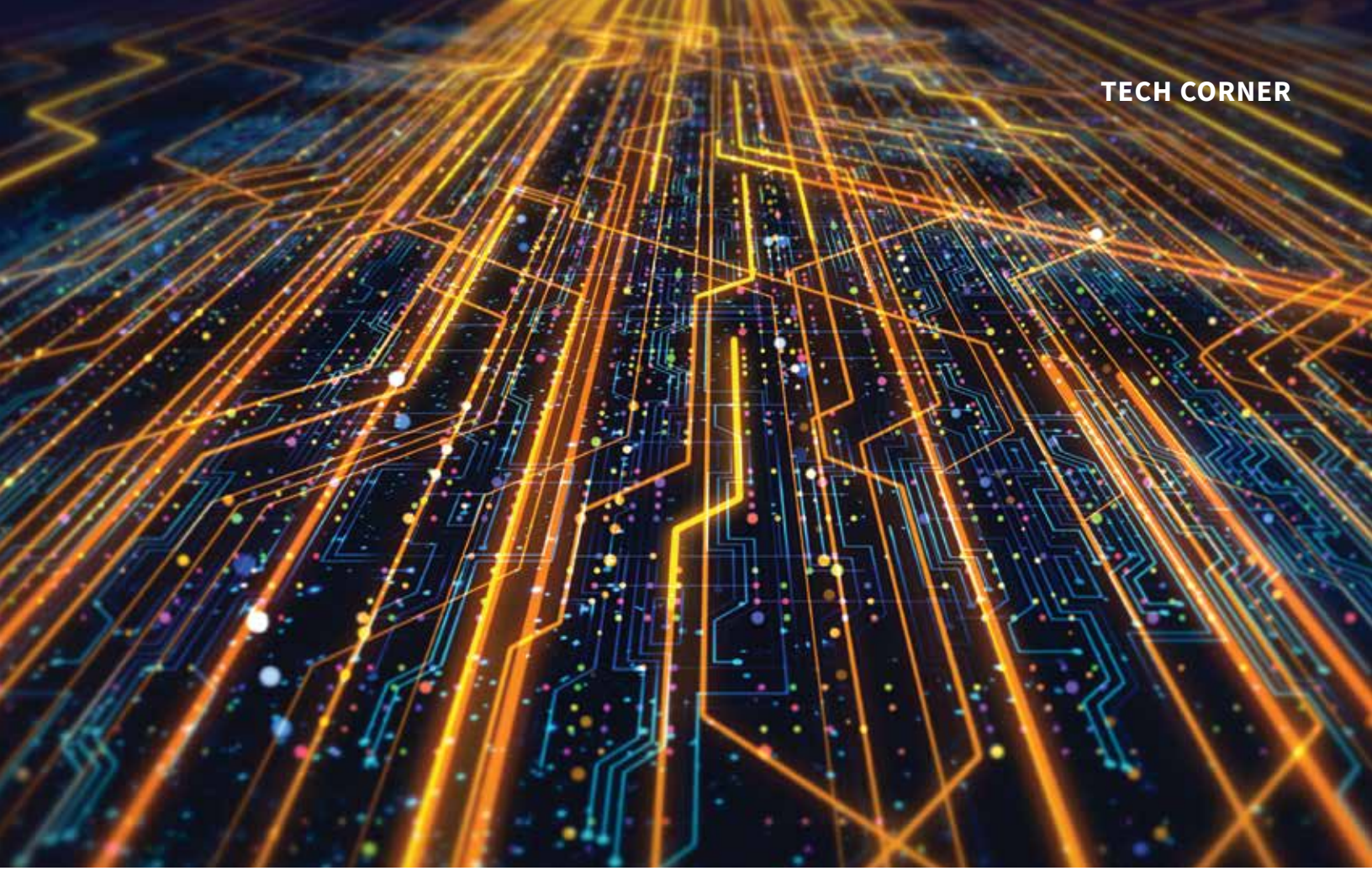

Increase your Tech IQ (Part Two)

By David Winters and Andy Foreman

Technology is constantly changing, and technology and litigation professionals use terms and jargon that can be confusing. In addition, even those of us who are comfortable with technology can, from time to time, use terms without knowing what they really mean. The goal of this article is to provide short definitions of some technology terms to help increase your technology IQ and understanding.

