

GLOSSARY

AACR2: *Anglo-American Cataloging Rules, 2nd Edition*. Content rules used in the creation of cataloging records.

AAT: *Art and Architecture Thesaurus*; a publication of the Getty Information Institute. a thesaurus for terms to describe art and architecture.

access point: a name, term, phrase or code that is used to search, identify or locate a file, document, record, or object.

acquisitions information: information about the acquisition of the collection or objects by the repository.

Acrobat: Adobe's electronic document format. Documents can be created from within a word processor, from postscript, or from scanned pages. The documents are highly portable, yet maintain the look of the original. Acrobat is especially useful in this area because Adobe makes the reader available for free.

administrative information: information regarding the administration of the collection or object. May include acquisitions information, provenance, use restrictions, access restrictions, copyright ownership, citation information, and general processing information. Administrative information can refer to all or part of a collection.

administrative metadata: metadata primarily intended to facilitate the management of resources.

angled brackets: an SGML/XML syntax convention to set apart a tag, < >.

ANSI: American National Standards Institute, an organization which accredits other standards development organizations.

APPM: *Archives, Personal Papers and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, by Steven L. Hensen. Published by SAA as a supplemental set of cataloging rules. APPM has been superseded by DACS.

ASP: Active server pages – web pages that use scripts to access information from a database stored on the server. See also html.

attribute: modifier for the meaning of elements, named properties of an element that may carry different values depending upon the context in which they occur.

authority control: the process of verifying and authorizing the choice of unique access points, such as names, subjects, and forms and assuring that the access points are consistently applied and maintained in an information retrieval system. See also controlled vocabulary.

authority file: a group of authority records searchable by all established headings and cross-references.

authority record: an entry that contains information about an access point. An authority record establishes the form of the heading, determines cross-references and relationships of the heading to other headings.

biographical/historical note: highlights of the life and activities of a person, family, or corporate body that generated the document described therein. A biographical/historical note is intended to provide contextual information for researchers.

bit depth: see *dynamic range*.

BMP: Windows Bitmap. Usually uncompressed but can be compressed (lossless). Up to 32 bit depth. Standards for Windows Imaging. Large file sizes. Not supported in some browsers and some non-Windows applications.

boilerplate text: standardized text used for labels and other text used for all digital files (i.e. copyright notice, citation format, etc.).

CCO: *Cataloging Cultural Objects.*

CDWA: *Categories for the Description of Works of Art*, a metadata standard for describing works of art for the purpose of art historical scholarship. Developed by the Getty Information Institute.

close tag: the tag that closes an element, also called an end tag.

component level: an EAD expression for the hierarchy of nested information in a finding aid.

compression: the re-encoding of data to make it smaller. Most image file formats use compression because image files tend to be large and consume large amounts of disk space and transmission time over networks.

controlled access: a list of index terms for a finding aid.

controlled vocabulary: formal limits on a vocabulary, useful for consistent use of vocabulary terms.

copyright term: the length of time during which the copyright is honored.

crosswalk: an authoritative mapping from the metadata elements of one scheme to the elements of another.

DACS: *Describing Archives: a Content Standard*, published and officially endorsed by the Society of American Archivists as an output-neutral content standard for archival description.

DAT: digital audio tape, a magnetic tape originally designed for use in audio applications, but now popular for storing data. Capacities range up to 12 gigabytes.

DCMI: Dublin Core Metadata Initiative, an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models, responsible for the maintenance of the Dublin Core metadata scheme.

derivative image: an image that has been created from another image. Usually involves a loss of information. Techniques to create derivative images include sampling to a lower resolution, using lossy compression techniques, or altering an image with image manipulation software during image processing.

descriptive metadata: metadata primarily intended to serve the purposes of discovery, identification, and selection.

digitization: the conversion from printed paper, film, or other media formats to an electronic format where the object is represented as either black and white dots, color or grayscale pixels, or 1s and 0s.

DjVu: An electronic document format primarily useful for scanning documents. Key features are support for different resolutions and compression types for photo areas of an image versus text. Uses a variant of JBIG2 compression from binary image data and wavelets for continuous areas such as photos. For more information, see <http://www.princetonimaging.com/djvu/>

DLT: digital linear tape, a fairly new high end tape format. Capacities range up to 35 gigabytes.

download: to transmit a file from one computer to another. Usually implies retrieving a file from a remote computer to a local one, or from a large computer to a smaller one. **FTP** is a commonly used command for this.

dtd: document type definition, the formal specifications and definitions of the structural elements and markup to be used in encoding specific types of documents in SGML/XML.

