Professional Perspective

Metadata Issues in Discovery

Robert Dibert, Connie Wilkinson-Tobbe, & Lindsay Graves, Frost Brown Todd

Bloomberg Law

Metadata Issues in Discovery

Contributed by Robert Dibert, Connie Wilkinson-Tobbe, & Lindsay Graves, Frost Brown Todd

Requests for production of metadata for electronically stored information (ESI) are commonplace in discovery today. Metadata discovery has evolved from a document-centric means of authenticating or impeaching authors or recipients, to pondering the mass of data beneath the surface of a file in an ocean of information. The potential relevance of metadata, and costs associated with its appropriate preservation, collection, and production, are continuing challenges in ESI's evolution.

Described broadly, metadata is, "data which gives information about other data," *United States v. Wehrle*, 985 F.3d 549, 554 (7th Cir. 2021). This article discusses the evolution of three categories of metadata, then some practical considerations for anticipating and addressing metadata issues in discovery.

Metadata Categories

File Metadata

File metadata—metadata embedded within electronic word processing and other user-created files ("documents")—have been subjects of discovery for decades. Information embedded in a computer file may include a file creation date, when it was last accessed or modified, a stated author, and perhaps previous versions or changes. This information is not usually displayed on a screen or print of the file. File metadata has been described as "not obvious to non-computer professionals." *Continental Group, Inc. v. Kw Property Management*, 622 F. Supp. 2d 1357, 1373 (S.D. Fla. 2009). It may be created, altered, or deleted without the knowledge of the user.

File metadata can be used to authenticate (or impeach) the stated date(s) or time(s) that a document was created, modified, or accessed. It has been used in court proceedings to establish that an individual received certain information during employment. It also has been used to show that ostensibly different documents originated from the same source.

Confirming whether a document was backdated to appear that it was created at an earlier time is another common usage. In one example, metadata analysis showed that recordings had not been changed between the time they were created and the time they were shared with defense counsel. *United States v. Woods*, 978 F.3d 554, 559 (8th Cir. 2020). In another, metadata showed that a retainer agreement "purportedly signed" on one date was not created until a later date. *SPV-LS, LLC v. Transamerica Life Ins. Co.*, 912 F.3d 1106, 1114 (8th Cir. 2019)

Increasing varieties in methods of communication and data storage have increased the potential scope and content of file metadata. In another example, metadata was determined to be "underlying data that is attached to a file that provides identifying information about the file," including such information as when the data file was produced and last modified, the type of device used to create it, and the program or firmware used to create the file. *United States v. Hager*, 710 F.3d 830, 832 n.2 (8th Cir. 2013).*United States v. Post*, 997 F. Supp. 2d 602, 603 (S.D. Tex. 2014).

Some types of files also may store certain system information within the file metadata, such as geolocation information that can pinpoint where a picture or other data-creating action occurred. See *United States v. Martinez*, 9 F.4th 24, 40 (1st Cir. 2021). In this case, the court focused on a video that was, according to the metadata, taken with a phone at a rest stop. "In most cases, this

information is automatically embedded in digital pictures unless the user opts out of the features that capture the information." *United States v. Post*, 997 F. Supp. 2d 602, 603 (S.D. Tex. 2014).

Although file metadata is not usually displayed on a screen or print, it is accessible to change in a user's discretion or other individual activity. Users may alter some metadata properties or create additional categories in a document. For example, a user may "provide a title, as well as tags or keywords that describe the video, and can also select pre-set categories describing the video, such as 'music,' 'faith' or 'politics.'" *UMG Recordings, Inc. v. Shelter Capital Partners LLC*, 718 F.3d 1006, 1012 (9th Cir. 2013).

File metadata also may be supplemented or otherwise altered in a business process. In a media case, an editor inserted metadata in a freelance article to generate website traffic. *Miami New Times, LLC Employer and The Newsguild-Communications Workers of America, AFL-CIO*, No. 12-RC-255122 (N.L.R.B. 7/15/2020) Alterations may be intentional, but also coincidental in that the mere act of copying an electronic file or forwarding an email may alter some of the related metadata.

System Metadata

File and system metadata have overlapped in the evolution of ESI. For this article, "system" metadata refers to information about an electronic document that is located on a computer device or in another location outside the document itself. See Nat'l Inst. of Standards and Technology, *Active File Identification & Deleted File Recovery Tool Specification* at 4 (March 24, 2009).