This article is the second installment of a two-part series. The first installment focused on both basic and advanced technology terms as well as terms used in the insurance and reinsurance industry. This installment focuses on technology terms that are often used in insurance and reinsurance litigation, with a particular focus on terms used in e-discovery and arbitrations.

Electronic Discovery in Litigation

Electronic discovery: Electronic discovery or (e-discovery) is simply the electronic form of traditional discovery in litigation. As with traditional discovery, parties engaged in electronic discovery preserve, gather, review, and produce information in response to discovery requests, but with electronic discovery, the information is in electronic format and all of those steps take place electronically. For example, a party might copy all emails from a certain person's email account (from a custodian, defined below), then review and produce the emails in their original format or as image files.

ESI: ESI is an acronym for electronically stored information, a term of art defined in the Federal Rules of Civil Procedure. Broadly speaking, ESI is information created, manipulated, transmitted or stored in an electronic medium or in a manner otherwise requiring the use of computer hardware or software. Everything saved on your computer is ESI. So is everything on your smartphone. The vast majority of documents produced or exchanged in modern arbitration will be ESI.

Gigabyte and megabyte: A byte is a term describing the quantity of electronic information expressed as a unit of memory size. A gigabyte is roughly one billion bytes of information; a megabyte is roughly one million bytes of information. Because ESI cannot always be measured in terms of pages of information (think videos or databases), parties in litigation sometimes use terms like gigabyte or megabyte to quantify electronic information. Even where ESI can be measured in pages, the number of pages of information contained with a gigabyte or megabyte will vary depending on the types of files in question. For example, a gigabyte could contains anywhere from 15,000 pages to more than 670,000 pages, depending on the file type.

Metadata: Metadata is data about data; it is the information that describes the characteristics of ESI, such as sender, recipient, author, the date a document was created, and the date a document was modified. An electronic document's metadata is stored within the document itself. For example, a Microsoft Word document may contain metadata such as who created it, when it was created, and when it was edited. If someone were to change the text of that Word document, even though it might not be evident later from the text that a change was made, the metadata will show that it was changed and when.

Where Files Are Found

Custodian: The custodian is most often the individual from whose files a group of records were collected or extracted. This person is not necessarily the author of the documents. For example, you're the custodian of all of the emails in your own email account, even though you received the ones in your inbox and only authored the ones in your sent folder.

The cloud: The cloud is a network of offsite electronic storage locations for holding ESI. The computers and servers that make up the cloud are networked and accessible through the internet. Information stored in the cloud is generally saved in more than one location to ensure that if one copy is lost—for example, due to malware or flooding—other copies remain.

Archive: A separate storage device or location for long-term preservation of ESI that is no longer actively used. This can also refer to the process of moving ESI to a long-term storage device or location.

Backup: A backup is a copy of one or more files created as an alternate in a different storage location—on a disk, a drive, or, increasingly these days, in the cloud—in case the original data is lost or becomes unusable.

How Files Are Organized

Document family: All parts of a group of files that are attached or connected to each other, such as an email and its attachments. For example, an email and its attachments together would be collectively called the "document family."

Parent document: Within a document family, the file to which other files are attached. In the case of emails, the email itself is the parent.

Child document: Within a document family, a file that is attached to another file. In the case of emails, an attachment to an email is a child.

Attachment: A file that is connected to another file either externally, such as a file connected to an email so as to send the file with the email, or a file embedded in another file, such as an image in a word processing document.