Dublin Core: metadata element set created to facilitate the discovery of electronic resources. Consists of core 15 elements and is typically used in conjunction with HTML. Maintained by the DCMI (<http://www.dublincore.org/>)

DVD: digital video disk, an optical storage medium that can store up to 4.7 gigabytes (single layer) 8.5 GB (double layer), 9.4 GB (double sided, single layer), or 17 GB (double sided, double layer). Transfer rates and seek times are similar to those of CD-ROMs for currently available drives. The DVD specs include higher level specs for audio and video capabilities.

dynamic range: the number of colors or shades of gray that can be represented by a pixel. The smallest unit of data stored in a computer is called a bit. Dynamic range is a measurement of the number of bits used to represent each pixel in a digital image. Also called bit depth.

EAD: Encoded Archival Description, an SGML/XML dtd for the description of archival finding aids that reflects the hierarchical arrangement of archival materials. EAD provides a framework for information storage, retrieval and display on the World Wide Web. Maintained by SAA with web support from the Library of Congress (<http://www.loc.gov/ead/>)

electronic document: a document that consists of 1s and 0s and requires hardware and software for access. Documents become more useful when stored electronically because they can be widely distributed instantly and allow searching. Best practice for the preservation of electronic document is still underdevelopment. HTML and PDF are well known electronic document formats.

element: an essential building block of metadata schemes that serves to identify and surround the content of sections of the metadata. Elements are constructed of a *open tag* and a *close tag*. Elements may contain other elements, attributes and values, PCDATA or be empty.

encoding rules: the syntax or prescribed order for the elements contained in the metadata description.

end tag: see *close tag*

entity: an independent file that is used to include external information.

finding aid: a tool used to communicate the contents of an archival collection, the finding aid typically includes administrative information, contextual information, scope and content information, intellectual organization and physical location information for archival and manuscript materials.

FPX: Flashpix, a file format that is 8-24 bit depth and uncompressed. Developed by Kodak. Flashpix can be compressed and audio can be embedded in images. It supports text fields, stores various resolutions in one file, has consistent color, but is not supported by most software.

G4 Compression: A compression technique used in Fax Group 4. It produces very good results for black and white images, and is frequently used as an option in TIFF files. It is also used in Adobe Acrobat (PDF) files.

GIF: Graphics Interchange Format. An 8-bit image file format that is commonly used on the Web. GIF uses LZW compression, which makes it good for color and grayscale images, but it does not compress as well as G4 for black and white. LZW is "lossless" which means it will not compress as well as JPEG, but will retain all of the images quality. PNG is designed to replace GIF.

grayscale: an image type that uses black, white, and a range of shades of gray. The number of shades of gray depends on the number of bits per pixel. The larger the number of shades of gray, the better the image will look, and the larger the file will be.

HTML: hypertext markup language; most common procedural markup language found on the Web. An international standard for coding text to make it appear with formatting on web pages. HTML includes the structure of documents (title, headings, etc.) and the formatting (bold, fonts, and font size). For example, `Headline` would make the word **Headline** appear in bold.

HTTP: Hypertext Transfer Protocol. The protocol designed to convert HTML code so web browsers can interpret and display web pages.

ICR: Intelligent Character Recognition. The processes of recognizing handwritten characters. Similar to OCR, but more difficult since OCR is from printed text. Used for forms you fill out that are then scanned to gather information you have provided on the form.

image capture: using a scanner, digital camera, or other device to create a digital representation of an object.

image file format: when a page is scanned, the page can be stored in a number of file formats. The type should be chosen based on the desired use of the image, and the software that will be used. Different file formats commonly use different methods of compression as well, and some types of images compress better using some formats rather than others.

image manipulation: making changes (i.e. tonal adjustments, cropping, moiré reductions, etc.) to an image using image processing software; altering the image from its original digital capture.

instance: the text and tags (excluding the dtd and related files) of an individual SGML/XML-encoded document, such as a single EAD-encoded finding aid.

interoperability: the ability of multiple systems, using different hardware and software platforms, data structures, and interfaces, to communicate, exchange, and share data.

ISAD(G): General International Standard for Archival Description, a general framework for archival description developed by the International Council on Archives.

ISBN: International Standard Book Number, an identifier for nonserial print publications.

ISO-9660: International Standards Organization 9660, *Information processing – Volume and file structure of CD-ROM for information exchange*. A file system format standard developed for CD-ROMs using the CD-XA encoding standard. It is supported by Microsoft operating systems, UNIX, and Macintosh.

JBIG: A “lossless” image compression format for binary (black and white) images. Compresses better than G4 by up to 25 percent. Also supports progressive encoding. Licensing issues have slowed its adoption for use.

JBIG2: A “lossy” image compression format for binary (black and white) images. A JBIG2 compressor identifies common objects (usually characters) in the image and creates a dictionary with references to those objects. Lossiness is induced by allowing similar objects to be represented by a single dictionary entry. This format is supported in PDF 1.4 and greater.