System metadata about an electronic document could reside within a device file/folder directory structure. It could be contained within a backup or temporary copy of the document or a separate database or other compilation that describes the document. System metadata could even exist in unallocated space on a device that used to contain a document that was "deleted," intentionally or otherwise. In this case, forensic analysis used to identify files that had been deleted using anti-forensic software. *Southern New England Tel. Co. v. Global Naps Inc.*, 624 F.3d 123, 143 (2d Cir. 2010) The "system" may be a specific device or shift between multiple systems such as internet- and non-internet-facing devices.

Beyond individual document metadata, system metadata also may contain information about the creation or usage of a file that is relevant to the matter. In one case, "… the computer stores unseen information about any given 'file'—not only metadata about when the file was created or who created it … but also prior versions or edits that may still exist 'in the document or associated temporary files on [the] disk'—further interspersing the data corresponding to that "file" across the physical storage medium." *United States v. Ganias*, 824 F.3d 199, 213 (2d Cir. 2016)

Court decisions have identified types of information such as geolocation, system updates, and privacy/security or other settings as relevant to issues in specific proceedings. For example, CSLI (cell site location information) is a type of metadata generated every time a cell phone connects to the nearest antenna. The cellular service provider retains a time-stamped record of that antenna. "CSLI can provide a detailed log of an individual's movements over a period of time." *United States v. Goldstein*, 914 F.3d 200, 202 (3d Cir. 2019)

In *United States v. Brown*, 826 F.3d 51, 53-54 (2d Cir. 2016), "examining metadata from one of the images, investigators examined an image's metadata and determined it had been taken using a Motorola Droid X cell phone. The metadata also revealed GPS coordinates associated with the image."

Such information could be relevant to determining issues such as jurisdiction, as well as actions or other behavior of individuals relating to the merits of the matter. For example, metadata from a user device or a server to which it had connected for data communication could show whether specific actions were taken within or directed at a particular jurisdiction.

Remote Metadata

For this article, remote metadata refers to file or system metadata that is created, modified, or otherwise stored or used within a hosted environment beyond the direct control of an individual user or related organization. For example, metadata could be compiled from internet subscribers and stored on the devices that host the user applications, but also collected for storage and possible use elsewhere. In one system the metadata may include "information related to the user, the data, the application, and the user environment; tracking the movement of the user from the user environment of the web-based computing platform to a second user environment of the web-based computing platform; and dynamically updating the stored metadata." *Leader Techs., Inc. v. Facebook, Inc.*, 678 F.3d 1300, 1302 (Fed. Cir. 2012).

In addition to direct information about the user, device, or connected system(s), remote metadata also could include behavioral or other interpretive information. Information could include search history, social media connections, or other information for a longer period than stored on a user's local device. Remote metadata also could include interpretations of such data, and use it to create new metadata for advertising, sentiment, or other analysis.

Recent research and development of biometric applications have sparked litigation over the scope of such activities. In *Vance v. Microsoft Corp.*, 525 F. Supp. 3d 1287, 1291 (W.D. Wash. 2021), "Those with the dataset, and the corresponding information, could 'identify the Flickr user who uploaded the photograph,' 'view the Flickr user's homepage,' and 'view each photograph's metadata, including any available [information] relating to where the photograph was taken or uploaded.'". Current attention to defending against cybercrime also may include the exploration of metadata that accompanies even encrypted internet data transmissions.

Beyond creation or usage, remote metadata may be edited or deleted as a matter of contract or discretion of the remote host. In one example, "[p]hotographs uploaded for use in advertisements are shorn of their metadata, thus removing from scrutiny information such as the date, time, and location the photograph was taken." *Doe v. Backpage.com*, LLC, 817 F.3d 12, 16-17 (1st Cir. 2016).

Finally, the dynamic and distributed nature of remote hosting could undercut the utility of what might have been file or system metadata. In 2017, a service provider responded to Department of Justice (DOJ) search warrants by stating in part that it could not identify, at any point in time, the exact location of email stored on its servers. Precise geolocation was stated to be infeasible because the information systems moved user data automatically—including movement across national borders—as a matter of efficiency. *In Re Search Warrant No. 16-960-M-1 To Google*, MJ No. 16-960 (E.D. Pa. 8/17/2017). "In addition, for some types of data—for example, a Word document attached to an email—the network breaks individual user files into component parts, or 'shards,' and stores the shards in different network locations in different countries at the same time."

