



Archaeology
Data Service

Migration and Normalisation of Files

Kieron Niven, Digital Archivist: Data Standards
22nd November 2022

**ADS follows a Preservation by
Normalisation & Migration
Approach.**

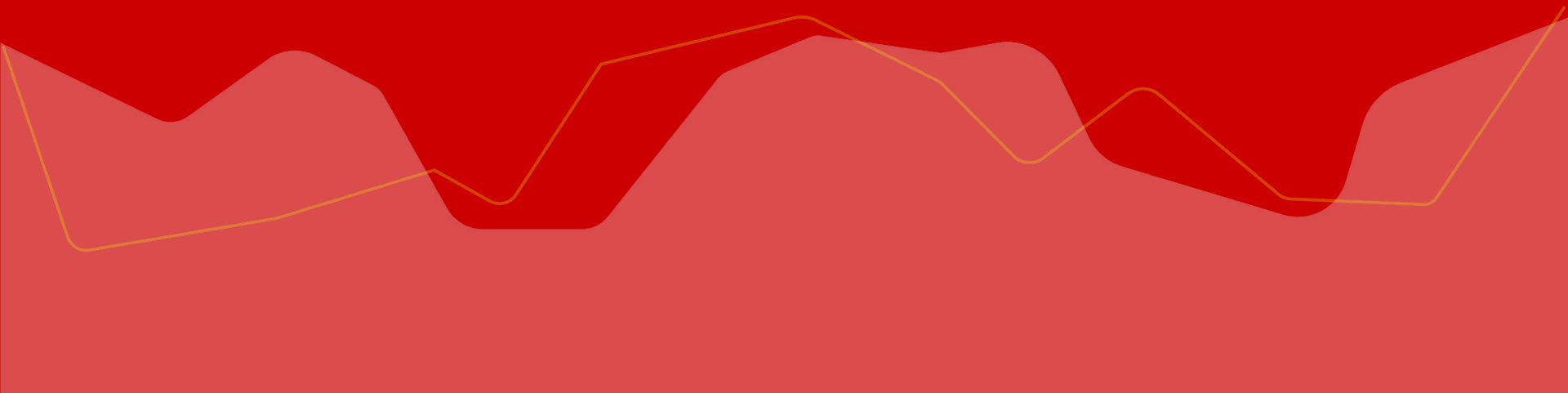
Other approaches are available...

Other Approaches

- Technology Preservation
- Refreshment
- Emulation

These other approaches all aim to preserve the **original bit stream** and maintain the **surrounding infrastructure** whereas file migration and normalisation focus on maintaining **content**.

Definitions



Migration

- *“The process of moving data from one information system or storage medium to another to ensure continued access to the information as the system or medium becomes obsolete or degrades over time.”*

Society of American Archivists 'Dictionary of Archives Terminology'.
<https://dictionary.archivists.org/>

Migration

Can include:

- migration through formats (e.g. rtf to docx)
- media migration (e.g. floppy disc to CD to DVD)

Key element is **movement** to maintain access.

Normalisation

- *"1. the process of converting a digital object into a persistent file format"*
- *2. transforming data to a canonical form"*

Society of American Archivists 'Dictionary of Archives Terminology'.
<https://dictionary.archivists.org/>

Normalisation

Can be viewed as a form of migration focused on data type (e.g. images) moving to a single, ideally stable, format.

