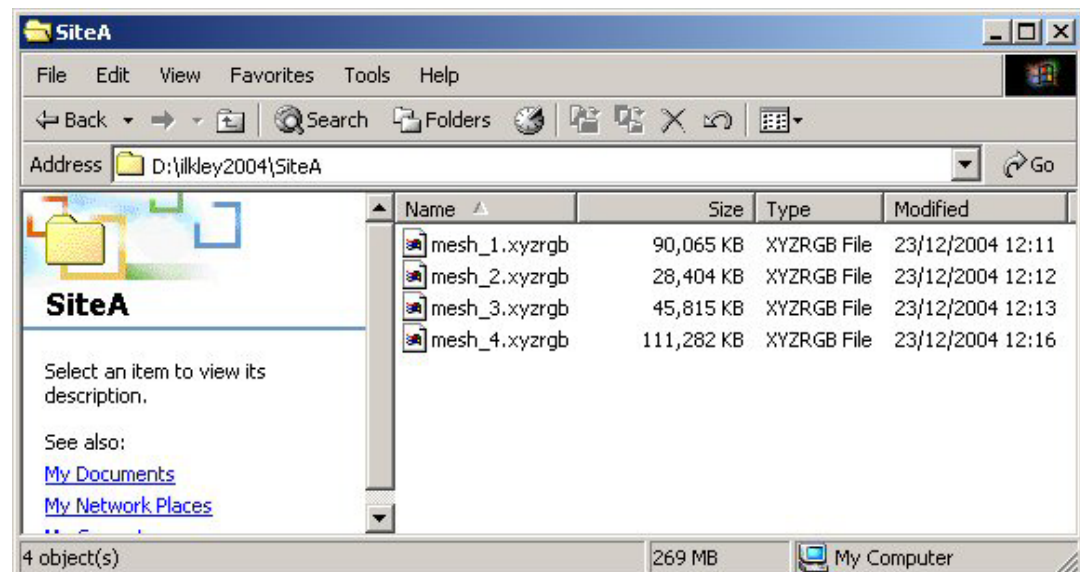
2. Can we read the CD/can we open the files?





SIP to AIP – Ingest checklist

3. Check file counts against documentation provided





SIP to AIP – Ingest checklist

4. Check documentation is adequate for the data provided



...a little bit more about documentation

Carrots:

- Doing the right thing
- Professionalism
- Reputation
- They may also be users of the archive
- Provide templates and guidance

Sticks:

- Undocumented data will make them look bad
- Funders may require them to archive data
- We can refuse to accept it



SIP to AIP – Ingest checklist

5. Run checksum on CD

```
Command Prompt
Microsoft Windows [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.

H:\>c:

C:\>FSUM "D:\\" C:\checksum.txt /R
MD5 Checksum calculation and verification ut
(C) 2003-2004 Kirill Zinov and Vitaly Rogoz

Calculating size of:
D:\

8x[.....]100%
*****

Calculation summary:
Processed 18 files in 6 folders with total size 634.56 Mb.
Elapsed time: 00:03:27 Average speed: 3.13 Mb\Sec.
Result written in C:\checksum.txt

C:\>_
```

