Terminology For The Creation, Care, And Storage Of Digital Materials

This guide provides definitions of commonly used terms for the creation, care, and storage of digital materials as covered within the NPS *Conserve O Gram* series. Links to the Conserve O Grams these terms appear in are provided. Also included are selected terms appearing in guidance on digitization projects from other sources, including the National Archives and Records Administration, Library of Congress, and Federal Agencies Digital Guidelines Initiative. Links to selected projects and organizations are provided at the bottom of this document.

Selected Definitions of Digitization and Preservation Terms

Backup – copy of a digital asset (image, document, media file, etc.) created as a replacement in case the original asset is lost or damaged; can also refer to an entire collection of backup files created at a single point in time (i.e. "Tuesday's backup")

Benchmark - quality model or standard used as a management tool for predicting (and judging) the results of imaging techniques. (COG 19/21)

Best practices – professionally developed guidelines for the development and management for digital data. (See the link to Federal Agencies Digitization Guidelines Initiative below). www.digitizationguidelines.gov

Bit depth - also known as dynamic range or color depth – a measurement that expresses the number of colors or tones that can be displayed in a digital image. The higher the number, the greater the range of colors that can be rendered. For instance, 4 bit = 16 colors and 8 bit = 256. 24 bit is known as "true color" and can render 16.7 + million colors. See also bitonal image, grayscale image, color image.(*COG* 19/21, *COG* 22/1)

Bitonal image – digital image consisting of only two tones/colors; in practical usage, this describes black and white images, but it can actually be any two colors. When scanning, the bitonal option may be described as "black and white" (as opposed to gray scale or color). (*COG* 22/1)

Blu-Ray disk, writable (BD-R) – optical media format that uses a more precise type of laser (blue in color) than CD or DVD, allowing a much larger amount of data to be written on a disk. (*COG* 22/5)

Born analog – content that was originally created using non-digital means, (such as books, reports, correspondence, maps, film photographs, or analog (tape) audio or video recordings); often used when referring to content that has since been digitized

Born digital – content originally created using digital means, such as a text document or a digital photograph.

Camera RAW – file format that holds the "raw" data captured directly from a camera's digital imaging sensor. These digital files hold the most data possible from an exposure, making them similar in purpose to film negatives. RAW files are proprietary (specific to the make and model of the camera) and must be converted to another format (e.g. TIFF or JPEG) to be editable. (*COG* 22/2)

Cloud storage – online, server-based storage of digital information. Usually managed by a third party and stores information across multiple servers kept in large data centers.

Color image – digital image composed of a color standard (such as RGB) with bit depths ranging from 8 to 24 bits. The RGB color standard has 8 bits for each channel (red, green, blue) for a total of 24 (3x8) bits. 24 bits allows the possibility of more than 16.7 million colors. See also bit depth. (*COG* 22/1)

CD – Compact Disc – optical storage media format consisting of a polycarbonate disk made up of complex laminate structures. Inexpensive and nearly universal, but has been largely eclipsed by optical storage media formats with much larger capacities. (*COG* 19/19)

CD-R – Compact Disc Recordable –type of recordable CD on which users can write data. Data is written using laser light to modify, or "burn" a layer of dye in the disk (*COG* 19/19, *COG* 22/5)

CD-ROM - Compact Disc Read-Only Memory –type of CD that provides access to data, but does not allow adding or editing data. See also CD.

CD-RW – Compact Disc Rewritable –type of recordable CD on which data can be written, modified and erased. Their complexity makes them less reliable (and less popular) than CD-R (*COG* 19/19, *COG* 22/5)

Compression - process of forcing more data into less space to speed processing, storage, and transmission. See also lossless compression and lossy compression. (*COG* 19/21, *COG* 19/22)

Computer (magnetic) tape – obsolete magnetic storage media composed of a thin, flexible magnetic medium. Originally stored on open reels.,In more modern applications it has been housed inside of plastic casings/cartridges. See also tape cartridge. (*COG* 19/11)

"**Dark" repository** – digital repository that has no public user access. A dark repository can be used solely for preservation purposes and be permanently dark, or it can be temporarily designated as dark while its contents are being added and verified prior to being made ready for public access.

