Introduction to Digital Preservation

TNA Training School

13th November 2024

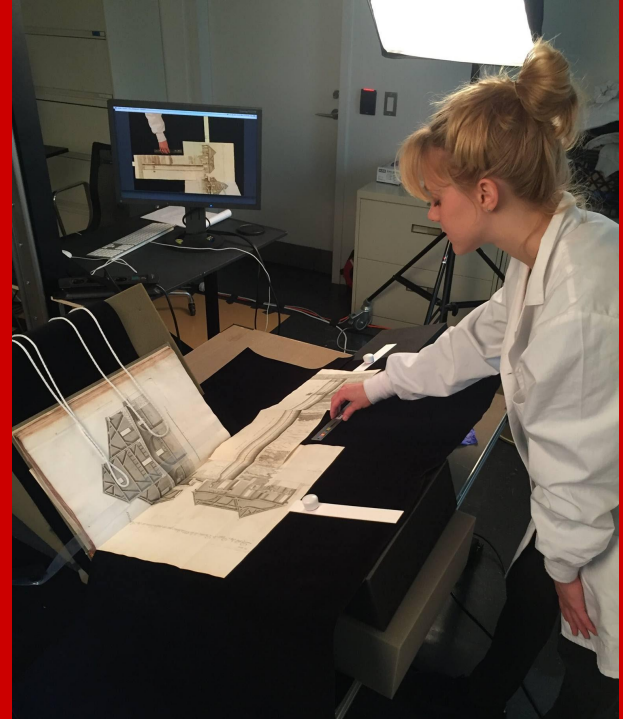*Olivia Foster*

Introduction to Digital Preservation

- Why is digital data at risk?
- What is digital preservation?
  - Digital preservation activities
  - Digital preservation strategies
- A case study of the loss of digital data

https://doi.org/10.5281/zenodo.11147887

Risks to physical archival material

- Loss of records

- Physical and chemical deterioration

- Dissociation of material and/or inadequate documentation

Is digitisation the solution?

Why is Digital Data at Risk?

- Depends on technology to access
- Inadequate documentation
- Risk of data loss
- Obsolescence of:
    - Media
    - Hardware
    - Software

https://angrysysops.com/2022/02/25/do-not-erase/

shutterstock.com · 1130375453

Obsolesce of Storage Media, Software and Formats

- Technology changes over time, creating barriers to access and reuse.

- Accessibility of software or file formats often require documentation to use and understand them.

https://www.123rf.com/photo_91808895_illustration-of-various-electric-appliances-storage-media.html

ads **Archaeology Data Service**

If it is adequately preserved, there are many benefits to digital data, in particular the ability to make content more accessible to users.

https://digitalbevaring.dk/

ads Archaeology Data Service

https://digitalbevaring.dk/
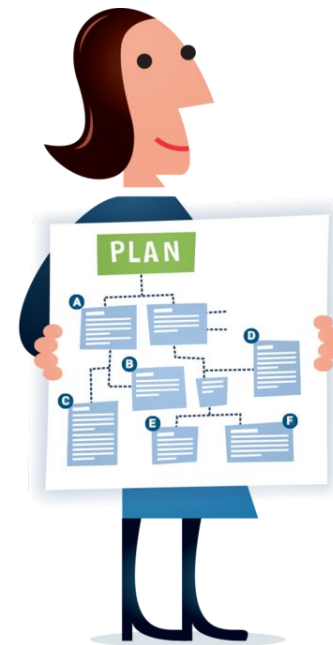
Digitisation ≠ **Digital Preservation**

Data creation / acquisition

Digital Preservation Policy: **Selection and retention**

- Not everything may need to be preserved
- Copyright, sensitive data and GDPR
- Financial and environmental cost of storing digital data