**The ADS approach primarily focuses on
Normalisation**

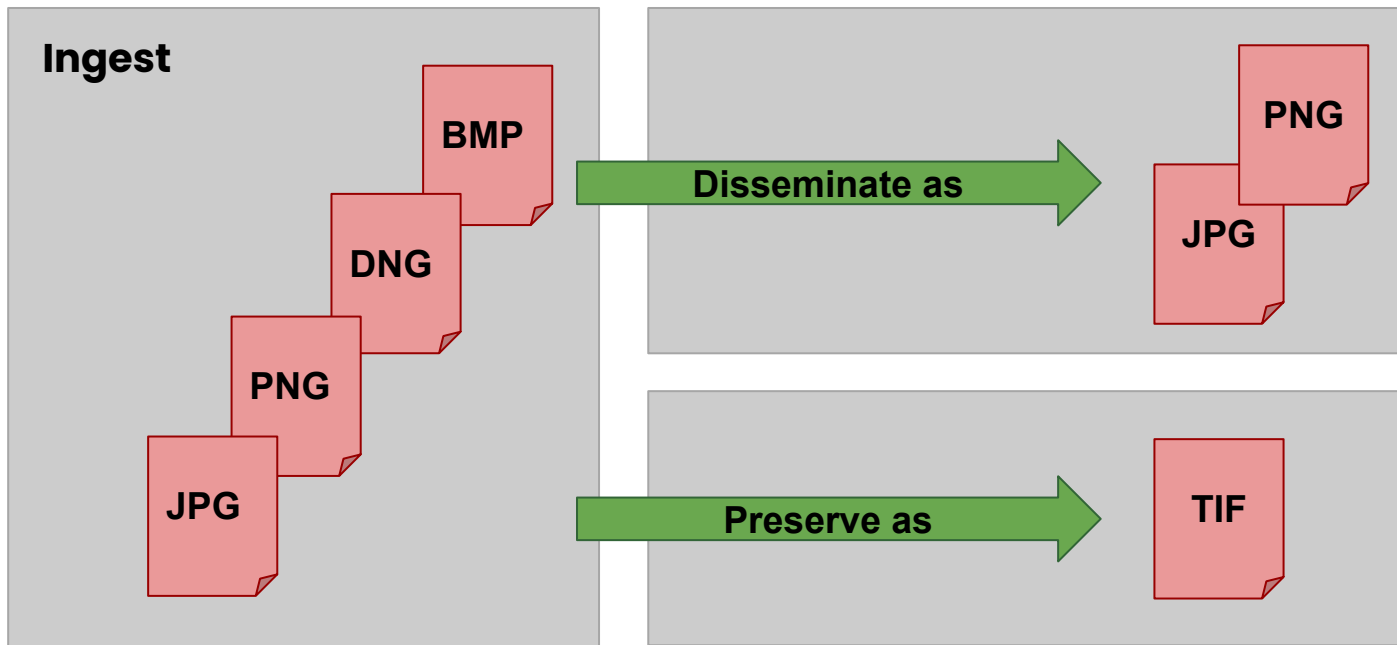
**ADS Preservation Policy (s. 5.2 Method of
Preservation) defines these terms
specifically in relation to ADS**

https://archaeologydataservice.ac.uk/attach/ADS_preservation_policy_v1-8.pdf

Normalisation

“Within the ADS, all digital objects of a particular type (e.g. raster images) are converted into a single chosen file format that is thought to embody the best overall compromise amongst characteristics such as functionality, longevity, and preservability. The formats chosen are set within the individual ADS *Data Procedures*.”