Derivative/proxy – lower resolution (or file size) version of high-resolution files used as preservation digital masters. Often compressed to allow for easy electronic sharing or access. Example: a lower-resolution JPEG version of a large, high-resolution TIFF digital image file. See also digital master. (*COG* 19/21)

Digital master – high-resolution, uncompressed (or minimally compressed) file. It serves as the "master" copy of a resource from which further copies (i.e. derivatives/proxies) are made. See also derivative/proxy. (*COG* 19/21)

"Dim" repository – digital repository that has limited public user access. Repositories may customize different aspects of user access; e.g. one repository may allow users to add, access,

and delete files, whereas another may only allow a user to search for files.

DPI (dots per inch) - in digital printing, the number of individual dots of ink that a printer produces within a linear inch of paper. A higher DPI generally produces a sharper, more detailed image. (*COG* 19/22, *COG* 22/3)

Digital storage media – physical device (disk, card, drive, etc.) that facilitates storage of and easy access to electronic data. It does not provide long-term archival storage. See also flash memory media, magnetic storage media, optical storage media. (*COG* 19/20, *COG* 22/5)

DVD-ROM - Digital Video Disc Read-Only Memory – an optical storage media format similar in technology and appearance to a CD-ROM, but with a larger storage capacity. As with a CD-ROM, users can access stored data but are not able to add, modify or delete data.

DVD-R/+R – Digital Video Disc Recordable – similar in technology to a CD-R, but offers a much larger storage capacity. (*COG* 22/5)

Digitization – creating a digital version or copy of analog source material. (*COG* 19/11)

Dublin Core® - a metadata schema widely used by museums and other cultural institutions due to its flexibility in describing different kinds of digital resources.

External hard disk (HDD) drive – device that stores data on a magnetic disk drive housed within a portable metal or plastic case. Connects to a PC through a cable. (*COG* 22/5)

External solid-state disk (SSD) drive – device that offers the large storage capacity of an

external hard disk (magnetic) drive, but uses flash memory instead of magnetic media. SSD drives lack the moving parts of magnetic hard-drives and are therefore less prone to shock/impact damage. (*COG* 22/5)

Flash memory media – ype of digital storage media where data is stored in memory chips and is modified electronically (no moving parts). See also external solid-state disk (SSD) drive, memory card, and USB flash drive. (*COG* 22/5)

Hardware, computer – physical components of a computer; used in conjunction with computer software.

Histogram, image – bar graph that displays the relative distribution of all tonal values of a digital image; the more symmetrical the graph is, the more balanced the tonal (light/dark) distribution is. This tool is available in SLR (Single Lens Reflex) digital cameras and most image editing software. (*COG* 22/4)

Interoperability – ability of different computer software and/or hardware systems to work together.

File format, digital – particular way in which information is encoded for storage as a computer file. There are many different formats, each offering different levels of quality, file size, and compatibility. See camera RAW, JPEG, and TIFF (*COG* 19/22, *COG* 22/2)

Floppy disk – obsolete form of magnetic storage media. Consists of a flexible magnetic disk housed inside of a plastic casing. Information on floppy disks should be migrated to an accessible format if possible. (*COG* 19/11)

Grayscale image – digital image composed of multiple bits of tonal information, usually between 2 and 8 bits, including black, white, and numerous shades of gray in between. (*COG* 22/1)

Image, digital – visual representation, rendered in a digital format, produced or captured by a digital device (can be born analog or born digital); can be used to describe photos, drawings, maps, plans, scanned text, etc.

JPEG - Joint Photographic Experts Group –file format and industry standard for digital image storage and delivery over the internet and for non-professional digital photography. JPEG compression is lossy but produces a reasonably clear image with a relatively small file size. (*COG* 22/2)

LOCKSS (Lots of Copies Keep Stuff Safe) – international community initiative that provides digital preservation tools and support to libraries and archives. The term is often referenced in the museum community as a reminder of the importance of redundancy. See also, redundancy.

