GLOSSARY OF COMMON E-DISCOVERY TERMS

*All definitions are reproduced with permission from the Sedona Conference and appear in The Sedona Conference Glossary: eDiscovery & Digital Information Management, Fifth Edition, 21 SEDONA CONF. J. 263 (2020), except those marked with a *. Certain definitions from the Glossary have had citations, cross-references, or alterations omitted for clarity and brevity. Access the complete Sedona Conference Glossary.*

Active data - Information residing on the direct-access storage media (disk drives or servers) that is readily visible to the operating system and/or application software with which it was created. It is immediately accessible to users without restoration or reconstruction.

Algorithm - With regard to electronic discovery, a computer script that is designed to analyze data patterns using mathematical formulas and is commonly used to group or find similar documents based on common mathematical scores.

Archival Data – Information an organization maintains for long-term storage and record-keeping purposes, but which may not be immediately accessible to the user of a computer system. Archival data may be written to removable media or may be maintained on system hard drives. Some systems allow users to retrieve archival data directly, while other systems require the intervention of an IT professional.

Artificial intelligence (AI) - A subfield of computer science focused on the development of intelligence in machines so that the machines can react and adapt to their environment and the unknown. AI is the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience. It attempts to approximate the results of human reasoning by organizing and manipulating factual and heuristic knowledge. Areas of AI activity include expert systems, natural language understanding, speech recognition, vision, and robotics.

Backup Tape - Magnetic tape used to store copies of electronically stored information, for use when restoration or recovery is required. The creation of

backup tapes is made using any of a number of specific software programs and usually involves varying degrees of compression.

Bates Number - Sequential numbering system used to identify individual pages of documents where each page or file is assigned a unique number. Often used in conjunction with a suffix or prefix to identify a producing party, the litigation, or other relevant information.

Boolean Search - Boolean searches use keywords and logical operators such as "and," "or," and "not" to include or exclude terms from a search, and thus produce broader or narrower search results.

Bring Your Own Device Policy (BYOD) - A policy whereby an organization specifies how personal computing devices, like smart phones, personal laptops, or portable tablets, can be used in the context of work for that organization, and may include provisions for the ownership and discoverability of the organization's data stored on the device. *See* The Sedona Conference, *Commentary on BYOD: Principles and Guidance for Developing Policies and Meeting Discovery Obligations*, 19 SEDONA CONF. J. 495 (2018), available at https://thesedonaconference.org/publication/Commentary_on_BYOD.

Chain of Custody - Documentation regarding the possession, movement, handling, and location of evidence from the time it is identified to the time it is presented in court or otherwise transferred or submitted; necessary to establish both admissibility and authenticity, and important to help mitigate risk of spoliation claims.

Clawback Agreement - An agreement outlining procedures to be followed if documents or electronically stored information are inadvertently produced; typically used to protect against the waiver of privilege.

Client Server - An architecture whereby a computer system consists of one or more server computers and numerous client computers (workstations). The system is functionally distributed across several nodes on a network and is typified by a high degree of parallel processing across distributed nodes. With client-server architecture, CPU intensive processes (such as searching and indexing) are completed on the server, while image viewing and Optical Character Recognition (OCR) occur on the client. This dramatically reduces network data traffic and insulates the database from workstation interruptions.

Cloud Computing - A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Continuous Active Learning (CAL) - A machine-learning algorithm that periodically analyzes users' decisions in order to rank unreviewed data, with the most likely desired data ranking first based on the users' previous decisions. See also Technology-Assisted Review.

Discovery on discovery* - Discovery on discovery is discovery requests directed at the manner and methods that opposing counsel used to locate, preserve, search, review, and produce relevant information.

Electronically Stored information (ESI) - As referenced in the U.S. Federal Rules of Civil Procedure, information that is stored electronically, regardless of the media or whether it is in the original format in which it was created, as opposed to stored in hard copy (i.e., on paper).

Email String - An electronic conversation between two or more parties via email. Also referred to as an email thread.

F-Measure - Also known as the F1 Score or the F Score, a measure of a search's accuracy calculated by using precision and recall. (Precision x Recall) / (Precision + Recall).

Hash Coding (also Hash Value, Hash) - A mathematical algorithm that calculates a unique value for a given set of data, similar to a digital fingerprint, representing the binary content of the data to assist in subsequently ensuring that data has not been modified. Common hash algorithms include MD5 and SHA.

Inactive data* - Files, fragments and artifacts that reside in unallocated and/or slack space in a data source including deleted files that have not yet been overwritten.

Internet of Things (IoT) - A catchall term used to describe a broad array of electronic devices, such as computers or sensors in cars, refrigerators, lights, or security systems, that are connected to the internet and may collect, store, and/or share information.

Legacy Data, Legacy System - Electronically stored information that can only be accessed via software and/or hardware that has become obsolete or replaced. Legacy data may be costly to re-store or reconstruct when required for investigation or litigation analysis or discovery.

Legal Hold - A communication issued as a result of current or reasonably anticipated litigation, audit, government investigation, or other such matter that suspends the normal disposition or processing of records. Legal holds may encompass procedures affecting data that is accessible as well as data that is not reasonably accessible. The specific communication to business or IT organizations may also be called a hold, preservation order, suspension order, freeze notice, hold order, litigation hold, or hold notice. *See* The Sedona Conference, *Commentary on Legal Holds, Second Edition: The Trigger & The Process*, 20 SEDONA CONF. J. 341 (2019), available at https://thesedonaconfer-ence. org/publication/Commentary_on_Legal_Holds.

Linear and Nonlinear Review - Performed by humans. Linear review workflow begins at the beginning of a collection and ad-dresses information in order until a full review of all information is complete. Nonlinear review workflow is to prepare only certain portions for review, based either on the results of criteria, such as search terms, technology-assisted review results, or some other method, to isolate only information likely to be responsive.