Metadata Issues in Discovery

Proportionality in Metadata Discovery

Metadata certainly has been within the scope of "nonprivileged matter that is relevant to any parties' claim or defense and proportional to the needs of the case." Fed. R. Civ. P. 26(b)(1). Indeed, the 2006 Advisory Committee Notes specifically stated that "Rule 34(a)(1) is intended to be broad enough to cover all current types of computer-based information, and flexible enough to encompass future changes and developments." Parties therefore can, and do, request metadata pursuant to Rule 34(a).

On the other hand, "whether material that falls within this term should be produced, and in what form, are separate questions that must be addressed under Rules 26(b), 26(c), and 34(b)." Similarly, the "addition of testing and sampling to Rule 34(a) with regard to documents and electronically stored information is not meant to create a routine right of direct access to a parties' electronic information system, although such access might be justified in some circumstances. Courts should guard against undue intrusiveness resulting from inspecting or testing such systems."

Metadata discovery may be unnecessary, secondary, or essential. Court decisions correspondingly have denied metadata discovery, permitted it, or permitted discovery subject to cost-shifting or other conditions.

Process in Metadata Discovery

The fact-specific nature of metadata and its relevance suggests that the subject might well be considered as early as the legal hold step of a potential proceeding. When a dispute reaches the point of "reasonable anticipation of litigation," it may be appropriate to consider whether metadata could be relevant to proving timing, jurisdiction or other issues in the matter.

Consideration of individuals and issues potentially relevant to the matter also may guide the consideration of relevant metadata types and locations. For matters potentially headed into federal court or other jurisdictions that have adopted discovery provisions of the federal rules of civil procedure, initial disclosure provisions of Fed. R. Civ. P. 26(a)(1)(A), -(3) also may help assess the potential significance of particular metadata. In *Niemi v. Burgess*, 874 F. Supp. 2d 1048 (D. Colo. 2012), file metadata was noted as used to identify authors and associated entities for purposes of injunctive relief.

The fragility of metadata may be a consideration as to process as well as proportionality. As noted, the act of copying an electronic document from one location to another may alter some of its properties. See, e.g., *CBT Flint Partners, LLC v. Return Path, Inc.*, 737 F.3d 1320, 1329 (Fed. Cir. 2013), where "the mere act of copying a file may destroy certain types of metadata"), and *Pitney Bowes Gov't Solutions, Inc. v. United States*, No. 10-257C (Fed. Cl. 8/19/2010), where saving an

electronic document to another location may alter metadata without changing substantive content. Transitory change, particularly if made in the usual course of business, therefore may eliminate the evidentiary significance of some metadata. See *United States v. Chavez*, 985 F.3d 1234, 1239 n.3 (10th Cir. 2021), where Chavez "failed to establish the authenticity and reliability of the metadata".

On the other hand, ESI may be self-authenticating if metadata shows that it was generated in the usual course of a technological process and not altered in a meaningful way. In *United States v. Hunt*, 534 F. Supp. 3d 233, 255 (E.D.N.Y. 2021) metadata showed times or dates of posting or transmission, or IP addresses–"those sorts of records would be self-authenticating."

Metadata might be deemed potentially relevant, but require a special collection or production process. In such an instance, a process outside the usual course of business or other reasonably accessible means could warrant cost-shifting as a matter of Rule 26 proportionality.

For example, one court denied a request for supplemental email production because metadata had not been included in the original request for production, there had been no showing that additional material facts would likely be found within the metadata, and the volume of emails produced was not so large that metadata was necessary to manage the production.

The court did require supplemental production of certain non-email documents with metadata, but required the moving party to bear the costs because the documents already had been produced in a searchable form. The moving party again failed to show the significance of the requested metadata, so the additional information seemed to be "at best, marginally relevant." *Aguilar v. Immigration and Customs Enf. Div. of U.S. Dep't of Homeland Sec.*, 255 F.R.D. 350, 360-362 (S.D.N.Y. 2008).