Lossless compression – type of compression that results in no loss of data; the entire original file can be recreated when the compression is removed. (*COG* 19/21)

Lossy compression – type of compression that results in some data loss; some of the original information is discarded to save space and cannot be retrieved. (*COG* 19/21)

Magnetic storage media – type of storage media where data is stored by altering the magnetic polarity on disks or tape, requiring that the tape or disk be moved or rotated. See also external hard disk drive, floppy disk, computer tape, tape cartridge, and ZIP disk. (*COG* 19/11, *COG* 19/20, *COG* 22/5)

Memory card – form of flash memory storage media that holds data intended for later transfer to a PC or other device. Most commonly used in digital cameras. (*COG* 22/5)

Metadata – textual information that describes a resource (physical or digital) regardless of its format. Sometimes descrbied as "data about data." Metadata usually includes management, provenance, rights, or organizational information.

Metadata schema – way of to structure or format metadata consistently; provides the location (order) and formatting for the information contained in metadata

Migration – process of converting files into new file formats or transferring onto new types of storage media in order to guard against data loss caused by technological obsolescence , or failure/degradation of media. (*COG 19/22*)

Obsolescence – the state, process, or condition of becoming obsolete (i.e. out-of-date). Electronic data created or stored on obsolete software or hardware can be difficult to preserve and/or migrate.

OCR (Optical Character Recognition) – technology that enables a scanner to recognize individual characters in a document, allowing a scan to produce actual editable, searchable text instead of a static image of text. (*COG 19/22*)

Optical storage media – type of digital storage media where information is read using a combination of lasers and optical sensors. See also, CD, CD-ROM, CD-R, DVD-ROM, DVD-R/+R, and Blu-Ray disk, writable (BD-

R). (*COG* 22/5)

Photograph - visual representation created as a result of [by / from] the reaction of radiant energy exposed to a light-sensitive media. Traditional (film) photography creates these visual representations on a physical film, while digital photography creates them using light-sensitive electronic sensors and memory chips.

Pixel - single picture element or tonal value in binary code; on a computer screen, it appears as a single "dot" or "point" and the smallest addressable element that makes up the displayed image. (*COG* 19/21)

PPI (Pixels Per Inch) – measure of image resolution that is represented by the number of individual square pixels per a linear inch within a digital range. PPI affects the quality of an image's printable size. (*COG* 22/3)

Proprietary – term that describes software or file formats that are specific to a particular program or vendor. Data loss can result if the program or vendor become defunct and data is not migrated. (*COG* 19/21, *COG* 19/22)

Recovery – also known as "restoration" – the process of copying backup files back to their original location; usually performed after data loss occurs

Redundancy – principle of creating duplicates of files so that if one copy of a file is lost or rendered inaccessible, one or more copies exist in other locations to replace it. Performing a backup results in redundancy.

Reformatting – copying informational content to a new medium or a new file format to improve access and/or protect the original from unnecessary touching/handling. (*COG* 19/10,

COG 19/11, *COG* 19/22)

Repository, digital – institution, organization or location where digital content is stored as part of a collection and made available for retrieval or use; sometimes referred to as a "digital library".

Resolution – measurement of an image's clarity/definition; when used in reference to scanning or printing, resolution is usually expressed in pixels (width by height – e.g. 1920 x 1280). Multiplying the width by the height gives the total number of pixels, often expressed as "megapixels" (for instance, 1920 x 1280 = 2.46 million pixels, or approximately 2.4 megapixels). (*COG* 19/21)

Software, computer – written instruction (or "code") that is run on computer hardware and enables it to perform tasks.

Scanner, flat-bed – digital imaging device that digitizes one item at a time on a flat surface. Primarily used to scan photographs or oversize records. (COG 19/22)

Scanner, sheet-fed/document – digital imaging device that scans documents assembled in stacks; because the documents are mechanically fed one-by-one and processed through the scanner, they are not suitable for photographs, or fragile, oversize, or bound items. (*COG* 19/22)

Scanner, slide – digital imaging device designed to scan transparent materials such as slides or negatives; may not work well if the original is opaque. (*COG* 19/22)

Scanning software – controls the scanner and manages bit depth, compression, enhancements, file formats, image manipulation,

resolution, and thresholding. Such software may be included with a computer's operating system, bundled with scanning hardware, or purchased separately. (*COG* 19/22).

