



CERP GLOSSARY

Active record:

A record used frequently in conduct of daily business.

Administrative Metadata:

Information needed to manage digital content and that is not part of the digital resource itself. Examples include acquisition date, copyright ownership, and disposition date.

Archival Information Package (AIP):

Originally accessioned digital content plus content converted to preservation format (such as xml) and associated metadata required for storage in a repository such as DSpace.

Archival record:

Information with legal, financial, administrative, or research value that should be kept permanently according to an organization's Records Retention and Disposition Schedule.

ASCII:

A text file where each character or space is represented by one byte encoded according to the ASCII (American Standard Code for Information Interchange) code. It preserves Latin-based alphabetical characters, punctuation marks, and some symbols and formatting.

ASP (Active Server Page):

This web page format uses scripting, normally VBScript or JavaScript code in combination with HTML, to dynamically generate a complete HTML page for display on the requesting web browser. The complete HTML is not generated until that page is requested by a web browser.

Audit Trail:

A record of actions performed on a computer system. It includes user identification as well as time and date information.

Authenticity:

A record that is what it purports to be and has not changed since its creation. Authentic e-mail includes the e-mail message as well as any attachments and its transmission data.

Born-digital:

Material (text, images, audio, video) that was created in a digital format. Not to be confused with digitized materials that have been converted from paper or other original type to a digital format by scanning or other methods.

CFM:

Cold Fusion template/page. Cold Fusion is a Macromedia web development application used to create dynamic web pages.

Convenience copy:

A copy of a record kept for reference and quick access.

CSV:

Comma Separated Values. Another name for comma-delimited text format. CSV preserves the data input (not formulae or formatting), allowing a spreadsheet or database to be recreated later.

DBF:

Database format used by various applications.

Descriptive metadata:

Information within and external to an electronic record that references selected components of its content for use in identifying or locating the record, such as a finding aid, a search term, or type media.

Digital curation:

Management and preservation of digital objects (data generated in binary code) over their lifecycle of current and future use, ensuring the data retains its authenticity, access, reproducibility, and longevity. It includes selection, appraisal, intellectual control, redundant storage, data migrations, bitstream preservation, and metadata capture and creation.¹

Digital obsolescence:

Digital data that was created in outdated programs or operating systems or on old media that is difficult or impossible to access in the current digital environment.

Digital record:

Information created or stored in a format that provides evidence of activities, events, decisions, programs, policies, or transactions. It may be born-digital or digitized.

Dissemination Information Package (DIP):

An information package, such as an e-mail accession or a journal, delivered from a digital repository upon request from an archivist or researcher.

Discovery:

Legal process in which one party to a lawsuit is required to furnish documents requested by the opposing side.

Disposition:

Routine, planned disposal of records by scheduled transfer (for permanent) or destruction (for non-permanent).

Document management system:

Computer software that files, routes, and retrieves documents created electronically regardless of the document's original format (Word, Excel, etc.).

DPI:

Dots per inch. A means of expressing the amount of information recorded in a digital image correlating to the resolution quality or density of the image.

DSpace:

Open-source content management software originally called Durable Space and developed by MIT and Hewlett-Packard for use in preserving, storing, and allowing access to digital information. Its community of users, primarily academic institutions, determines their own policies for deposit, storage, and retrieval; however, preservation is at the bitstream level with only a few formats renderable. See <http://www.dspace.org/>.

Dublin Core:

ISO/ANSI standard (15836/Z39.85) that defines metadata elements used to describe and provide access to online resources. Elements include title, creator, subject, publisher, date, etc. See <http://dublincore.org/>.

EAD (Encoded Archival Description):

EAD is the non-proprietary standard for-encoding finding aids for use in an online environment.

E-Discovery:

Legal process in which one party to a lawsuit, or an organization subject to governmental regulation, is required to furnish documents generated and/or maintained in electronic formats to the opposing side upon their request.

8.3:

The MS-DOS file-naming convention of eight characters followed by a period (.) and three final characters. The three final characters are popularly used as acronyms for the file format of the electronic document. For example, "demo.ppt" is a Microsoft PowerPoint document. PPT would be the ".3" expression, or the acronym for a PowerPoint file.

Electronic Communications Privacy Act (ECPA):

Federal law that defines invasion of privacy regarding electronic communication, including e-mail, cellular telephones, pagers, etc.

Electronic document management system:

Computer program that enables an organization to manage its electronic documents from creation through storage and retrieval. (Note: this is not the same as archiving electronic documents.)

Electronic record:

Information created or stored in an electronic form that provides evidence of activities, events, decisions, programs, policies, or transactions. Electronic records include born-digital, digitized, and non-digital content such as video tapes.

Electronic signature:

According to the New York Electronic Signatures and Records Act, an electronic signature is "an electronic identifier, including without limitation a digital signature, which is unique to the person using it, capable of verification, under the sole control of the person using it, attached to or associated with data in such a manner that authenticates the attachment of the signature to particular data and the integrity of the data transmitted, and intended by the party using it to have the same force and effect as the use of a signature affixed by a hand."

eml:

E-mail format used by Microsoft Outlook Express and other e-mail applications.