File Processing and Review

OCR: Optical character recognition (OCR) is a process by which a computer analyzes images of text, such as scans of printed pages, to convert the images into electronic text that can be electronically searched using search queries. OCR does not always work perfectly—for example, handwriting can be particularly difficult if not impossible to convert into searchable text. But since OCR is dependent on current technology and the bounds of artificial intelligence, it is getting better all the time.

Search: The process of looking within a data set for data that matches, or is related to, specified criteria (a query). Searches can be as simple as a keyword search, where the search looks for a specified word, or as sophisticated as a concept search, where the search identifies documents based on a query but where the query term itself is not in the document.

De-duplication: Also referred to as de-duping, de-duplication is the process of comparing electronic documents or files to identify and/or remove duplicate records. For example, if emails from multiple custodians who exchanged emails with each other are collected, de-duplication can remove copies of the same email that appear across multiple different email files—in the sent folder of one and the inboxes of others.

Predictive coding: Also known as "technology assisted review" (TAR) or "computer assisted review" (CAR), predictive coding is a tool involving the use of computer software that is used in litigation to manage the review of large quantities of ESI. Predictive coding is a process of categorizing documents by extrapolating the tagging decisions of a reviewer across a larger data set, using artificial intelligence. It is an iterative tool that increases accuracy throughout the reviewing and tagging process. Predictive coding can be used as a more economical alternative to having attorneys review each and every document collected for production.

Redact: To intentionally conceal or censure, usually via an obscuring overlay or by removing data or portions of a document considered privileged, proprietary, confidential or otherwise objectionable. Redaction of ESI can be somewhat more complicated than with paper documents—it is not always enough simply to hide the text being redacted, as files can contain text that is not visible but that can nevertheless be discovered.

Producing Documents in Litigation

Parties in litigation produce documents in a variety of different ways. To illustrate, imagine a Microsoft Word document. The following terms illustrate different ways this same Microsoft Word document could be produced in litigation.

Native format: An electronic file that is in the same format in which it was created. A native format file can contain metadata about its creation that might be lost if the file is converted to an image format for production, although the metadata can be separately recorded and produced with the image file and production of a native file can alter its metadata. In the case of our Word document example, an electronic copy of the original Microsoft Word document file is in native format.

PDF: Portable document format (PDF) is a file format consisting of either an electronic image of text (or text and graphics) or searchable text, when the PDF is output directly from a native document. While a PDF can be created from a document in native format, a PDF file can also be created by scanning a hard-copy document, which results in a PDF that contains an image of the document. For our Word document example, it would be possible either to "print" the Word document "to PDF," resulting in a file with searchable text, or print the Word document on paper and then scan it into a PDF.

TIFF: Tagged image file format (TIFF) is a type of image file that can be created directly from a file in native format or by scanning a hard-copy document. Unlike a PDF, even a TIFF created directly from a native document will not contain searchable text. Parties in discovery sometimes exchange ESI in the form of single-page TIFFs, where each TIFF is an image of a single page. As with PDFs, our example Word document could be converted to a TIFF directly or printed and scanned to become a TIFF.

Legacy data: Data for which the format has become obsolete. This can include file formats for programs that have been discontinued—think of a document created with outmoded word processing software, or a spreadsheet created using Lotus-1-2-3—as well as media in formats that are no longer used, such as floppy disks and tape backups.

The definitions used in this article were derived from a variety of sources, including the following: The Federal Rules of Civil Procedure; Legal Dictionary at Law.com, available at https://dictionary.law.com; "How Many Pages in a Gigabyte?" available at https://www.lexisnexis.com/applieddiscovery/lawlibrary/whitePapers/ADI_FS_PagesInAGigabyte.pdf; "The Tech Terms Computer Dictionary," available at https://techterms.com; and "52 eDiscovery Terms You Should Know," available at https://cdslegal.com/wp-content/uploads/2012/07/52-ediscovery-terms-you-should-know.pdf.

David Winters and **Andy Foreman** are partners at Porter Wright Morris & Arthur LLP. They are trial lawyers who represent insurance and reinsurance companies. They have had extensive experience with the technological issues that arise in litigation and arbitration.