Digitisation
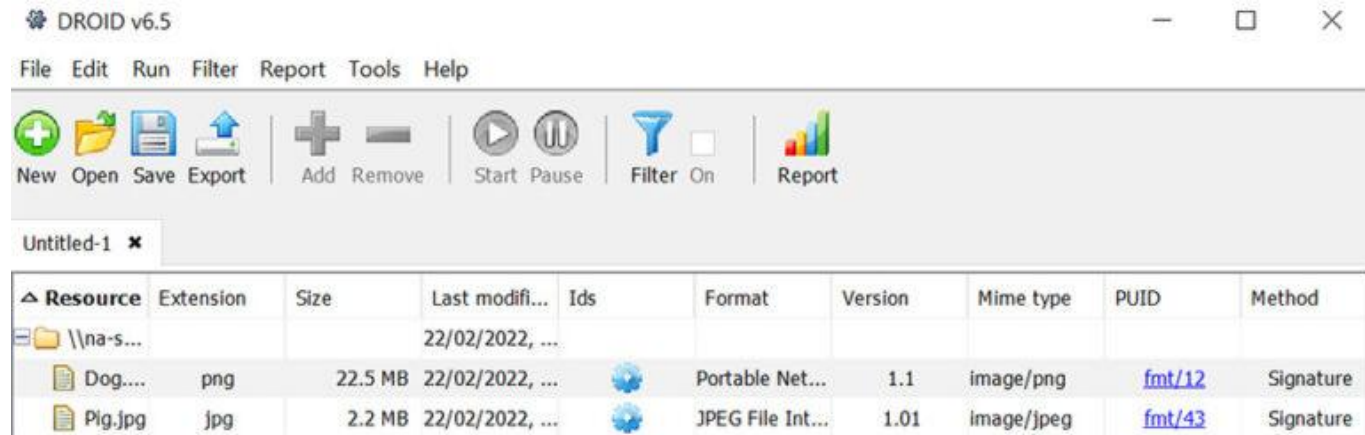
Digital data

Digital preservation activities

Digital Preservation Activities: Appraisal

DROID (Digital Record Object Identification) - a software tool developed by The National Archives to perform automated batch identification of file formats

Digital Preservation Activities: Preservation actions

- **Replication** of files (creating back-ups)
- **Normalisation or Migration** of files to different formats
- **Emulation** – re-creating the original operating environment.

Preserving content vs preserving look and feel of a resource

https://megadrive-emulator.com/platform-game/sonic-the-hedgehog

Digital Preservation Activities:
Creating documentation

- File level metadata

- Collection level ('discovery') metadata

- Using controlled vocabularies (e.g Wikidata)

- Documenting preservation actions

https://www.wikidata.org/wiki/Q437356

Object Id	Caption	Object Representation	Subject Keyword	Period	Period Dates	Creator	Location	Location Coordinates / Extent	Creation Date
2282009	North-east facing shot of deposits (5) and (3) (1m scale)	USH22_282.jpg USH22_282.jpg USH22_282.tif	Watching Brief Archaeological Investigation Deposit	Other: 19th Century Other: 20th Century Other: 21st Century	1800, 2022	Archaeological Research Services Ltd	Other: 76 Union Street Other: Harthill	Type: OSGB, Easting: 449381, Northing: 380700	10-Oct-2022

Digital Preservation Activities: Publishing data

- Ensuring data is findable on our website

- Providing data to download in an accessible way

- Creating persistent identifiers (DOI: Digital Object Identifier)

The British Institute for Libyan and Northern African Studies (BILNAS) (2024) *Digital Archive from Sabratha Excavations: The Kathleen Kenyon and John Ward-Perkins Collection, 1948-1951* [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1122174
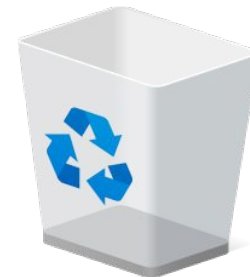
Digital Preservation Activities: Data disposal?

- Retention periods: Does everything need to be preserved forever?

- Sensitive data and GDPR

- Embargoed or restricted access to datasets

Council for British Research in the Levant Research Reports
Council for British Research in the Levant, 2019. (updated 2022) https://doi.org/10.5284/1053684. **How to cite using this DOI**

Introduction
Metadata
Usage Statistics

Data copyright © Council for British Research in the Levant unless otherwise stated

This work is licensed under a Creative Commons Attribution 4.0 International License.

This is a landing page for doi:10.5284/1053684 (ADS archive arch-3514-1)

This dataset has been deaccessioned at the request of the depositor. The ADS has no files to download for this collection

For more information on CBRL publications, please visit the Council for British Research in the Levant Website.

Digital preservation is an **active process** of data management, appraisal and curation.
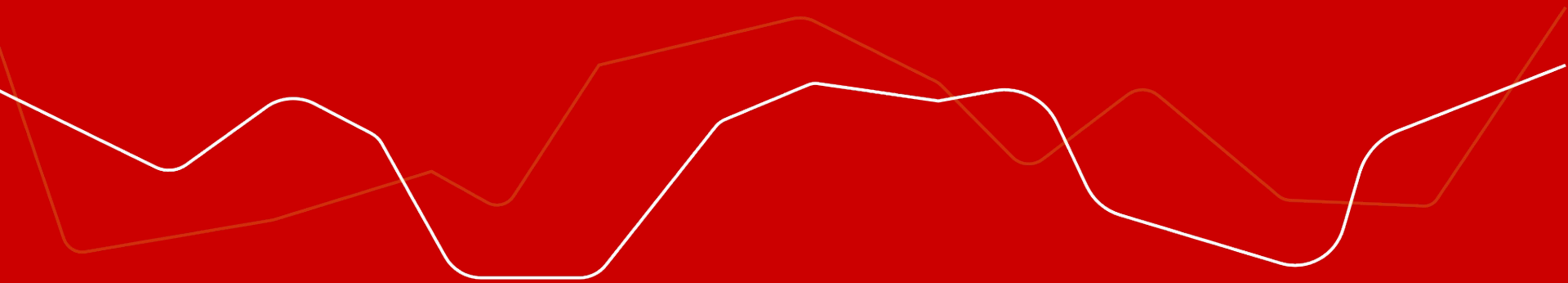
Digital Preservation Models