Emulation:

Way of mimicking hardware or software so other processes think that the original equipment or system is still operating in its original form.

Encryption:

Method of hiding electronic information by encoding it so that only authorized persons who have the decryption code may access the data.

Enterprise Content Management (ECM):

Use of technology to manage an organization's information flow from creation through storage. The term typically is used when referring to a company that provides software that captures, preserves, and retrieves electronic records. ECM also often includes management of digital rights, web content, and records retention.

E-Sign (Electronic Signatures in Global and National Commerce Act):

Federal law that gives electronic signatures the same legal status as handwritten signatures with regard to electronic transactions.

Format:

Type of computer file, e.g. Microsoft Excel or JPEG image.

HTML:

HyperText Markup Language, is a markup language for Web pages.

HTML Help Compiled Help File (CHM):

Microsoft format for online help files.

Information Technology (IT):

The system that handles information generated or stored through computers and telecommunications. Also known as Information Services (IS) or Management Information Services (MIS).

Integrity:

Confirmation that a record has not been altered, intentionally or accidentally, since its creation or receipt.

Internet Header:

Metadata viewable through e-mail software tools that gives information in addition to that shown in an e-mail message body. The Internet Header gives IP addresses of sending and receiving computers, date and time stamps, and other details which may authenticate the message.

JPEG/JPG:

JPEG is a lossy compression technique for color images developed by the Joint Photographic Experts Group. File sizes can be reduced, but with a loss in detail. JPG is an alternate representation of JPEG.

LAN (Local Area Network):

A network of personal computers, usually within each location of an organization, that allows transmission of data within the network.

Life Cycle Management:

Retaining or destroying documents when they reach a pre-determined age and in accordance with government regulations, legal or financial guidelines, or an organization's internal policies regarding records retention.

MARC:

MAchine Readable Cataloging, a format for structured descriptive bibliographic, authority, classification, and holdings data. (Based on ANSI Information Interchange Format standard Z39.2.) See <http://lcweb.loc.gov/marc/>.

MDB:

Format for Microsoft Access database (2003 and earlier).

MBOX:

A generic format for e-mail messages. All messages in an MBOX mailbox are concatenated and stored as plain text in a single file.

Metadata:

Internal metadata is information inherent within a digital document automatically produced when an electronic document is created, sent, modified, or received that describes its subject, date created, sender, recipients, etc. External metadata refers to preservation, technical, and descriptive information not part of the document itself that is created by a document creator, archivist, or other user. Metadata is used to identify, manage, preserve, and access digital information and includes format, size, accession source and date, disposal date, migration requirements, etc.

METS (Metadata Encoding and Transmission Standard)

An XML format used for depositing text and image digital content and encoding its descriptive, administrative, and structural metadata necessary for managing digital accessions in a digital repository and for sharing that content with other repositories and users. A METS document is usually a required component of SIPs, AIPs, and DIPs.

Migration:

The process of transferring data from one electronic format to another, usually from older technology to newer. This is done to preserve information that might otherwise be lost as the old format becomes obsolete.

MIME (Multipurpose Internet Mail Extension):

The standard encoding method for e-mail attachments most frequently used.

.msg:

A proprietary binary e-mail format used by Microsoft Outlook.

Near-line storage:

Storing information in an electronic format apart from the e-mail system, such as on a desktop computer's hard drive or a shared drive. E-mail remains somewhat functional.

Official copy (also known as record copy):

Original record or a copy that is retained in compliance with an organization's Records Management Policy and Records Retention Schedule. If the e-mail is created within the organization, the sender usually maintains the official copy. When it is received from outside the organization, the primary recipient usually holds the official record.

Official record:

Information created or received in the course of conducting an organization's business, and required by law or deemed appropriate to be preserved, either short or long term.

Off-line storage:

Storing information outside an electronic environment, such as on paper copies, magnetic tape, optical disk, or computer-output-to-microfilm.

Online storage:

Storage of e-mail, metadata, and attachments within the e-mail system currently being used by an organization. E-mail remains fully functional, i.e., it can still be forwarded, replied to, etc.

Open Archives Initiative (OAI):

An organization that developed and published application-independent interoperability standards to facilitate management and sharing of online content from harvested metadata. See <http://www.openarchives.org/>.

OAIS (Open Archival Information System) reference model:

Model serves as a reference for long-term preservation and access of digital materials in a repository: how digital objects can be prepared, placed in an archive, and stored, maintained, and retrieved. Many in the cultural heritage field have adopted it for their digital preservation efforts because of its flexibility and acceptance.

Parser:

A computer program that interprets digital data input such as e-mail text and converts it to XML or other format.

Portable Document Format (PDF):

Software developed by Adobe Systems that operates on several platforms (Mac, Windows, UNIX, etc.) and converts a variety of formats including Microsoft Word, Publisher, and PowerPoint, into a file that usually looks almost exactly like the original. The PDF version loses some automatically generated metadata and may lose some special formatting such as underlining. PDF is an open standard under the International Organization for Standardization (ISO) 32000.