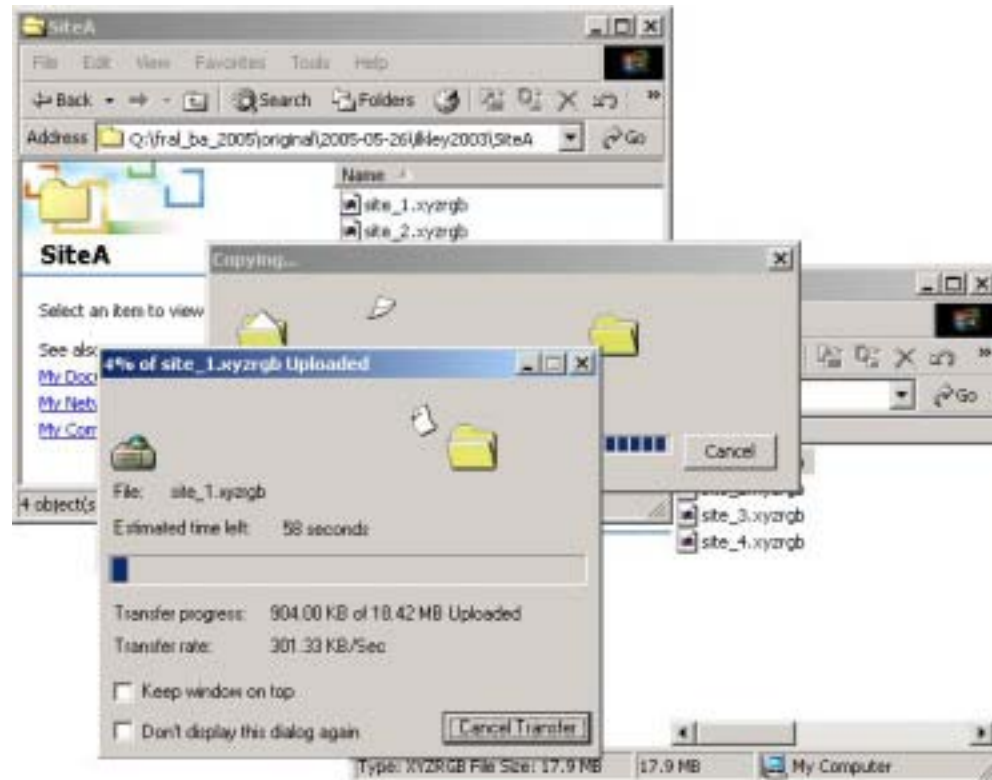
```
checksum.txt - Notepad
File Edit Format Help

: Generated by FastSum - an extremely fast checksum utility.
: Go to www.fastsum.com to get your own copy for free.
:
9F7ECF7ECBDA1A819CB2D400B824CD69 *ilkley2003\siteA\site_1.xyzrgb
CD9FE4E21E32D6C8313E580180C88ABB *ilkley2003\siteA\site_2.xyzrgb
CCCFCFB5E1B6083488A3067C89928011 *ilkley2003\siteA\site_3.xyzrgb
EE89727738A6C0443D9E5839685076A2 *ilkley2003\siteA\site_4.xyzrgb
8DA1A86DBD0374A8FF03C3DF0D9B5AB0 *ilkley2003\siteB\site_B_Trimmed.xyzrgb.gz
3B860E41ECEB4BD9A9762414631C7ED7 *ilkley2004\siteA\mesh_1.xyzrgb
E10E5A8B5169E64BA5C0081C0A8DAA30 *ilkley2004\siteA\mesh_2.xyzrgb
A1EED094121A9E680C6CA839C9799BC5 *ilkley2004\siteA\mesh_3.xyzrgb
A1A0EA4923CB97460167AD65D73A8386 *ilkley2004\siteA\mesh_4.xyzrgb
27D4A416A714F23C52D1FB6B5B455604 *ilkley2004\siteB\Mesh.xyzrgb.gz
```



SIP to AIP – Ingest checklist

6. Copy data to server





SIP to AIP – Ingest checklist

7. Run another checksum and compare

```
checksum.txt - Notepad
File Edit Format Help
: Generated by FastSum - an
: Go to www.fastsum.com to
:
9F7ECF7ECBDA1A819CB2D400B824
CD9FE4E21E32D6C8313E580180C8
CCCFCFB5E1B6083488A3067C8992
EE89727738A6C0443D9E58396850
8DA1A86DBD0374A8FF03C3DF0D9E
3B860E41ECEB4BD9A9762414631C
E10E5A8B5169E64BA5C0081C0A8DAA30
A1EED094121A9E680C6CA839C9799BC5
A1A0EA4923CB97460167AD65D73A8386
27D4A416A714F23C52D1FB6B5B455604
```

