

Preservation planning: Normalizing file formats

Defining normalization

- What is it?
 - Normalization means converting ingested objects into a small number of pre-selected formats
- Why do it?
 - Some formats are easier to preserve than others
 - A smaller number of formats means fewer preservation actions required

Normalization vs. migration

- Migration is similar to normalization in that it involves converting ingested objects into preservation-friendly formats
- Unlike normalization, migration is typically done only when the format is at risk of obsolescence
- Migration as a strategy means adopting a wait and see approach

Disadvantages of normalization

- It requires more planning up front to implement
- Re-normalization may be required as better target formats or conversion tools become available

Advantages of normalization

- Taking preservation action on ingest helps define and manage risk
 - Adopting a wait and see approach means putting off an *undefined* amount of work for an *indefinite* period of time at an *unknown* cost
- Normalization does not preclude the future use of migration or other strategies such as emulation

Criteria for choosing formats

1. The format must be non-proprietary

- There must be no associated licenses or patents or the possibility of there being such licenses or patents in the future

Criteria for choosing formats

2. There must be freely available specifications

- A specification is a document that explains exactly how the format is structured and rendered
- This specification must be freely available to all and not subject to copyright or other restrictions

Criteria for choosing formats

3. The format should be widely endorsed and/or adopted

- Other established repositories should be using or have endorsed the format
- Formats that have been approved as international standards are particularly desirable

Criteria for choosing formats

4. For images and audio files there should be no compression
5. For video files any compression should be lossless

Criteria for choosing formats

6. There should be writing and rendering tools available for the format
 - Idealized standards must be matched by practical tools
 - The tools must reliably meet the requirements of the format specifications and must produce normalized objects that are faithful representations of the original objects

Choosing formats

- The Archivemata approach: develop media type preservation plans
 - That is, break the various formats into groups and develop normalization plans for each group
 - <http://www.archivemata.org>