

## Texas Heritage Digitization Initiative Glossary

### **administrative metadata**

metadata used for the administration of a collection or digital object. May include acquisitions information, provenance, use restrictions, access restrictions, copyright ownership, citation information, and general processing information.

### **archival image**

an image kept for preservation purposes. Current best practices for archival images call for uncompressed TIFF (Tagged Image/Interchange File Format) or another uncompressed, nonproprietary format. Sometimes called the “digital master” or “master image.”

### **bit depth**

the number of colors or shades of gray that can be represented by a pixel. Also called color or pixel depth.

### **bitonal**

an image mode that captures each pixel of an image as either a white (255) or black (0) value. Bitonal scanning is sometimes used in scanning typed text or line art.

### **collection**

a group of resources brought together by shared content, provenance, ownership, subject, etc. Digital objects can simultaneously belong to multiple collections.

### **compression**

the re-encoding of data to make it smaller. May be “lossy,” discarding information from the original file, or “lossless,” which preserves the original information.

### **controlled vocabulary**

a list of terms that may be used to describe resources in order to ensure consistency and improve information retrieval.

### **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.

### **digital asset management system (DAMS)**

a software system, frequently in the form of a database, used to organize and process digital assets like images, documents and presentations.

### **digital object**

an item or resource in digital format. May be the result of digitization or may be “born digital.” Sometimes referred to as a resource when metadata has been assigned.

### **digitization**

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

**DNG**

Digital Negative, a publicly available archival format for the raw files generated by digital cameras, developed by Adobe Systems. DNG is used by some camera manufacturers, and a free DNG converter is available from Adobe to translate RAW files from other manufacturers into DNG.

**DPI (dots per inch)**

a measure of print resolution, i.e, how many individual dots are produced within one linear inch.

**Dublin Core**

metadata element set created to facilitate the discovery of electronic resources, maintained by the Dublin Core Metadata Initiative (DCMI). Consists of core 15 elements and is typically used in conjunction with HTML.

**file format**

the underlying structure of a digital file, often characterized by its extension (e.g., .pdf, .jpg, etc.). Different types of files, such as text, images, audio, etc. require different formats. Some formats use compression to minimize file size, while others preserve all the information in the original.

**grayscale**

an image mode that captures 256 shades of gray (8-bit). Every pixel of a grayscale image has a brightness value ranging from 0 (black) to 255 (white). Grayscale mode is commonly used in scanning black and white photographs.

**harvester**

a client application that issues OAI-PMH requests. A harvester is operated by a service provider as a means of collecting metadata from repositories.

**image capture**

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

**image manipulation**

making changes to an image using image processing software; altering the image from its original digital capture.

**interoperability**

the ability of multiple systems, using different hardware and software platforms, data structures, and interfaces, to communicate, exchange, and share data. See also OAI-PMH and Z39.50.

**JPG, JPEG**

Joint Photographic Experts Group, or an image file format commonly used for photographs, based on a standard developed by that group. JPEG2000 is a newer standard based on different technologies.

**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 crosswalk**

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

**metadata harvesting**

a technique for extracting metadata from individual repositories and collecting it in a central catalog to facilitate search interoperability. See also OAI-PMH

**metadata scheme**

a set of metadata elements and rules for their use that has been defined for a particular purpose. Examples include MARC, MODS, Dublin Core, EAD, and many others.

**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.

**OAI-PMH**

The Open Archives Initiative's Protocol for Metadata Harvesting, an application-independent interoperability framework based on metadata harvesting, maintained by the Open Archives Initiative.

**OCR**

Optical Character Recognition, a process that produces a page of text from an image file. Accuracy rates vary greatly depending on the quality of the image.

**pixel**

short for picture element, the building block of digital images. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it.

**PPI (pixels per inch)**

resolution of a computer display, image scanner or digital camera, measured in terms of the number of pixels displayed per linear inch of display.

**preservation metadata**

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

**quality control**

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

**RAW**

the unprocessed data from the image sensor of a digital camera. For most digital cameras, RAW is a proprietary format, and the data must be processed and converted to an RGB format such as TIFF or JPEG before it can be manipulated by a bitmap graphics editor, printed, or displayed by a typical web browser. See also DNG.

**repository**

a storage system for digital resources and/or metadata, such as a DAMS. In OAI-PMH, a network accessible server that can process OAI-PMH requests.

**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 the file size. Can be expressed as a ratio (640 x 480 pixels) or in terms of pixels per inch (ppi).

**rights metadata**

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

**structural metadata**

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

**surrogate**

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

**technical metadata**

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

**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.

**unique identifier**

a name or number that unambiguously identifies an item within a repository or database. Unique identifiers can be self-describing (e.g., ISBNs) or assigned.

**URI**

Uniform Resource Indicator, a formatted string that serves as an identifier for a resource, typically on the Internet. The most common type of URL is the URL (Uniform Resource Locator) used in HTML.

**Z39.50**

An ANSI/NISO standard protocol for system-to-system search and retrieval. This standard is commonly used for the interchange of information in library catalogs and other databases.