Glossary

Below is a list of terms that recur in the minimum specifications for the content types covered by these guidelines. For a general glossary of terms please refer to the Federal Agencies Digitization Guidelines Initiative (FADGI) Glossary available at: http://www.digitizationguidelines.gov/glossary.php [1]

ACCESS (copy) is also known as the User/Patron format.

BLACKLIGHT

BORN-DIGITAL are assets that originated in digital form, such as Web sites, wikis, e-books, digital sound recordings, and email.

CCITT Group IV is an image compression schema based on the "Comité Consultatif International Téléphonique et Télégraphique"), a telecommunications standard created in 1956

CHECKSUM is a function used for validating data integrity. Also referred to as MD5 (*Message-Digest algorithm 5).* An algorithm or formula is applied against the source (typically a file and its content, such as the image of a scanned page from a book) in order to generate a unique, 128-bit hash value often called a *checksum*. In digital preservation processes, the MD5 checksum from when the content was created is compared to another checksum created after the content has been received or stored over a period of time. The values are compared and, if they match, this indicates that the data (e.g. the scanned page image) is intact and has not been altered.

COLOR (specifications)

FEDORA (http://www.fedora-commons.org/ [2]) (Flexible Extensible Digital Object Repository Architecture) is a software framework to construct and maintain repositories of digital objects.

HYDRA (http://projecthydra.org/ [3]) is an open source respository software solution.

JPEG (Joint Photographic Experts Group) is the name of the group that developed the standard. JPG is a compression method for images.

JPEG 2000 is a wavelet-based image compression standard. It was created by the Joint Photographic Experts Group committee in the year 2000 with the intention of superseding their original discrete cosine transform-based JPEG standard (created about 1991). The standardized filename extension is JP2.

LADYBIRD (http://web.library.yale.edu/lit/projects/ladybird [4]) is Yale University Library's home-grown ingest tool for cataloging non-MARC records for digitized materials for display in a digital library interface on the web.

LZW (Lempel-Ziv-Welch) is a universal lossless data compression algorithm created by Abraham Lempel, Jacob Ziv, and Terry Welch.

MASTER (copy) is also known as Preservation Master or Archival Master.

OCR (Optical Character Recognition), computer software designed to convert images of text (usually captured by a scanner) into machine-editable text

PDF (Portable Document Format) is a file format, created by Adobe Systems, for document exchange in a manner independent of the application software, hardware, and operating system.

PRINT is determined when reformatting a fragile, brittle, or otherwise vulnerable volume characterized as such because of its physical condition.

PROCESSED (copy) is also known as the Mezzanine copy or the Access master or the Working master format.

RATIO

RESOLUTION

RETENTION

TIFF (Tagged Image File Format) is recognized as the best format for preservation and technical longevity.

TXT (.txt) is a file format used for textual documents usually containing very little formatting.

UTF-8 (UCS Unicode Transformation Format—8-bit) is a form of encoding that is backwards compatible with ASCII. The encoding standard is capable of displaying in email and in Internet browsers the standard 128 ASCII characters for English as well as Latin alphabet characters with diacritics, Greek, Cyrillic, Coptic, Armenian, Hebrew, and Arabic characters.

Effective Date: March 19, 2014

Source URL:https://web.library.yale.edu/digital-initiatives/digitization-standards-and-guidelines/glossary

Links
[1] http://www.digitizationguidelines.gov/glossary.php [2] http://www.fedora-commons.org/ [3] http://projecthydra.org/ [4] http://web.library.yale.edu/lit/projects/ladybird