

Appendix E: Glossary

This glossary is intended to define terms of art used in this white paper or common to the disciplines of records management and information technology as they relate to topics covered here, including the identification, collection, and analysis of information and records for investigation and litigation. This glossary is not comprehensive or exhaustive of such terms. References to “DoD 5015” refer to Department of Defense “Design Criteria for Electronic Record Management Software Applications” (October 2003).

Active Data: Active Data is information residing on the direct access storage media (disk drives or servers) of computer systems, which is readily visible to the operating system and/or application software with which it was created and immediately accessible to users without restoration or reconstruction.

Active Records: Active Records are those Records related to current, ongoing or in-process activities and are referred to on a regular basis to respond to day-to-day operational requirements. An active record resides in native application format and is accessible for purposes of business processing with no restrictions on alteration beyond normal business rules. *See* Inactive Records.

Ambient Data: *See* Residual Data.

Application: An application is a collection of one or more related software programs that enables a user to enter, store, view, modify or extract information from files or databases. The term is commonly used in place of “program,” or “software.” Applications may include word processors, Internet browsing tools and spreadsheets.

Archival Data: Archival Data is information that is not directly accessible to the user of a computer system but that an organization maintains for long-term storage and record-keeping purposes. Archival data may be written to removable media such as a CD, magneto-optical media, tape or other

electronic storage device, or may be maintained on system hard drives or network servers.

Archive, Electronic Archive: Archives are long term repositories for the storage of records. Electronic archives preserve the content, prevent or track alterations and control access to electronic records. *See* the discussion of electronic archives in the Technical Appendix, Appendix D.

Attachment: An attachment is a record or file associated with another record for the purpose of storage or transfer. There may be multiple attachments associated with a single “parent” or “master” record. The attachments and associated record may be managed and processed as a single unit. In common use, this term refers to a file (or files) associated with an e-mail for transfer and storage as a single message unit. Because in certain circumstances the context of the attachment—for example, the parent e-mail and its associated metadata—can be important, an organization should consider whether its policy should authorize or restrict the disassociation of attachments from their parent records.

Attribute: An attribute is a characteristic of data that sets it apart from other data, such as location, length, or type. The term attribute is sometimes used synonymously with “data element” or “property.”

Author or Originator: The author of a document is the person, office or designated position responsible for its creation or issuance. In the case of a document in the form of a letter, the author or originator is usually indicated on the letterhead or by signature. In some cases, the software application producing the document may capture the author's identity and associate it with the document. For records management purposes, the author or originator may be designated as a person, official title, office symbol or code. (DoD 5015)

Backup Data: Backup Data is information that is not presently in use by an organization and is routinely stored separately upon portable media. Backup data serves as a source for recovery in the event of a system problem or disaster. Backup data is distinct from "Archival Data."

Backup Tape Recycling: Backup Tape Recycling describes the process whereby an organization's backup tapes are overwritten with new data, usually on a fixed schedule determined jointly by records management, legal and IT sources. For example, the use of nightly backup tapes for each day of the week with the daily backup tape for a particular day being overwritten on the same day the following week; weekly and monthly backups being stored offsite for a specified period of time before being placed back in the rotation.

Backup tapes: See Disaster Recovery Tapes.

Compact Disk (CD): A type of optical disk storage media, compact disks come in a variety of formats. These formats include CD-ROMs ("CD-Read-Only-Memory") that are read-only; CD-Rs ("CD-Recordable") that are write to once and are then read-only; and CD-RWs (CD-Read-Write") that are write to in multiple sessions.

Computer Forensics: Computer Forensics (in the context of this document, "forensic analysis") is the use of specialized

techniques for recovery, authentication and analysis of electronic data when an investigation or litigation involves issues relating to reconstruction of computer usage, examination of residual data, authentication of data by technical analysis or explanation of technical features of data and computer usage. Computer forensics requires specialized expertise that goes beyond normal data collection and preservation techniques available to end-users or system support personnel, and generally requires strict adherence to chain-of-custody protocols.

Custodian: See Record Custodian.