When a matter has been resolved, the prevailing party may recover certain "costs" of the proceeding beyond cost-shifting that might have occurred during discovery. 28 U.S.C. §1920; Fed. R. Civ. P. 54(d)(1). "To the extent that a party is obligated to produce (or obligated to accept) electronic documents in a particular format or with particular characteristics intact (such as metadata), the costs to make duplicates in such a format or with such characteristics preserved are recoverable as 'the costs of making copies ... necessarily obtained for use in the case.'" *CBT Flint Partners, LLC v. Return Path, Inc.,* 737 F.3d 1320, 1328 (Fed. Cir. 2013).

See also, *Race Tires America, Inc. v. Hoosier Racing Tire Corp.*, 674 F.3d 158, 171 n. 11 (3d Cir. 2012), where "costs of conversion to an agreed-upon production format are taxable as the functional equivalent of 'making copies.' It is all the other activity, such as searching, culling, and deduplication, that are not taxable." One court observed that recovery of costs for "actual electronic copies must be limited to data ingestion or extraction as a substitute for physical copying." *Absolute Activist Value Master Fund Ltd. v. De*vine, No: 2:15-cv-328, p. 22 (M.D. Fla. 2019).

In sum, cost recovery for metadata and other ESI discovery may approximate a three-tiered framework: copying electronic documents as a whole from their source (such as a computer hard drive), organizing the electronic documents into a database and analyzing them for discoverability, and producing the electronic documents on some media or other method of delivery. "The Federal Circuit determined … that activities falling within the second category are most likely not taxable as costs, whereas activities within the first category are mostly taxable and activities within the third category are taxable." *United States ex rel. Saldivar v. Fresenius Med. Care Holdings, Inc.*, 291 F. Supp. 3d 1345, 1352 (N.D. Ga. 2017).

Thus, a starting and perhaps defining question for shifting or recovering costs for metadata discovery may be whether a process associated with metadata discovery is undertaken voluntarily, or in response to a demand for additional discovery. If a party has collected ESI in bulk and is administering the matter through a process which inherently may include metadata discovery, then cost-shifting or recovery may not be available. If specific additional collection, processing, and production are deemed necessary in the context of the matter, then those additional costs may be recoverable either as a matter of Rule 26 proportionality or Rule 54 cost recovery.

Templates for Metadata Discovery

Metadata discovery is dependent upon the facts and potential evidence at issue in each case. Federal agencies have incorporated metadata discovery into their procedures for both litigation and administrative proceedings. For example,

the Federal Trade Commission (FTC) template, Preparing Native Files, includes metadata elements that the FTC may consider potentially relevant to its handling of unfair or deceptive trade practices:

Document Info/Metadata	Description	Concordance Field Name
Beginning Bates Number	The beginning bates number for the document	BEGBATES
Ending Bates number	The ending bates number for the document	ENDBATES
Page Count	The total number of pages in the document	PGCOUNT
Custodian	The name of the original custodian of the file	CUSTODIAN
Creation Date	The date attachment was saved at the location on the electronic media for the first time	CREATEDATE
Creation Time	The time the attachment was saved at the location on the electronic media for the first time	CREATETIME
Modified Date	The date/time the attachment was last changed, and then saved	MODDATE
Modified Time	The time the attachment was last changed, and then saved	MODTIME
Last Accessed Date	The time the attachment was last opened, scanned, or even "touched" by a user or software activity	LASTACCDATE
Last Accessed Time	The time the attachment was last opened, scanned, or even "touched" by a user or software activity	LASTACCTIME
Size	The amount of space the file takes up on the electronic media. Usually recorded in kilobytes, however may be reported in single bytes	FILESIZE
File Name	The name of the attachment including the extension denoting the application in which the file was created	FILENAME
Originating Path	File path of the file as it resided in its original environment.	FILEPATH
Production Link	Relative path to production media of submitted native files. Example: FTC-001\NATIVE\001\FTC-0003090.xls	NATIVELINK
Hash	The SHA (Secure Hash Algorithm) or MD5 Hash for the original native file if available	HASH

These categories include what could be self-authenticating metadata of individual documents as they were collected, and also the folder context in which they were located at the time of collection.