Load file - A file that relates to a set of scanned images or electronically processed files, and that indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends. A load file may also contain data relevant to the individual documents, such as selected metadata, coded data, and extracted text. Load files should be obtained and provided in prearranged or standardized formats to ensure transfer of accurate and usable images and data.

Meet and confer* - A requirement in some jurisdictions that parties to a suit must meet and discuss various matters and attempt to resolve disputes without court action.

Metadata - The generic term used to describe the structural information of a file that contains data about the file, as opposed to describing the content of a file. See System-Generated Metadata and User-Created Metadata. For a more thorough discussion, *see* The Sedona Conference, *The Sedona Guidelines: Best Practice*

Guidelines & Commentary for Managing Information & Records in the Electronic Age, Second Edition (November 2007), available at https://thesedonaconference.org/publication/Guidelines_for_Managing_Information_and_Electronic_Records, and The Sedona Conference, *Commentary on Ethics & Metadata*, 14 SEDONA CONF. J. 169, available at https://thesedonaconference.org/publication/Commentary_on_Ethics_and_Metadata.

Mirror image - A bit-by-bit copy of any storage media. Often used to copy the configuration of one computer to an[o]ther computer or when creating a preservation copy.

Modern attachments* - Also known as embedded files or "pointers", they are essentially hyperlinks that direct users to related electronic resources or documents. Modern attachments are often found in collaborative applications or chat platforms, where they point to documents that are stored on a shared network or cloud.

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 may be converted to a neutral format as part of the record acquisition or archive process. Static format (often called imaged format), such as TIFF or PDF, is designed to retain an image of the document as it would look viewed in the original creating application but does not allow metadata to be viewed or the document information to be manipulated unless agreed-upon metadata and extracted text are preserved. In the conversion to static format, some metadata can be processed, preserved, and electronically associated with the static format file. However, with technology advancements, tools and applications are increasingly available to allow viewing and searching of documents in their native format while still preserving pertinent metadata. It should be noted that not all electronically stored information may be conducive to production in either the native format or static format, and some other form of production may be necessary. Databases, for example, often present such issues.

Operating System (OS) - The operating system provides the software platform that directs the overall activity of a computer, network, or system and on which all other software programs and applications run. In many ways, choice of an operating system will affect which applications can be run. Operating systems

perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files and directories on the disk, and controlling peripheral devices such as disk drives and printers. For large systems, the operating system has even greater responsibilities and powers—becoming a traffic cop to make sure different programs and users running at the same time do not interfere with each other. The operating system is also responsible for security, ensuring that unauthorized users do not access the system. Examples of computer operating systems are UNIX, DOS, Microsoft Windows, LINUX, Mac OS, and IBM z/OS. Examples of portable device operating systems are iOS, Android, Microsoft Windows, and BlackBerry. Operating systems can be classified in a number of ways, including: multi-user (allows two or more users to run programs at the same time; some operating systems permit hundreds or even thousands of concurrent users); multiprocessing (supports running a program on more than one CPU); multitasking (allows more than one program to run concurrently); multithreading (allows different parts of a single program to run concurrently); and real time (instantly responds to input; general-purpose operating systems, such as DOS and UNIX, are not real time).

Precision - When describing search results, precision is the number of true positives retrieved from a search divided by the total number of results returned. For example, in a search for documents relevant to a document request, it is the percentage of documents returned that are actually relevant to the request. *See* The Sedona Conference, *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 15 SEDONA CONF. J. 217 (2014), available at https://thesedonaconfer-ence.org/publication/Commentary_on_Search_and_Retrieval_Methods.

Privilege log* - A privilege log is a document that describes documents or other items withheld from production under a claim that the documents are "privileged" from disclosure due to the attorney–client privilege, work product doctrine, joint defense doctrine, or some other privilege.

Proportionality* - A global "cost-benefit" analysis conducted by courts, which assesses the appropriateness of a discovery request by weighing the needs of the case and the importance of the information against the burden of producing it.

Quick Peek - An initial production whereby documents and/or electronically stored information are made available for review or inspection before being reviewed for responsiveness, relevance, privilege, confidentiality, or privacy.

Recall - When describing search results, recall is the number of documents retrieved from a search divided by all of the responsive documents in a collection. For example, in a search for documents relevant to a document request, it is the percentage of documents returned compared against all documents that should have been returned and exist in the data set. *See* The Sedona Conference, *Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 15 SEDONA CONF. J. 217 (2014), available at https://thesedonaconfer-ence.org/publication/Commentary_on_Search_and_Retrieval_Methods.

Subject Matter Expert (SME)* - The physical custodian or subject-matter expert on the contents of the record who is responsible for the lifecycle management of the record. This may be, but is not necessarily, the author of the record.

Spoliation* - The destruction of records or properties, such as metadata, that 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.

Supervised Learning - Use of machine learning to analyze data, using training examples that have been coded by humans, such as categorization.

Technology-Assisted Review (TAR) - A process for prioritizing or coding a collection of electronically stored information using a computerized system that harnesses human judgments of subject-matter experts on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection. Some TAR methods use algorithms that determine how similar (or dissimilar) each of the remaining documents is to those coded as relevant (or nonrelevant) by the subject-matter experts, while other TAR methods derive systematic rules that emulate the experts' decision-making processes. TAR systems generally incorporate statistical models and/or sampling techniques to guide the process and to measure overall system effectiveness.

Unstructured Data - Free-form data that either does not have a data structure or has a data structure not easily readable by a computer without the use of a specific program designed to interpret the data; created without limitations on formatting or content by the program with which it is being created. Examples include word-processing documents or slide presentations.