**Tool #14: Digital Media and Files: Free Tools and Glossary**

*Free downloadable tools for managing and preserving digital assets.*

* **MediaInfo**: a free, open source platform that displays some of the most relevant technical and tag data for video and audio files (format, duration, bit rate, language, etc.) <http://mediaarea.net/en/MediaInfo>
* **VLC**: a free, open source cross-platform multimedia player and framework that plays most multimedia files (MPEG-2, MPEG-4, H.264, MKV, WebM, WMV, MP3, etc.) as well as DVDs, Audio CDs, VCDs, and various streaming protocols. Available for computer operating systems as well as mobile devices. <http://www.videolan.org/vlc/index.html>
* **FFMpeg**: free, open source multimedia framework, able to decode (i.e. decompress or make larger)**,** encode (i.e. compress or make smaller)**,** transcode (i.e. convert file from one format to another)**,** mux (i.e. combine multiple data inputs to a single output)**,** demux (i.e. separate one input into multiple outputs)**,** stream**,** filter (i.e. scale, speed up, etc.) and play many formats.<https://www.ffmpeg.org/about.html>
* **Better File Renamer**: file renaming utility. Can use metadata (when file was created, what format it is, who created it, etc.) to rename many files at once, create file names with detailed information in them, and more. <http://www.publicspace.net/ABetterFinderRename/>
* **IsoBuster**: Data recovery software; can recover data from some unreadable discs and drives <http://www.isobuster.com/>
* **ExifTool**: a platform-independent Perl (programming language) library plus a command-line application (in which the user issues commands by entering lines of text) for reading, writing and editing meta information in a wide variety of files. <http://www.sno.phy.queensu.ca/~phil/exiftool/>
* **Hashdeep**: a program to compute, match, and audit hashsets. Matching and auditing can compare sets of data and alert you if anything in a certain hashset has moved or changed. More information on hashsets (aka checksums) below in the glossary. <http://md5deep.sourceforge.net/>

**Glossary**

* **Analog**: Often used as shorthand for any format, material type, or technology that is not digital. For instance, a film camera is analog, as opposed to a digital camera. VHS tape is analog, but MiniDV tape is digital.
* **Born Digital**: A digital object that has never had an analog form. They differ from documents, movies and photographs that may have been scanned or converted to a digital format.
* **checksum**: (or hash sum or hash value) is a unique code (the digital “fingerprint”) created from running an algorithm off of data. Since alterations to data will change the code, it’s used for the purpose of [detecting errors](http://en.wikipedia.org/wiki/Error_detection) which may have been introduced during its [transmission](http://en.wikipedia.org/wiki/Telecommunication) or [storage](http://en.wikipedia.org/wiki/Computer_storage). Checksums can be used to check digital files to ensure they are not corrupted.
* **Decompression:**The process used to restore data to uncompressed form after compression.
* **Digital Provenance**: Information on the origin of a digital object and also on any changes that may have occurred over the course of its life cycle.
* **Emulation**: A means of overcoming technological obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers.
* **Format:** 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, and process, and save the information. It is dictated by the application program which created the file, and the operating system under which it was created and stored. Some file formats are designed for very particular types of data, others can act as a container for different types. A particular file format is often indicated by a file name extension containing three or four letters that identify the format.
* **fixity**: the property of a digital file being fixed, or unchanged. Fixity checking is the process of verifying that a digital object has not been altered or corrupted, often with use of checksums.
* **Format/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
* **Hash sum:** see checksum
* **Lossless Compression:** A mechanism for reducing file sizes that retains all original data.
* **Lossy Compression:** A mechanism for reducing file sizes that typically discards data.
* **metadata**: “data about data,” which can be used to describe either content, or technical and structural aspects of the way the content is held
* **migration**: transferring data to a new format or carrier to ensure content preservation and accessibility. This process can introduce data loss.
* **Open Format**: In a computer environment, an open format is a data format that is not considered proprietary and is free of commercial ownership or patents. Typically the technical specifications for the format are also publicly available, allowing users to alter and develop the format to suit their specific needs.
* **Open Source**: Open source refers to software in which the source code is available to the general public for use and/or modification from its original design. Open source code is typically created as a collaborative effort in which programmers improve upon the code and share the changes within the community.
* **Resolution**: The measure of how well audio, video, or film can faithfully portray images or sound. Picture cell (pixel) density and bit depth are the units of measure for individual images. Sampling rate and bit depth are the units of measure for moving images and sound.
  + **Bit Rate**: The amount of sample data that is collected per unit of time. This is usually expressed as ‘bits per second’ (bps) or ‘Megabits per second’ (Mbps), and is the result of sampling rate multiplied by the number of data bits per sample (or bit depth), plus any additional data such as tracking information.
  + **Bit Depth**: The number of data bits used for each individual sample. The amount of bits determines the number of discrete levels possible to place a sample within. 8-10 bit is common for video; 16-24 bit is common for audio.
  + **Sampling Rate**: The frequency at which information from the original recording is sampled or collected from a continuous signal. The rate is given in hertz (Hz) which equals cycles per second. One common example is the audio CD which has a sampling rate of 44,100 Hz, or 44.1 kHz.

**Sources**

[National Archives: Terms Used in the Preservation of Audio, Video and Motion Picture Film](https://www.archives.gov/preservation/formats/glossary.html)

[Digital Preservation Coalition Glossary](https://www.dpconline.org/handbook/glossary#B)

[University of Minnesota Digital Preservation Glossary](https://www.lib.umn.edu/dp/glossary)

[National Digital Stewardship Alliance Glossary](https://ndsa.org/glossary/)

[University of Michigan Library Digital Preservation Glossary](https://www.lib.umich.edu/preservation-and-conservation/digital-preservation-glossary)

[Society of American Archivists Glossary](https://www2.archivists.org/glossary/terms)