Similarly, parties also often agree to produce certain metadata as part of an ESI protocol or order entered by the court at the beginning of a proceeding. This may include common metadata used for authentication, such as: [Custodian, Date Sent, Time Sent, Date Received, Time Received, Filename, Author, Date Created, Date Modified…]

Authentication Issues in Metadata Discovery

Whatever the nature or purpose, metadata as well as other items of evidence must be authenticated as the first step to admissibility. Federal Rules of Evidence 901(a) states generally that "To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is." Rule 901(b) provides 10 example methods of authentication, including personal knowledge by a testifying witness, and association with a reliable pattern or practice of behavior. Rule 901(b)(10) concludes by allowing authentication by any other "method of authentication or identification allowed by a federal statute or a rule prescribed by the Supreme Court."

Rule 902 then states that certain "items of evidence are self-authenticating; they require no extrinsic evidence of authenticity in order to be admitted." In 2017, the Rule was amended to recognize self-authentication for certain types or processes of electronic records, including a "record generated by an electronic process or system that produces an accurate result" or "Data copied from an electronic device, storage medium, or file, if authenticated by a process of digital identification[.] Rule 902(13-14). In both situations, the authentication must be "shown by a certification of a qualified person that complies with the certification requirements of Rule 902(11) or (12). The proponent must also meet the notice requirements of Rule 902(11)." Rule 902(11) certification, in turn, is based upon proof that the evidence has resulted from a regularly conducted activity, such as usual course of business recordkeeping. One example of such recordkeeping "is typically a comparison of metadata hash values." *State v. Sassarini*, 452 P.3d 457, 471 (Or. App. 2019).

However, self-authentication through certification usually is limited to situations where the metadata is offered to prove the existence -- not content -- of an electronic record. "[S]uch a certification serves a limited role: it simply shows that a record was made at or near a certain time, that the record was kept in the course of a regularly conducted business activity, and that the making of the record was a regular practice of that activity." *United States v. Hunt*, 534 F. Supp. 3d 233, 254 (E.D.N.Y. 2021). To certify substantive content, the proof must include a showing that the accuracy of content is validated as part of the regularly conducted activity. See *U.S. v. Browne*, 834 F.3d 403, 410 (3d Cir. 2016) ("reliability of business records is said variously to be supplied by systematic checking, by regularity and continuity which produce habits of precision, by actual experience of business in relying upon them, or by a duty to make an accurate record as part of a continuing job or occupation") (citation omitted). If the certification does not contain that level of detail, then authentication must be made through Rule 901. See also, *United States v. Farrad*, 895 F.3d 859 (6th Cir. 2018) (holding that the District Court erred in allowing self-authentication of social media photos, but the error was harmless because enough supporting evidence had been provided to authenticate them under Rule 901(a)).

Assessing the sufficiency of authentication, courts have identified metadata elements that are not enough, in themselves, to prove a record. In addition to potential date/time instabilities discussed previously in this article, the Court in *Arconic Corp. v. Novelis Inc.* 2022 WL 409488, *5 (W.D. Pa. 2/10/2022), stated that "The author field of the metadata is not evidence of the actual preparer of a document ... When a document is created in Microsoft Word, the user name entered on the *File* > *Options* > *General* tab is automatically added as the author in the metadata ... The "author" field of the metadata is not authenticated[.]"

Authentication therefore can be a first -- but not conclusive -- step in the admissibility of metadata or any other evidence. Even where "photos and their date and time stamps are deemed authentic, [a party] may still elicit and present evidence regarding metadata reliability and accuracy." *Tamares Las Vegas Props. v. Travelers Indem. Co.*, 117 Fed. R. Evid. Serv. 1582, 2022 WL 476093, *5 (D. Nev. Feb. 16, 2022).

Takeaways

The long-expanding scope, variability and use of metadata suggest that its discovery is a case-by-case undertaking. Indeed, significant decisions affecting the value and use of metadata in proceedings might be made by implication during basic recordkeeping for a business operation. When a situation has reached the point of an actual or reasonably anticipated proceeding, some questions to consider about metadata might include:

- Does metadata matter to any issues of authentication, timing, or merits of legal issues in the proceeding?
- Do any service provider agreements allow separate creation of metadata, or direct or derivative use of metadata that might be relevant to issues in the proceeding?
- Does the proceeding involve sequence or timing of activity related to any electronic communications or devices?
- Does the proceeding involve any particular types of information that are based upon use of custom applications?
- What costs or other issues might be involved in the identification and preservation of metadata for the proceeding?