checksum_Q.txt - Notepad

```
File Edit Format Help
: Generated by FastSum - an extremely fast checksum utility.
: Go to www.fastsum.com to get your own copy for free.
:
8DA1A86DBD0374A8FF03C3DF0D9B5AB0 *ilkley2003\siteB\site_B_Trimmed.xyzrgb.gz
site_1.xyzrgb
site_2.xyzrgb
site_3.xyzrgb
site_4.xyzrgb
Mesh.xyzrgb.gz
mesh_2.xyzrgb
mesh_3.xyzrgb
mesh_4.xyzrgb
mesh_1.xyzrgb
```

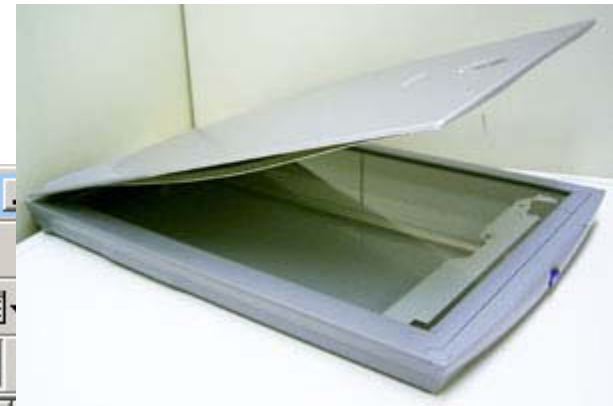
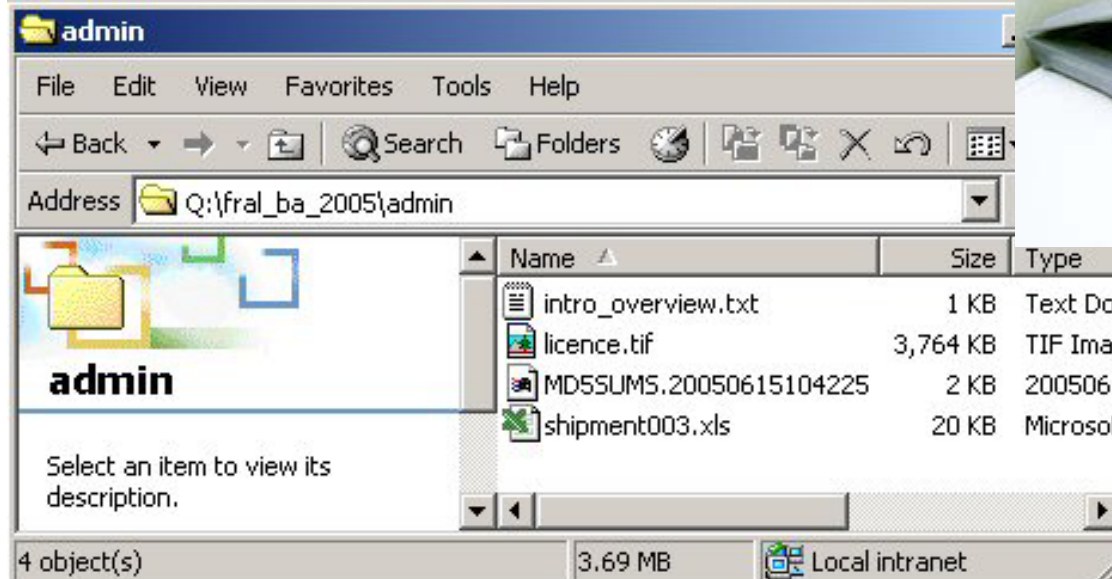
Copyright ©1998 - 2004
Salty Brine Software
all rights reserved.

www.FolderMatch.com



SIP to AIP – Ingest checklist

9. Scan licence agreement and store with data





SIP to AIP – Ingest checklist

10. Record receipt of dataset in management database

AHDS Collections management system.