Preservation Metadata:

Technical information required for managing and preserving digital assets over time to ensure the digital objects remain viable. It includes documentation of preservation actions such as migration, as well as collection and rights management information.

Personal Storage File (PST):

Microsoft Outlook proprietary format that creates one file containing all selected e-mail messages and attachments. It is stored outside of the e-mail server.

Record:

Formal or informal information generated within an organization or received by it during its course of business. A record may be in various forms whether printed or electronic, including book, CD/DVD, e-mail, instant message, map, memory card or stick, handwritten notes, memos, and sketches, photograph or other image, spreadsheets, audio or video tape, voice-mail (See Official Record).

Records Management Application (RMA) or Records Management System (RMS): Electronic document management system with an added feature that applies the organization's retention schedule to determine how long to retain a particular record. The purchasing organization usually works with the software provider to assign recognition identifiers (such as keywords in e-mail subject headings) and retention criteria.

Records Management Policy:

A formal, written document containing an organization's procedures for managing records of its activities. It typically includes guidelines regarding which records to retain, the length of time they should be kept, the manner in which they should be organized, and the procedures for disposing of them or transferring them to an archive.

Records retention schedule:

A list of an organization's records by record type that indicates how long each type should be retained.

Refreshing:

The process of transferring data from one electronic media to another, usually from older technology to newer. This is done to preserve information that might otherwise be lost as the old media deteriorates or becomes obsolete.

Retention period:

An organization's pre-determined "expiration" dates – the point in time when a record may be destroyed. Financial, legal, and governmental requirements, in addition to the organization's administrative needs and the historical value of the records, are considerations in establishing retention periods.

RGB:

Red, Green, Blue components of a color TIFF image

RTF:

Rich Text Format. A format standard which embeds basic formatting instructions in an essentially ASCII document. Margins, font style, indentation and other formatting instructions are supported.

Schema:

An expression of data structure and content in tagged format, usually in XML, that enables machines to perform tasks ordered by human computer operators.²

Security log:

A record of access, attempted access, and use of a computer system automatically kept by security software such as a virus protection program.

Signature line:

Lines of user-determined text, usually containing name, title, organization name, and contact details, set to be automatically entered by an e-mail client at the end of an outgoing message. (Note: this is not the same as an electronic or digital signature.)

SMTP (Simple Mail Transport Protocol):

Commonly used rules for e-mail transmission through the Internet.

Source file:

Digital files as originally created or deposited/donated to an archive or other repository.

Spoliation:

Unauthorized, whether accidental or deliberate, destruction of records pertinent to lawsuits or regulatory body investigations, or potential suits or investigations.

Structural Metadata:

Information about the divisions, views, extent, sequence, use, and relationship between parts of a compound object, such as pages and chapters of a book, table of contents, PDF file for download and printing, TIFF file for display, etc.

Submission Information Package (SIP):

Source data and relevant metadata provided to an archive by the data creator or a person or entity acting on the creator's behalf.

SWF:

Shockwave file format, commonly referred as a Flash component, used by Macromedia's Flash player application. An increasingly popular plug-in, or supplemental application, used with web browsers.

Tags:

Symbols used in electronic documents that instruct a program how to display the documents, e.g., font type and size.

TCP (Transport Control Protocol):

The rules that enable computers to communicate with each other through the Internet.

Text file:

An electronic file that can be read by many computer programs because it consists solely of ASCII characters and formatting. Extension is TXT.

TIFF:

Tagged Image File Format, a popular format for storing bit-mapped images, supports black-and-white, grayscale, and color images.

Unicode:

A character encoding standard developed by the Unicode Consortium. By using more than one byte to represent each character, Unicode enables almost all of the written languages in the world to be represented by using a single character set.

URL:

Universal Resource Locator. The "address" of an Internet-accessible document. Most frequently begins with `http://...` but also includes `ftp://...` and `telnet://...`

W3C (World Wide Web Consortium):

The organization responsible for managing standards for the WWW.³

WinHelp:

HLP is a proprietary format for online help files that can be displayed by the Microsoft Help browse.

XML:

Extensible Markup Language. A non-proprietary text format that is self-describing and flexible, making it attractive as a preservation format. XML is derived from the Standard Generalized Markup Language (SGML).

XHTML:

Extensible Hypertext Markup Language. This information standard essentially expresses HTML code in an XML syntax. XHTML 1.0 has been recognized by the Internet-related vendors as the successor to HTML 4.0 and is the equivalent of the most recently adopted HTML 4.1 protocol.

Sources

Society of American Archivists

http://www.archivists.org/glossary/term_details.asp?DefinitionKey=1185.

W3Schools. Refsnes Data.

http://www.w3schools.com/site/site_glossary.asp.

File Extension Source.

<http://www.filext.com/>

¹ Texas Digital Library, *Journal of Digital Information*, Vol. 8, No. 2 (2007),
(<http://journals.tdl.org/jodi/article/view/229/183>).

² W3C, XML Schema 2000

³ W3Schools. Refsnes Data. http://www.w3schools.com/site/site_glossary.asp.

Version 2 - October 2008

Version 1 - September 2008