JPG, JPEG: Joint Photographic Experts Group. An 8-24 bit image file format that is best suited for photographs. It supports “lossiness,” which means that it will throw away some detail in order to achieve better compression. It has variable amount of compression to vary quality and file size. It

does not work well for text. Widely used as a delivery format.

JPEG 2000: An image format that provides the inclusion of metadata and structural elements for the image within the code stream.

LCNAF: Library of Congress Name Authority File. A controlled vocabulary used for the names of persons, corporate bodies, uniform and series titles available at <http://authorities.loc.gov/>

LCSH: *Library of Congress Subject Headings*, A controlled vocabulary used for creating subject terms and geographical terms.

link: encoding that is used for navigation. The link is seen on the browser-side and allows the user to "click" on it to go somewhere else in the document or on the internet.

MARC: Machine-Readable Cataloging. Data structure standard used in Integrated Library Systems (ILS) for Online Public Access Catalogs (OPACs).

metadata: structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource.

metadata harvesting: a technique for extracting metadata from individual repositories and collecting it in a central catalog to facilitate search interoperability.

metadata scheme: a set of metadata elements and rules for their use that has been defined for a particular purpose

metalanguage: a language used to describe other languages. SGML and XML are examples of metalanguages.

METS: Metadata Encoding and Transmission Standard, a specification for structural metadata.

migration: a digital preservation technique to preserve the integrity of digital files by transferring them across hardware and software configurations and subsequent generations of computer technology.

navigation: moving around a document or the internet

nesting: the way in which SGML/XML sub-elements may be contained within other elements to create a multilevel document.

noise: data or unidentified marks picked up in digital capture or data transfer that do not correspond to the original.

OAI: Open Archives Initiative, an organization that maintains a protocol for harvesting metadata from distributed repositories.

OCR: Optical Character Recognition, a process that produces a page of text from an image file.

open tag: the tag that opens an element, also called a *start tag*.

parent element: an element that may contain other elements, referred to as subelements of the parent element.

parse: a check against the XML syntactic rules. See also *validate*.

PCD: ImagePac, PhotoCD Lossy compression. 24 bit depth. Has 5 layered image resolutions. Used mainly for delivery of high quality images on CD.

PCT: PICT Compressed. Mac standard. Up to 32 bit. Supported by Macs and a highly limited number of PC applications.

PDF: Portable Document Format, 4-64bit depth. Uncompressed. Used mainly to image documents for delivery. Need plug-in or adobe application to view. Adobe's Portable Document Format, the term Adobe uses to describe Acrobat files. See also *Acrobat*.

PCDATA: parsable character data, i.e. text

pixel: short for picture elements, which make up an image. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it.

PNG: Portable Networks Graphics, lossless compression. 24 bit. Replaced GIF due to copyright issues on the LZW compression. Some programs cannot read it.

portable: to be functional across differing types of computers and operating systems. This can be used to describe programs or electronic documents.

preservation metadata: metadata primarily intended to help manage the process of ensuring the long-term preservation and usability of digital information resources.

progressive encoding: a method by which multiple resolutions of the same image is stored in the same image file. Imaging systems can efficiently serve lower-than-maximum resolutions with images encoded this way. Total file size is increased, but smaller amounts of data can be transmitted to clients.

proofing: a service by which the resulting OCR text or PDF file is repaired for errors induced by the electronic process.

provenance: history of ownership of materials prior to acquisition by the current institution.

qualifier: see *refinement*

quality control: techniques used to ensure that high quality is maintained through the various stages of digitization.

quantization: to reduce the number of colors or shades of gray in an image, with the goal being to reduce the file size while maintaining image quality. Also used to display images with more colors than are available on the display device.

Refinement: In Dublin Core and other metadata schemes, a term that restricts the

meaning of an element or identifies the encoding scheme used in representing the value of the element (also known as a qualifier).

resolution: the number of pixels (in both height and width) making up an image. The more pixels, the higher the resolution; the higher the resolution, the greater its clarity and definition and the greater the file size. Can be expressed as a ratio (640 x 480 pixels) or in terms of dots per inch (dpi). It is recommended that you use between 72 and 100 dpi for images that will be displayed on the screen, and 300 dpi for images that will print on common inexpensive printers.

rights metadata: metadata primarily intended to enable the management of rights related to information resources; a type of administrative metadata.

RLG: Research Libraries Group, a not-for-profit membership organization of libraries, archives, museums, and other cultural heritage institutions now part of OCLC.

RLIN: Research Libraries Information Network, a cataloging system and union catalog run by RLG.