[About](#) [People](#) [Tracking](#) [Accessions](#) [User Services](#) [Admin](#) [Assist](#) [Logout](#)

Accession details: Fading rock-art landscapes: evaluating the database in rock-art landscape research

Associated collection	1000431	view	Delivery media	
Associated tracking record	1002040	view	CD-ROM	1
Centre	AHDS Archaeology		Original files	
Checksum generated?	Yes		XYZ (xyz)	10
Type of Checksum	MD5		Added	30-Mar-2006 (afa2)
Virus checked?	Yes			
Date AIP sent to Exec	Not sent			
GPMD metadata generated?	No			
GMF2 metadata generated?	No			
Identifiers				
AIPID (arch)	Arch-431-1			
CollID (arch)	431			
Accessioned	26-May-2006 (afa2)			
Documentation notes				
No documentation notes listed				



SIP to AIP – Ingest checklist

11. Create receipt to send to depositor

Dear _____,

Fading Rock Art Landscapes

Thank you for the deposit of your digital data sets in respect of the above project received 26th May 2005.

We confirm that the data received consists of:

1 CD-ROM containing 10 files

with the following data files:

10 xyz files

Should you have any queries concerning the data or its accessioning please do not hesitate to contact us.

Yours sincerely,

Jenny Mitcham
Curatorial Officer
Archaeology Data Service



Assembly review +
Management function

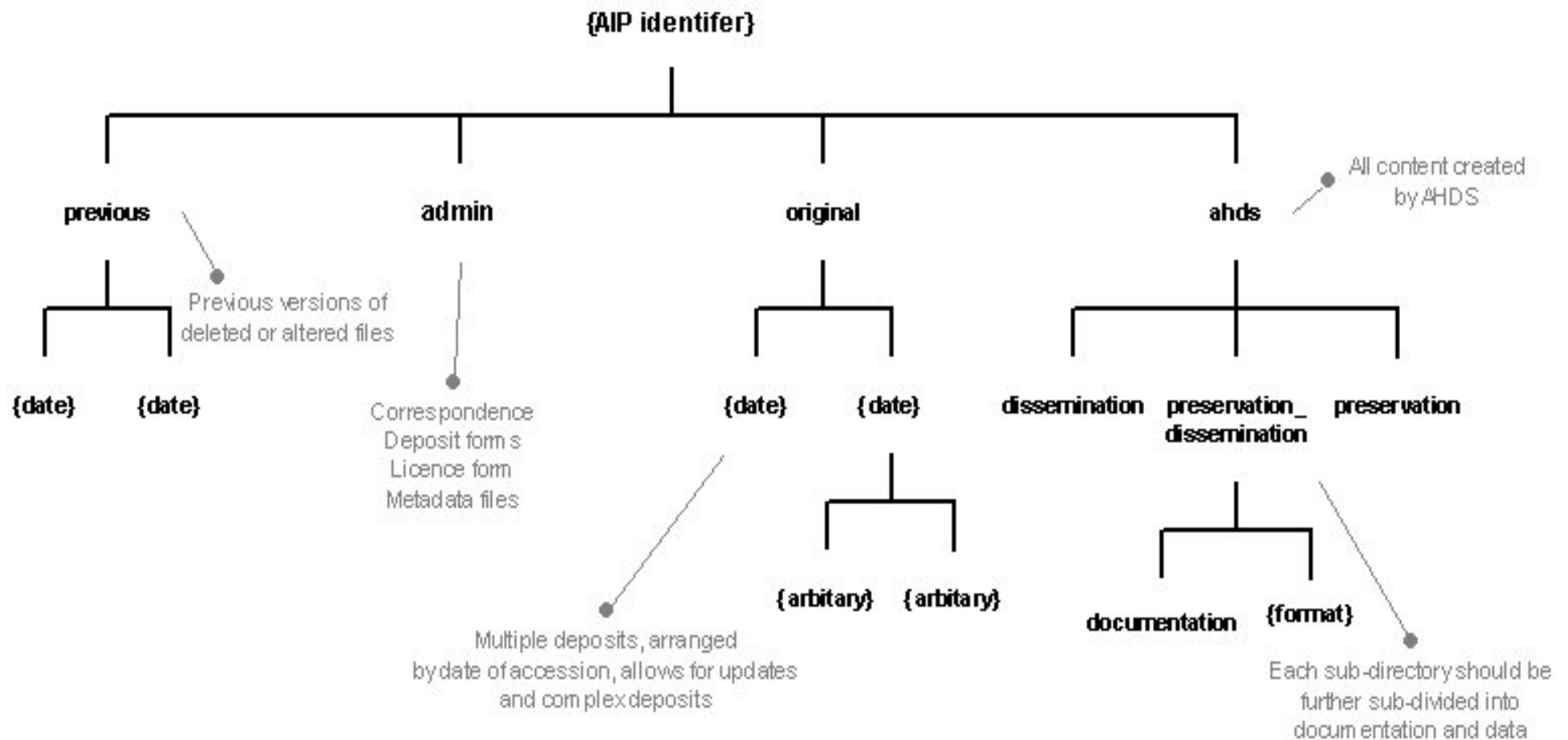
AIP: assembly

Assemble parts of AIP:

- Original submission
- Admin documents and metadata
- Preservation copies of data
- Dissemination copies of files
- Transformation metadata



AIP directory structure





Procedural review +
Management function

AIP: Procedures

What goes where

Format and transformation manuals


Documenting transformations

Administrative information



AHDS Preservation Manuals

<http://ahds.ac.uk/preservation/ahds-preservation-documents.htm>

**arts and humanities data service**

AHDS Preservation

AHDS Preservation Policies