Data Element: A combination of characters or bytes referring to one separate piece of information, such as name, address, or age. (DOD 5015)

Database Management System

(DBMS): A software system used to access and retrieve data stored in a database. (DOD 5015)

Database: In electronic records, a set of data elements, consisting of at least one file or of a group of integrated files, usually stored in one location and made available to several users. (DOD 5015)

De-Duplication: De-Duplication ("De-Duping") is the process of comparing electronic records based on their characteristics and removing or marking duplicate records within the data set.

Delete, Deletion: The process of permanently removing, erasing or obliterating recorded information from a medium, especially a reusable magnetic disk or tape. (DOD 5015) Deletion is the process whereby data is removed from active files and other data storage structures on computers and rendered inaccessible except by using special data recovery tools designed to recover deleted data.

Deleted Data: Deleted Data are data that existed on the computer as live data and

which have been deleted by the computer system or end-user activity. Deleted data may remain on storage media in whole or in part until they are overwritten or “wiped.” Even after the data have been wiped, directory entries, pointers or other information relating to the deleted data may remain on the computer. “Soft deletions” are data marked as deleted (and not generally available to the end-user after such marking), but not yet physically removed or overwritten. Soft-deleted data can be restored with complete fidelity.

Disaster Recovery Tapes: Disaster Recovery Tapes are portable media used to store data for backup purposes. *See* Backup Data.

Disposition: The final business action carried out on a record. This action generally is to destroy or archive the record. Electronic record disposition can include “soft deletions” (*see* Deletion), “hard deletions,” “hard deletions with overwrites,” “archive to long-term store,” “forward to organization,” and “copy to another media or format and delete (hard or soft).”

Distributed Data: Distributed Data is that information belonging to an organization which resides on portable media and non-local devices such as remote offices, home computers, laptop computers, personal electronic assistants (“PDAs”), wireless communication devices (*e.g.*, Blackberry), internet repositories (including e-mail hosted by internet service providers or portals and web sites) and the like. Distributed data also includes data held by third parties such as application service providers and business partners. In the event of litigation, distributed data may present additional issues for collection and analysis. *Note:* Information Technology organizations may define distributed data differently (for example, in some organizations distributed data includes any non-server-based data, including workstation disk drives).

Draft Record: Draft records can include working files such as preliminary drafts, notes, supporting source documents and similar materials. Organizations may determine that drafts should be retained if (1) they contain unique information including the substantive mental impressions of the author as to a business policy, decision, action or responsibility; or (2) they reflect substantive comments, annotations or comments by persons other than the author concerning a business policy, decision, action or responsibility; or (3) they are transmitted, circulated or made available to persons other than the author for business purposes such as approval, comment, action, recommendation or follow-up.

Electronic Mail: Electronic Mail, commonly referred to as “e-mail,” is an electronic means for communicating information under specified conditions, generally in the form of text messages, through systems that will send, store, process, and receive information, and in which messages are held in storage (until the addressee accesses them).

Electronic Mail Message: A document created or received via an electronic mail system, including brief notes, formal or substantive narrative documents, and any attachments, such as word processing and other electronic documents, which may be transmitted with the message. 36 CFR 1234.2, reference (aa). (DOD 5015)

Electronic Record: Information recorded in a form that requires a computer or other machine to process it and that otherwise satisfies the definition of a record. (DOD 5015)

File Plan: A document containing the identifying number, title, description and disposition authority of files held or used in an office. (DOD 5015)

Forensic Copy: A forensic copy is an exact copy of the entire physical storage

media (hard drive, CD-ROM, DVD-ROM, tape, etc.), including all active and residual data and unallocated space on the media. Forensic copies are often called “image or imaged copies”.

Format: The internal structure of a file, which defines the way it is stored and used. Specific applications may define unique formats for their data (e.g., “MS Word document file format”). Many files may only be viewed or printed using their originating application or an application designed to work with compatible formats. Computer storage systems commonly identify files by a naming convention that denotes the format (and therefore the probable originating application) (e.g., “DOC” for Microsoft Word document files; “XLS” for Microsoft Excel spreadsheet files; “TXT” for text files; and “HTM” (for Hypertext Markup Language (HTML) files such as web pages). Users may choose alternate naming conventions, but this may affect how the files are treated by applications.

