



Digital Curation 101

MIGRATE

About *Migrate*

Topics:

- Migration
- Migration in practice
 - Application
 - Issues
- Migration changes data

Migration

Migrate is an occasional action in the curation lifecycle, usually as an outcome of decisions made at the *Preservation Action* stage. Its activities are:

- Migration of data to a different format.

Migration refers to transferring digital materials from one technology (hardware or software generation) to another, or from one format to another, preserving their intellectual content and retaining the ability for clients to retrieve, display, and otherwise use them.

Migration is one of two primary preservation methods (emulation is the other). Other preservation methods depend on migration to some extent. It is a well understood process, carried out frequently, for example when systems are upgraded. It is applied in response to changes in the storage environment or changes in the technology platform to counteract hardware or software obsolescence.

Migration in practice

Migration is different from refreshing, which is a simpler process of maintaining the bit stream by copying it to a new, usually more durable, media. Migration has several varieties, including:



Digital Curation 101

- simple version migration (for example, within one software product – for instance new versions of Microsoft Word work with earlier version of Word, but these are usually only limited to a few generations back)
- migration to newer or standard file formats, when obsolescence occurs or is imminent
- migration on request: migration occurs when access to digital material is requested.

A commonly-encountered example of migration on request is of seeking access to documents created in earlier software packages. Some software provides this ability. For example, the open source software suite OpenOffice¹ supports files created in the word-processing software Microsoft Word 6.0/95/97/2000/XP) (.doc and .dot), Microsoft Word 2003 XML (.xml), Microsoft WinWord 5 (.doc), StarWriter formats (.sdw, .sgl, and .vor), AportisDoc (Palm) (.pdb), Pocket Word (.psw), WordPerfect Document (.wpd), WPS 2000/Office 1.0 (.wps), DocBook (.xml) and others. It also provides similar capabilities for other office suite software. Read more in a posting by Chris Rusbridge on the Digital Curation Blog².

Migration has a long history and is the principal preservation technique applied to date. Considerable expertise exists in migration of some kinds of materials and formats, particularly document and datasets.

Migration has well-established procedures and uses relatively simple technology. It is considered to be a reliable way to preserve the intellectual content of simple digital objects (for example, page-based documents).

Application

When migrating data the following should be followed:

- Select acceptable data structures or file formats: for example, use non-proprietary, well-documented data format standards (examples include ASCII/UNICODE and XML) to increase the chance of future recoverability
- Document migration procedures fully in the metadata
- Retain the original bitstream where possible
- Apply strict quality checking procedures.

Migration tools are available to assist the process.

¹ <http://www.openoffice.org/>

² <http://digitalcuration.blogspot.com/2008/03/migration-on-request-openoffice-as.html>



Digital Curation 101

Issues

These issues are acknowledged for migration:

- It is best applied to less complex digital objects and to collections of materials that are in uniform formats; it does not currently handle complex materials well
- It needs to occur on a regular basis throughout the life of the materials
- Because it needs to occur frequently, migration has significant ongoing costs attached to it
- Repeated migrations, each making small changes, may accumulate into major alterations to the data.

Migration changes data

Every migration changes data. Repeated migrations, each making small changes, may accumulate into major alterations to the data. The result is likely to be information loss, and/or changes in system behaviour. Chris Rusbridge describes a commonly-experienced migration result:

So if I migrate from those PowerPoint 4 files to today's PowerPoint, and then from today's to tomorrow's PowerPoint, and then from tomorrow's to the next great thing, I will introduce cumulative errors whose impact I will only be able to assess at some horribly cringe-making moment, like in the middle of a presentation using a host's machine. So the best way to do migration is to start from the original file and migrate to today's version. Always.

Because functionality is lost and integrity compromised as a result of migration, care must be taken to strict quality checking procedures, for example, to compare the original bitstream and the migrated bitstream.