Three Legged Stool:

- ■ Technology (tools, security, storage and Back-Up

- ■ Organisation (policy, strategy, procedures, staffing)

- ■ Resources (business planning, cost modelling and funding)
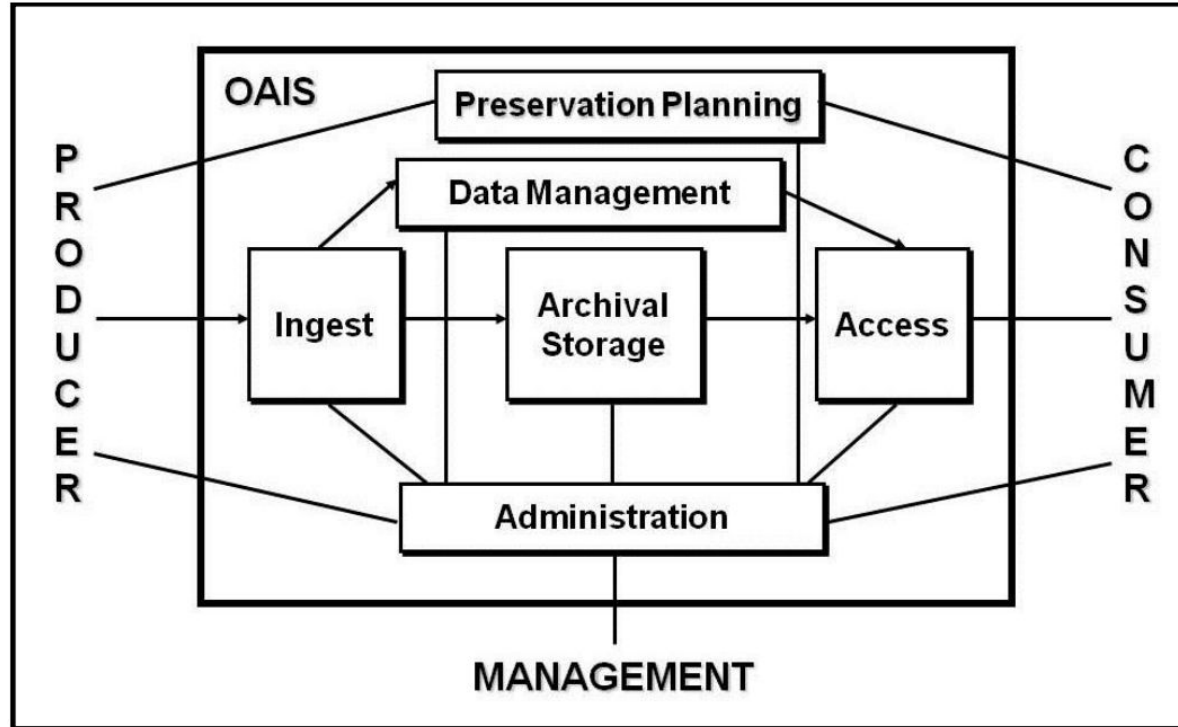
Open Archival Information System (OAIS) Model

ads Archaeology Data Service

Data creation, e.g. digitisation

CONCEPTUALISE

DISPOSE

TRANSFORM

CREATE OR RECEIVE

CURATE

ACCESS, USE & REUSE

PRESERVATION PLANNING

DESCRIPTION

Data (Digital Objects or Databases)