SAA: Society of American Archivists.

scanning: see *digitization*

schema: in XML a way of defining a document type used as an alternative to the dtd.

scheme: a formally defined set of metadata elements or fields.

scope and content note: note containing information regarding the scope and content of an archival collection. This note is written to provide an overall description of the collection and includes such information as significant topics covered by the collection, significant individuals, associations, or corporations, societies, and events documented by the collection, and the extent to which the material covers these

topics. It can also include the media the collection exists in and its organization.

semantics: the definitions of the meaning of metadata elements, as opposed to the rules for encoding or representing the values of the elements, see also *syntax*.

server: host computer for web pages

SGML: Standard Generalized Markup Language, SGML is a platform-neutral standard for creating documents. It is a series of rules that define document structures.

skew: during printing or scanning, the degree to which the page is not vertical. Deskewing is a process where the computer detects and corrects the skew in an image file.

source code: the code (usually HTML) behind any web page viewed by a browser. To see the source code of a page in Internet Explorer, right click on the page and select View Source or click on View on the tool bar and then select View Source. In Netscape, it is referred to as Page Source.

SPF: Still Picture Interchange Format (SPIFF), Official JPEG format. Lossless compression, supports text datafields, thumbnails, alternative color spaces. There is not a lot of support for this format, but it is designed to be read by applications that can handle jpg.

start tag: see *open tag*.

structural metadata: metadata that describes the internal organization of resource and its place in an external organization, including any relationships it has with other resources.

subelement: an element that is available within one or more other elements. In EAD, every element except the document element <ead> is a subelement of one or more *parent elements*.

surrogate: a secondary object meant to substitute for the original, such as a photograph of an artwork used in place of the artwork.

syntax: how a metadata scheme is structured for exchange in a machine-readable form, including the rules regarding that structure rather than their meaning. Common syntaxes include MARC, SGML, and XML. See also *semantics*.

tag: the syntactic expression of an element; tag and element are used interchangeably, although tag refers to the actual representation of the element, while element refers to the intellectual content of the tag.

tag library: a document that lists the names of the SGML or XML elements and attributes alphabetically, along with their definitions, tag names, and rules for their use.

technical metadata: metadata primarily intended to document the creation and characteristics of digital files.

TGA: TARGA format, compressed or uncompressed. Up to 32 bit, common in animation packages, good for interchange.

thesaurus: a *controlled vocabulary* with syndectic structure in which all allowable terms are given and relationships between terms are shown.

thresholding: when converting a pixel from grayscale to black and white, the threshold is the gray value above which will be considered white and below or equal to which will be considered black.

TIF, TIFF: Tagged Image File Format, an industry standard image file format. Uncompressed, originally developed for desktop publishing. 1 to 64 bit depth, used mostly for high quality imaging and archival storage. Generally non-compressed and high quality, including large file sizes. Most TIFF readers only read a maximum of 24-bit color. Delivery over web is hampered by file

sizes, although LZW compression can reduce these file sizes by 33%, it should not be used for archival material. It is unique in that it incorporates multiple compression techniques, allowing the user to specify the best format for a type of image, and that one file can contain multiple images.

unicode: syntactic representation of special characters to eliminate conflict between XML syntax and textual content. A complete set of Unicode values is available at:

<http://www.ncecho.org/ncead/documents/unicode.htm>

URL: Uniform Resource Locator – the “address for a web site” (Ex.: <http://www.ncecho.org/Guide/toc.htm>) HTTP is the method of connection; www.ncecho.org is the name of the host computer or server, also known as the domain name; /Guide/ is the particular directory on that computer; and index.htm is the specific file. .htm is the kind of file, also called the file extension. Note that URLs are a specific kind of URI (Uniform Resource Indicator).

validate: to assure that a document conforms to the rules in a *dtd*. See also *parse*.

value: the specific expression of an attribute.

vocabulary: the universe of values that can be used for a particular metadata element.

W3C: World Wide Web Consortium, an international committee working to provide vision and standards for the internet.

wrapper element: an element designed only as a container for other elements. Wrapper elements may have attributes and values but must contain one or more sub-elements in order to include text.

XHTML: eXtensible Hypertext Markup Language, an emerging markup language that combines information about the structure of a document and the structure of the data. The purpose of XHTML is to allow the exchange of information from different types of database.

XML: eXtensible Markup Language, a way of coding text to allow for content searching and manipulation. An adaptation of SGML for the use on the Web.

Z39.50: An ANSI/NISO standard protocol for system-to-system search and retrieval. Also International Standard, ISO 23950 “Information Retrieval (Z39.50): Application Service Definition and Protocol Specification” This standard is commonly used for the interchange of information in library catalogs and other databases.

zooming: make an image appear larger (zoom in) or smaller (zoom out) by re-displaying the image at different resolutions. Higher resolutions will make the image appear larger and easier to read.