General Preservation Reports

Preservation Bibliography

Depositing Data

Related Projects

Printer friendly version
Text version

AHDS Preservation
c/o AHDS Executive
26 - 29 Drury Lane
3rd Floor
LONDON, WC2B 5RL
Tel: 0207 848 1988
Fax: 0207 848 1989
Email the AHDS

AHDS Repository Policies and Procedures

Policies and procedures that guide the operation of the AHDS digital repository

- AHDS Preservation Policy [PDF version \(58kb\)](#) | [RTF version \(138kb\)](#)


The AHDS series of **Preservation Handbooks** identify significant properties to preserve digital objects, in accordance with "a migration based preservation strategy." While created with this internal purpose in mind, they may well be of use to other digital repositories.

Nine are in place. The rest will be published throughout Autumn 2005.

Audio	PDF version (55kb)	RTF version (138kb)
Moving images	PDF version (58kb)	RTF version (138kb)
Plain Text		
Digital By-products	PDF version (22kb)	RTF version (138kb)
Marked-Up Text		
Binary Text / Word Processor Documents		
Relational Databases		
Statistical Data Files	PDF version (92kb)	RTF version (138kb)

AHDS Preservation Handbook

Arts and Humanities Data Service
<http://ahds.ac.uk>



Preservation Handbook

Digital Audio

Author	Gareth Knight & John McHugh
Version	1
Date	26 July 2005
Change History	



DIP: Preparing for Dissemination / Access

Service review + helpdesk +
Management function

Community interaction

Design and implementation

Internal procedures and checklists

Quality Checks   

Depositor review / peer review

Release



DIP: Access

<http://ads.ahds.ac.uk/catalogue/>

http://ads.ahds.ac.uk/catalogue/resources.html?spitalfields_var_2001

Database	Codes	Notes
Burial Catalogue	Codes (340)	
Slaves Catalogue	Codes (340)	
Epiphany Register	Key (340)	340
Marriages Register	Notes (340)	340



...then continuous active management