REPRESENTATION INFORMATION

and

and

APPRAISE & SELECT

MIGRATE

COMMUNITY WATCH & PARTICIPATION

STORE

PRESERVE

INGEST

REAPPRAISE

PRESERVATION ACTION
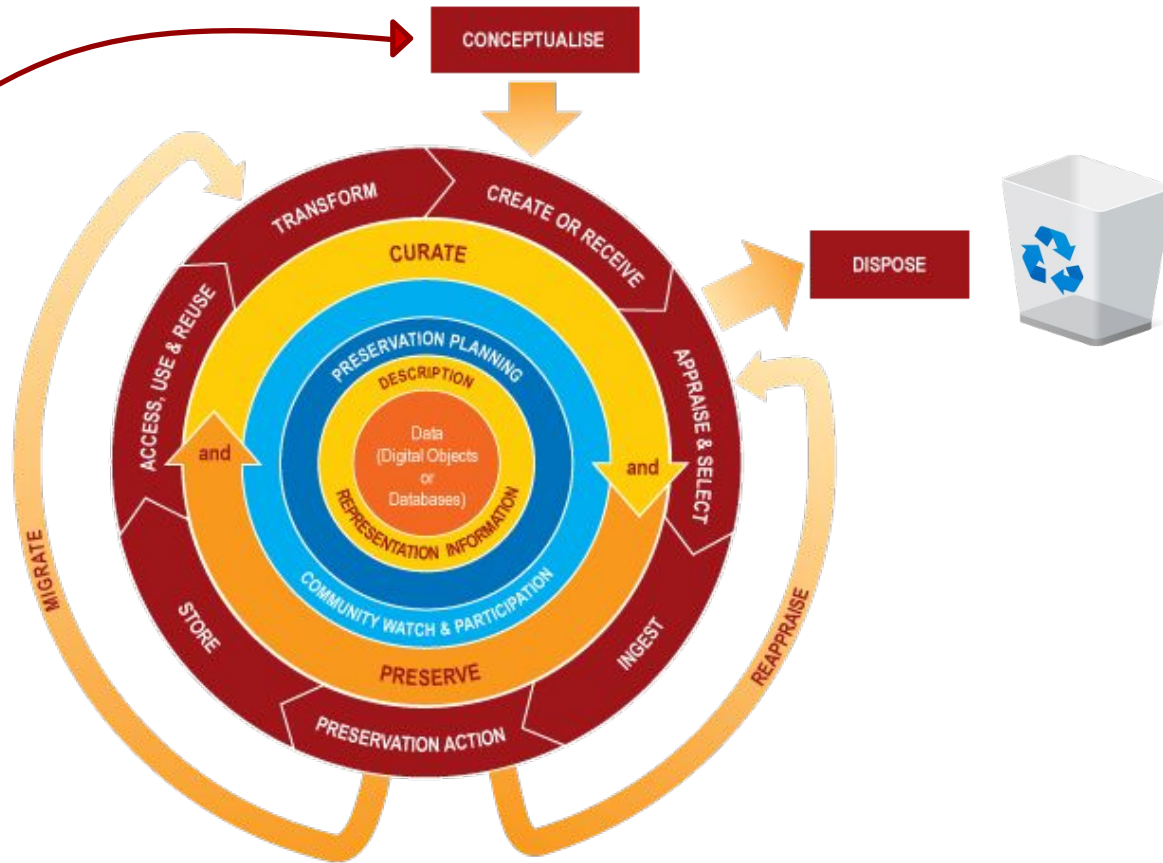
The DCC Curation Lifecycle Model

The Newham Archive: A Case Study of the Loss of Digital Data

The Archaeological Service was closed down in 1998 and the digital archive was passed to the ADS. The digital archive represented all the work that was digitised during Newham Archaeological Service fieldwork and post-excavation analysis, along with project designs over a period of about ten years.

This archive was delivered to the ADS on 230 floppy disks containing over 6000 files and totalling over 130Mb of data.

The Newham Archive: A Case Study of the Loss of Digital Data

Issues with data:

■ Data in proprietary file formats that have gone out of use

■ Non-existent data or project documentation.

Resulting in:

■ 10-15% of the files are inaccessible and the data the floppy disks contain are effectively lost

■ A number of 'orphaned' datasets, including a large cemetery database, which have been rescued but have little reuse potential.

Although the data was considered 'archived', much of it was inaccessible

Resources:

- Introduction to digital preservation handbook by the Digital Preservation Coalition (DPC)- https://www.dpconline.org/handbook/introduction

- Archaeology Data Service / Digital Antiquity Guides to Good Practice - https://guides.archaeologydataservice.ac.uk/g2gpwiki/

- National Archives Blog - PRONOM: A database centenary (an overview of PRONOM, DROID and file format identification) - https://blog.nationalarchives.gov.uk/pronom-a-database-centenary/

- DROID (Digital Record Object Identification) tool developed by The National Archives - https://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/

Any questions?

Archaeology Data Service
Department of Archaeology
University of York
The King's Manor
Exhibition Square
York, YO1 7EP

www.archaeologydataservice.ac.uk

help@archaeologydataservice.ac.uk