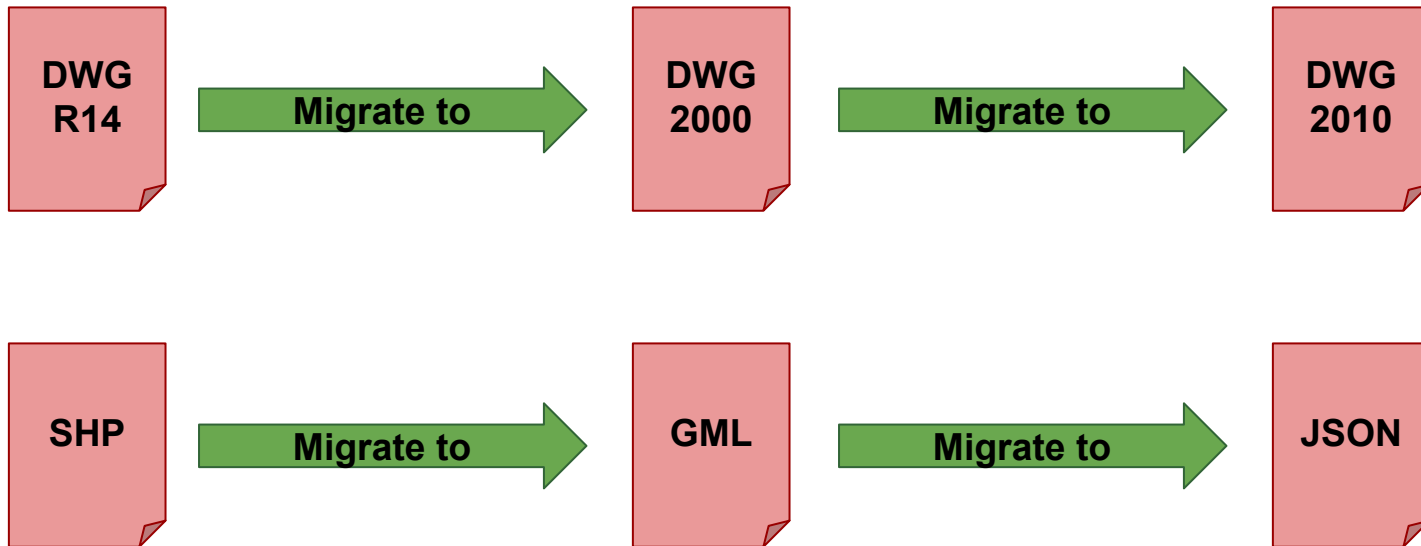
Normalisation



Migration

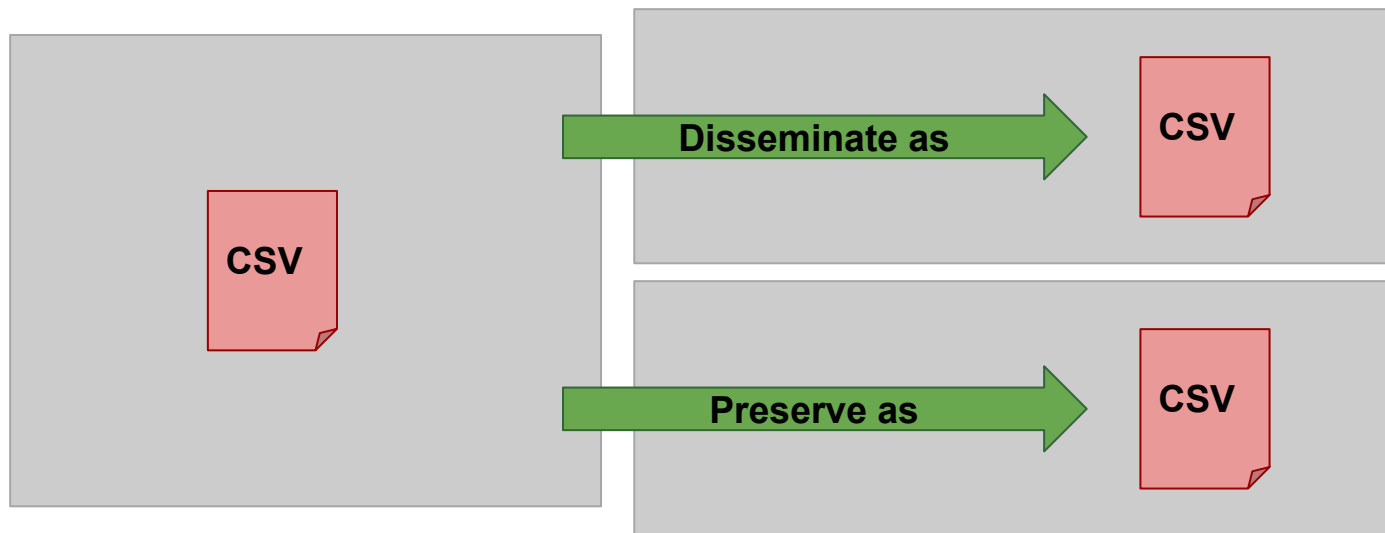
“As well as normalisation data may be migrated to other formats as part of longer-term, and repository wide, reviews of data types and standards. Data may be migrated through successive versions of a format. Version migration may be the only option for preserving proprietary formats that do not migrate to open standards. This is only practical where the software using proprietary formats is widely used within a community and accessible (affordable) to an archive.”

Migration



Replication

“migration between **media** that leaves data (the bit stream) totally unchanged.”



In an ideal world:

We'd only need to do **Replication**

- Ingest formats that are already suitable for preservation and dissemination.
- Require no further work for preserving or disseminating.

We aim at Normalisation with Migration as a 'fall back' for certain formats.

In reality:

Archaeologists work with a **range of formats**, many of which are:

- **Proprietary**
- use **compression**
- or can only be accessed using specialist software.