Hold: See Legal Hold.

Image Copy, Imaged Copy: See Forensic Copy.

Inactive Record: Inactive records are those Records related to closed, completed, or concluded activities. Inactive Records are no longer routinely referenced, but must be retained in order to fulfill reporting requirements or for purposes of audit or analysis. Inactive records generally reside in a long-term storage format remaining accessible for purposes of business processing only with restrictions on alteration. In some business circumstances, inactive records may be re-activated.

Information: For the purposes of this document, information is used to mean both documents and data.

Instant Message, Instant Messaging (“IM”): Instant Messaging is a form of electronic communication, which involves immediate correspondence between two or

more users who are all online simultaneously. Some IM communications (peer-to-peer) may not be stored on servers after receipt.

Janitor Program: An application which runs at scheduled intervals to manage business information by deleting, transferring, or archiving on-line data (such as e-mail) at specific points in time. Janitor programs are sometimes referred to as “agents”—software that runs autonomously “behind the scenes” on user systems and servers to carry out business processes according to pre-defined rules.

Legacy Data, Legacy System: Legacy Data is information in which an organization may have invested significant development resources and which has retained its importance but has been created or stored by the use of software and/or hardware that has become obsolete or replaced (“legacy systems”). Legacy data may be costly to restore or reconstruct when required for investigation or litigation analysis or discovery.

Legal Hold: A legal hold is a communication issued as a result of current or anticipated litigation, audit, government investigation or other such matter that suspends the normal disposition or processing of records. The specific communication to business or IT organizations may also be called a “hold,” “preservation order,” “suspension order,” “freeze notice,” “hold order,” or “hold notice.”

Lifecycle: The records lifecycle is the life span of a record from its creation or receipt to its final disposition. It is usually described in three stages: creation, maintenance and use, and archive to final disposition.

Metadata: Metadata is information about a particular data set which describes how, when and by whom it was collected, created, accessed or modified and how it is formatted (including data demographics such as size, location, storage requirements and media

information). *See* Technical Appendix D for discussion of Metadata.

Migration: Moving files to another computer application or platform which may require changing their formats.

Mount, Mounting: The process of making off-line data available for on-line processing. For example, placing a magnetic tape in a drive and setting up the software to recognize or read that tape. The terms “load” and “loading” are often used in conjunction with, or synonymously with, “mount” and “mounting” (as in “mount and load a tape”). “Load” may also refer to the process of transferring data from mounted media to another media or to an on-line system.

Native Format: Electronic documents have an associated file structure defined by the original creating application. This file structure is referred to as the “native format” of the document. Because viewing or searching documents in the native format may require the original application (for example, viewing a Microsoft Word document may require the Microsoft Word application), documents are often converted to a vendor-neutral format as part of the record acquisition or archive process. *Cf.* “Static” formats.

Near-line data storage: Storage in a system that is not a direct part of the network in daily use, but that can be accessed through the network. There is usually a small time lag between the request for data stored in near-line media and its being made available to an application or end-user. Making near-line data available will not require human intervention (as opposed to “off-line” data which can only be made available through human actions).

Official Record Owner: *See* Record Owner.

Off-line data: The storage of electronic data outside the network in daily use (*e.g.*, on backup tapes) that is only accessible through the off-line storage system, not the network.

On-line storage: The storage of electronic data as fully accessible information in daily use on the network or elsewhere.

Preservation Notice, Preservation Order: See Legal Hold.

Record: Information, regardless of medium or format, that has value to an organization. Collectively the term is used to describe both documents and electronically stored information.

Record Custodian: A records custodian is an individual responsible for the physical storage and protection of records throughout their retention period. In the context of electronic records, custodianship may not be a direct part of the records management function in all organizations. For example, some organizations may place this responsibility within their information technology department, or they may assign responsibility for retaining and preserving records with individual employees. For this reason, this publication discusses the possibility of having a content custodian and a technology custodian.

Record Lifecycle: The time period from when a record is created until it is disposed.

Record Owner: The records owner is the subject matter expert on the content of the record and is responsible for the lifecycle management of the record. This may be, but is not necessarily, the author of the record.