TIFF – Tagged Image File Format –flexible and very widely used file format for digital images. TIFF files can be uncompressed or compressed (lossy or lossless). Often used for master or archival copies of files. (*COG* 22/2)

Tape cartridge – magnetic storage media consisting of spooled magnetic (computer) tape housed inside a plastic casing/cartridge. Usually used by IT professionals for critical back-ups of computer systems. (*COG* 19/11, *COG* 22/5)

Trusted Digital Repository (TDR) – digital repository with a focused mission to provide long-term, reliable access to the digital materials under their care. TDRs have attributes and conform to standards that are desirable (long-term) when choosing a digital repository or digital archiving system. See also repository, digital.

Trustworthy Repositories Audit and Certification (TRAC) – checklist and set of criteria that has become the de facto standard for assessing the reliability, commitment, and readiness of institutions to assume long-term preservation responsibilities for a repository or archive.

USB flash drive – also known as a "thumb drive" or "jump drive" –small, extremely portable flash memory-based storage media drive that is compatible with any PC with a USB port. (*COG* 22/5)

ZIP disk – obsolete form of magnetic storage media similar in construction to a floppy disk, but with a much higher storage capacity. Information on ZIP disks should be migrated to a more accessible format if possible. (*COG* 22/5)

NPS Online Digital Management Resources

Natural Resource Information Portal (NR Info Portal) – Used to share natural resources information and research. Currently available only to NPS employees; public access is in development

NPS Denver Services Center Technical Information Center (e-TIC) – (http://etic.nps.gov/) – online component of the NPS central repository for NPS-generated drawings, specifications, scientific and technical reports. Currently available only to NPS employees; public access is in development.

NPSFocus – (http://focus.inside.nps.gov) – allows NPS to upload images, documents, and other resources. (http://npsfocus.nps.gov) allows the public to view them.

NPS Library Information Center (Voyager) – (www.library.nps.gov/) – NPS online library catalog. Available to the public.

NPS SharePoint Portal – (http://portal.nps.gov) – Online digital collaboration system available only to NPS employees; allows easy sharing of documents and managing of projects).

Selected Resources Offering Guidance and Terminology Related to the Planning and Implementation of Digitization Projects

Archives & Museum Informatics
www.archimuse.com
(Conferences, consulting, publishing, and training for cultural resources professionals.)

Digital Curation and Preservation Bibliography
http://digital-scholarship.org/dcpb/dcpb.htm
(References that aid in understanding digital curation and preservation.)

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)
www.repositoryaudit.eu
(An audit-based risk-management tool used to select an appropriate digital repository based on a particular institution's needs and capabilities.)

Federal Agencies Digitization Guidelines Initiative (FADGI)
www.digitizationguidelines.gov/
Glossary: www.digitizationguidelines.gov/glossary.php
(Collaborative effort by federal agencies to define common guidelines, methods, and practices for digitizing historical content.)

Library of Congress, LOC.
Building Digital Collections: A Technical Overview
http://memory.loc.gov/ammem/about/techIn.html
(Information on standards and practices used by LOC in digitizing, providing access to, and preserving *American Memory* materials.)

Museum Computer Network. www.mcn.edu
(Provides opportunities to explore and disseminate new technologies and best practices in the museum field.)

National Archives and Records Administration, NARA.
Bibliography of Glossaries
www.archives.gov/research/alic/reference/archives-resources/glossaries-for-archivists.html
(Wide-ranging compilation of glossaries for archivists and related cultural information professionals)

National Digital Stewardship Alliance (NDSA) Glossary. www.digitalpreservation.gov/ndsa/ndsa-glossary.html

National Initiative for a Networked Cultural Heritage, NINCH. http://www.ninch.org
(Coalition of organizations aimed at providing leadership from the cultural community for digital issues.)

Society of American Archivists, SAA.
Glossary of Archival and Records Terminology
http://www2.archivists.org/glossary
(References for archivists and other associated professions.)

Lauren Finn
Museum Technician
Mesa Verde National Park