DIGITAL CURATION & PRESERVATION GLOSSARY

The Basics

Digital curation is all about maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire life cycle. Digital curation builds upon the underlying concepts of digital preservation while emphasizing opportunities for added value and knowledge through annotation and continuing resource management.

The term *digital preservation* encompasses all of the activities, policies, strategies and actions required to ensure that the digital content designated for long-term preservation is maintained in usable formats, for as long as access to that content is needed or desired, and can be made available in meaningful ways to current and future users, for as long as necessary regardless of the challenges of media failure and technological change. Digital preservation goals include ensuring enduring usability, authenticity, discoverability, and accessibility of content over the very long term.

Best practices are procedures and guidelines that are widely accepted because experience and research has demonstrated that they are optimal and efficient means to produce a desired result.

A *workflow* consists of the tasks, procedural steps, organizations or people, information and tools needed for each step in a process.

Digital Content

Digital content refers to any item created, published or distributed in a digital form, including, but not limited to, text, data, sound recordings, photographs and images, motion pictures and software. Used interchangeably with *digital materials*.

Born-digital content has never had an analog form. *Born-digital* materials differ from analog documents, movies and photographs that were *digitized*; that is, scanned or converted to a digital format.

Digitization is the process of creating digital copies or "surrogates" by scanning or otherwise converting analog materials.

Electronic records are the records created digitally in the day-to-day business of an organization, such as word processing documents, emails, databases, or intranet web pages.

Bits and Bytes

A *bit* is the smallest unit of information that a computer can work with. Each bit is either a "1" or a "0".

Bit-level preservation is the basic level of preservation of a digital resource (literally, preservation of the bits forming a digital resource). Bit-level preservation may include maintaining onsite and offsite backup copies, virus checking, fixity checking, and periodic refreshment to new storage media.

A *byte* is a unit of digital information and measure of data volume, normally equivalent to eight bits.

Kilobyte (KB) = 1,000 bytes
Megabyte (MB) = 1,000 kilobytes
Gigabyte (GB) = 1,000 megabytes
Terabyte (TB) = 1,000 gigabytes

Files and File Formats

A *file format* is a standard way that information is encoded for storage in a computer file. It tells the computer how to display, print, process, and save the information. It is dictated by the program that created the file and the operating system under which it was created and stored. A file format is often indicated by a file name extension containing three or four letters, e.g. .tif, .pdf, .jpg.

Preservation master: Digital content targeted for preservation that is considered the master version of the intellectual content of a digital resource. Preservation masters generally do not undergo significant processing or editing. Preservation masters are often used to make other copies including reproduction and distribution copies.

Access copy: A copy made from a digital object that is intended for use, such as online display or transmission over email.

Backup copy: An additional copy of a digital asset made to protect against loss due to unintended destruction or corruption of the primary set of digital assets. The essential attribute of a backup copy is that the information it contains can be restored in the event that access to the master copy is lost.

More on Digital Preservation

Digital preservation is the maintenance and management of digital objects over time so that they can be accessed and utilized by future users. Because of the relatively short lifecycle of digital information, preservation is an ongoing process.

Digital provenance is information about the origin of a digital object and any changes that may have occurred over the course of its life cycle.

A *digital repository* is an organization or department responsible for the intake and maintenance of digital objects.

Content migration is the process of transferring content between storage types, formats, or computer systems.

A *dark archive* does not grant public access and only preserves the information it contains. Access to dark archive data is either limited to a few specific individuals or completely restricted to all.

Format or technology obsolescence occurs when a piece of software or hardware is no longer in wide use or available at all. This causes it to be difficult or impossible to use any files that depend on this software or hardware.

Media deterioration or degradation refers to the breakdown of an analog object that holds digital objects, potentially causing the objects on the media to no longer be retrievable.

OAIS is an acronym that stands for Open Archival Information System. The OAIS framework consists of an organization of people and systems who have accepted the responsibility to preserve information and make it available for a certain group of people. It does not offer a definitive guideline for how a digital repository should act or what it should do, but instead gives the digital preservation community a common language and outlook for talking about digital preservation.

Authenticity and Integrity

Authenticity tells us that the digital material is what it purports to be. In the case of born digital and digitized materials, it refers to the fact that whatever is being cited is the same as it was when it was first created, unless the accompanying metadata indicates any changes. Confidence in the authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made.

A *fixity check* is a method for ensuring the integrity and authenticity of a file and verifying it has not been altered or corrupted.

A *checksum* is a unique numerical signature derived from a file. Checksums are used in fixity checking in order to compare copies.

Metadata

Metadata is a Latin term meaning "information about information." In the digital realm, metadata is data that describes key information about digital objects (image files, text files, digital audio/video) and, when appropriate, the original objects they represent.

> *Administrative metadata:* Information needed to help manage the digital object, such as copyright and preservation information.
> *Descriptive metadata:* Information used to search for and locate an object such as title, author, subjects, keywords, and publisher.
> *Structural metadata:* Information on how the digital object is organized. This may include the page or chapter order of a book, its table of contents or indexes. Structural metadata is often used by software programs.
> *Technical metadata:* Information about aspects of the object related to its file format or the software used to create the file. This may include things like the scanning equipment used to create a digital object and the settings used to create or modify it.

PREMIS is an acronym that stands for Preservation Metadata: Implementation Strategies. PREMIS metadata structures and describes what sort of preservation actions have been done to a digital object.

XML stands for Extensible Markup Language. One of the most common ways used to represent metadata.

Sources

- Digital Preservation Coalition https://www.dpconline.org/handbook/glossary
- National Digital Stewardship Alliance http://ndsa.org/glossary/
- University of Michigan Libraries
  https://www.lib.umich.edu/preservation-and-conservation/digital-preservation/digital-preservation-glossary
- University of Minnesota Libraries https://www.lib.umn.edu/dp/glossary
- Yale University Libraries https://guides.library.yale.edu/digitalpreservation/glossary