Record Series: A description of a particular set of records within a file plan. Each category has retention and disposition data associated with it, applied to all record folders and records within the category. (DOD 5015)

Records Hold: *See* Legal Hold.

Records Management: Records Management is the planning, controlling, directing, organizing, training, promoting and other managerial activities involving the life-cycle of information, including creation,

maintenance (use, storage, retrieval) and disposition, regardless of media.

Records Manager: The records manager is responsible for the implementation of a records management program in keeping with the policies and procedures that govern that program, including the identification, classification, handling and disposition of the organization's records on all media throughout their retention life. The physical storage and protection of records may be a component of this individual's functions, but it may also be delegated to someone else. *See* Records Custodian.

Records Retention Period, Retention Period: The length of time a given records series must be kept, expressed as either a time period (*e.g.*, four years), an event or action (*e.g.*, audit), or a combination (*e.g.*, six months after audit).

Records Retention Schedule: A plan for the management of records, listing types of records and how long they should be kept; the purpose is to provide continuing authority to dispose of or transfer records to historical archives.

Records Store: *See* Repository for Electronic Records.

Record Submitter: The Record Submitter is the person who enters a record in an application or system. This may be, but is not necessarily, the author or the record owner.

Recover, Recovery: *See* Restore.

Report: Formatted output of a system providing specific information.

Repository for Electronic Records: Repository for Electronic Records is a direct access device on which the electronic records and associated metadata are stored. (DoD 5015) Sometimes called a "records store" or "records archive."

Residual Data: Residual Data (sometimes referred to as "Ambient Data") refers to data that is not active on a computer system.

Residual data includes (1) data found on media free space; (2) data found in file slack space; and (3) data within files that has functionally been deleted, in that it is not visible using the application with which the file was created, without use of undelete or special data recovery techniques.

Restore: To transfer data from a backup medium (such as tapes) to an on-line system, often for the purpose of recovery from a problem, failure, or disaster. Restoration of archival media is the transfer of data from an archival store to an on-line system for the purposes of processing (such as query, analysis, extraction or disposition of that data). Archival restoration of systems may require not only data restoration but also replication of the original hardware and software operating environment. Restoration of systems is often called "recovery".

Retention Schedule: *See* Records Retention Schedule.

Sampling: Sampling usually (but not always) refers to the process of testing a database for the existence or frequency of relevant information. It can be a useful technique in addressing a number of issues relating to litigation, including decisions about what repositories of data are appropriate to search in a particular litigation and determinations of the validity and effectiveness of searches or other data extraction procedures. Sampling can be useful in providing information to the court about the relative cost burden versus benefit of requiring a party to review certain electronic records.

Slack Space: A form of residual data, slack space is the amount of on-disk file space from the end of the logical record information to the end of the physical disk record. Slack space can contain information soft-deleted from the record, information from prior records stored at the same physical location as current records, metadata fragments and other

information useful for forensic analysis of computer systems.

Spoliation: Spoliation is the destruction of records which may be relevant to ongoing or anticipated litigation, government investigation or audit. Courts differ in their interpretation of the level of intent required before sanctions may be warranted. *See* Guideline 3.

Static formats. “Static” formats (often called “imaged formats”) are designed to retain a “picture” of the document as it would look viewed in the original creating application but do not allow manipulation of the document information; such formats may be well-suited for many records and litigation uses where access to document metadata and preservation of original document structures are not important. *Cf.* Native Formats

Suspension Notice, Suspension Order: See Hold.

System: A system is: (1) a collection of people, machines and methods organized to perform specific functions; (2) an integrated whole composed of diverse, interacting, specialized structures and sub-functions; and/or (3) a group of sub-systems united by some interaction or interdependence, performing many duties but functioning as a single unit.

Version, Record Version: A particular form of or variation from an earlier or original record. For electronic records, the variations may include changes to file format, metadata or content.

Vital Record: A record that is essential to the organization’s operation or to the reestablishment of the organization after a disaster.

Web site: A collection of Uniform Resource Indicators (URIs, including URLs (Uniform Resource Locators)) in the control of one administrative entity. May include different types of URIs (*e.g.*, file transfer protocol sites, telnet sites, as well as World Wide Web sites).