The basis of a migration/normalisation strategy is:

- to assess what formats are being created by your designated community.
- Look at how suitable these formats are for **preservation** and **reuse**.
- Where necessary, identify pathways between these formats and preservation formats.
- Assess your ability to support these formats: to assess and migrate (software, expertise).

Developing a migration/normalisation strategy:

- You need to find a balance between being able to support the data being deposited while also making preservation practical (costs, staffing, etc.).
- Avoid Worst-case Scenario – you have to accept (and support) all formats!
- Look to other repositories for the types of information and formats they accept (likely to be similar by subject area and data type).

- Define your ingest formats – generally the fewer the better and based on pathways to preservation formats
- Look at general DP guidelines on what makes a good sustainable format (DPC, LOC).
- Look at data type specific guidelines and what is involved in preserving specific formats (complexity, long-term cost of software etc.)
- Look at what is required by the community your repository is supporting (de facto standards e.g. AutoCAD)
- Expertise – do you have the skills to assess and migrate the data? You may want depositors to normalise formats (limit ingest)

ADS Instructions for Depositors

- An example of what formats ADS ingest.
- Each of these files has a pathway to pres. and diss versions defined in our Data Procedures.

<https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

Significant Properties

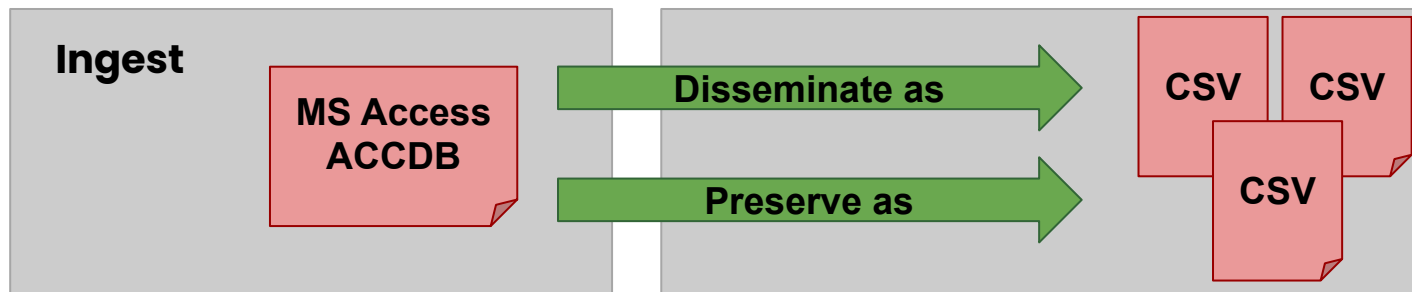
- In addition to the nature of formats, a key consideration is the preservation of **Significant Properties** when migrating files
- Important aspects of files that should be maintained regardless of the format
 - e.g. documents (paging, layout); images (colour, resolution)
- Highlighted in most guidance documents

Metadata

- Metadata also plays a key role in data migration and normalisation.
- It is important to know what processes have been undertaken on files.
- We record:
 - What the process is, when, by who
 - The software and OS used
 - Any issues encountered
- PREMIS metadata standard (Object, Environment, Event, Agent). <https://www.loc.gov/standards/premis/v3/index.html>

Normalisation can be complex

- Sometime normalisation is more complex than a 1:1 process e.g. databases, spreadsheets
- Metadata records the relationships between the original files and their archival versions



Aim for 'once' but plan for 'many'

- An ongoing process – includes **technology watch**, ongoing monitoring and assessment of formats and software
- **Preservation Policy: 5.7. Standards and preservation watch:**
"It may become necessary for a number of reasons including: Version change (many formats change or evolve over time); Format obsolescence (a format is or is becoming deprecated); Another format becomes a more attractive preservation option"
- Following widely accepted standards allows you to benefit from the expertise of larger organisations

Resources:

Guides to Good Practice:

<https://guides.archaeologydataservice.ac.uk/>



Sustainability of Digital Formats: Planning for Library of Congress Collections

<https://www.loc.gov/preservation/digital/formats/index.html>

DPC Technology Watch Reports and Guidance Notes

<https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>



Archaeology
Data Service

Migration and Normalisation of Files

kieron.niven@york.ac.uk



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