IN RE: UBER TECHNOLOGIES, INC., PASSENGER SEXUAL ASSAULT LITIGATION

This Order Relates to: ALL ACTIONS

Case No. 23-md-03084-CRB (LJC)

United States District Court, N.D. California

Filed April 23, 2024

Cisneros, Lisa J., United States Magistrate Judge

ORDER RESOLVING OUTSTANDING ESI PROTOCOL DISPUTES Re: Dkt. 499

The Court previously resolved most of the parties' disputes as to their competing proposed protocols for electronically stored information (ESI) in Pretrial Order (PTO) No. 9. Dkt. 345. The Court provided instructions to the parties for further investigations and meet-and-confer regarding the cloud-stored documents issues, metadata fields, and related provisions of the ESI protocol. Id. at 26.[1] The parties now submit a Joint Discovery Letter regarding their remaining disagreements, which the Court addresses herein.[2] Dkt. 499.

Relevant to the current dispute concerning the ESI protocol in this case, Uber uses Google Workspace. See Decl. of William Anderson ¶ 3 (dkt. 499-4). Google Workspace provides Uber with a suite of cloud-based web applications, including Gmail, Google Chat, Google Drive, and Google Vault. Decl. of Jamie Brown ¶ 4 (dkt. 499-6). Uber utilizes Google Vault as an information governance and e-discovery tool for its Google Workspace data. Anderson Decl. ¶ 14. Through Google Vault, Uber retains and holds, among other things, users' Gmail messages and Google Drive files. Id. Google Vault can act as a clearinghouse to process data for discovery purposes. Decl. of Philip Favro ¶ 18 (dkt. 262-8). Uber's standard discovery process involves exporting a custodian's Gmail messages and Google Drive files from Google Vault. Anderson Decl. ¶ 14. Data not stored using Google Vault is sometimes referred to as "active data". See Brown Decl. ¶¶ 26, 30; see also Joint Discovery Letter at 2 (Plaintiffs distinguishing between "Active Google Email" versus "Google Vault Email," and "Google Drive Documents" versus "Google Vault Documents.").

Of principle concern here, a Gmail or Google Chat message that contains a hyperlink to a document is referencing a Google Drive document that may still be evolving. A recipient or others may modify that referenced document, which is centrally located so multiple people can access and edit it. Furthermore, Google Vault does not export, collect, or connect the contemporaneous versions of hyperlinked documents with the corresponding emails or messages in which they are found. Rather, when a hyperlinked Google Drive document is exported from Google Vault, the current version of that document is exported. If a Google Drive document archived using Google Vault was edited after the email with the hyperlink to the document was sent, then the Google Vault export will not reflect the version of the document that existed at the time of the email. For data archived using Google Vault, and no longer in the active Google Workspace, there is a manual process in place to identify a historic version of a hyperlinked Google Drive document contemporaneous with the email communication. Anderson Decl. ¶¶ 17– 45.[3]

Certain technologies have been developed to link email and chat messages to Google Drive documents, but there are limitations. Metaspike's Forensic Email Collector (FEC) program can retrieve active Google Email and contemporaneous versions of linked Google Drive documents, but it does not have the ability to do the same with Google Email and Drive documents archived using Google Vault. Supplemental Decl. of Favro ¶ 13 (dkt. 499-5). Uber's e-discovery vendor, Lighthouse, has developed a tool, Google Parser, that extracts specific links to Google Drive documents from email and chat messages and certain metadata. Brown Decl. ¶ 20. Google Parser facilitates the grouping together of a message and document stored in Google Drive for purposes of review and production, and it contains certain metadata fields relevant to search, review, and production of messages. Id. ¶¶ 22–23. However, there is no evidence that this technology, which is an extraction tool, has been refined and deployed to collect contemporaneous versions of hyperlinked documents archived with Google Vault.

In sum, the briefing and evidence, as well as related case law, have made clear that cloud computing and document retention through Google Drive and Google Vault introduce a host of challenges to producing hyperlinked documents from Google Drive and other sources. See Nichols v. Noom Inc., No. 20CV3677LGSKHP, 2021 WL 948646, at *2 (S.D.N.Y. Mar. 11, 2021) (recognizing that "complex questions about what constitutes reasonable search and collection methods" result from "the changing nature of how documents are stored and should be collected."). Yet, contemporaneous versions of hyperlinked documents can support an inference regarding "who knew what, when." An email message with a hyperlinked document may reflect a logical single communication of information at a specific point in time, even if the hyperlinked document is later edited. Thus, important evidence bearing on claims and defenses may be at stake, but the ESI containing that evidence is not readily available for production in the same manner that traditional email attachments could be produced. See Favro Decl. ¶¶ 13–15 (describing the challenges associated with collecting specific iterations of hyperlinked documents compared with traditional email attachments).

I. CLOUD-STORED DOCUMENTS AND RELATED HYPERLINKS

Uber's Proposal - Uber's Proposed Section 17:

Cloud Stored Documents Uber will make reasonable and proportionate efforts to preserve the metadata relationship between email messages with links to files on Google Drive, to the extent Uber's vendor for processing and managing the documents to be reviewed and produced in this action possesses technology that enables it to maintain such a relationship. Defendants may use Lighthouse's "Google Parser" for this purpose. Notwithstanding that Uber agrees to make reasonable and proportionate efforts in this regard, because of technological limitations inherent in the processing of emails containing embedded links, it shall not be presumed that all emails containing links to files on Google Drive will be produced with a metadata relationship between the parent email and the linked document. To the extent the Receiving Party believes that there is a lack of a metadata relationship between a specific email message and a specific linked document, the Receiving Party may notify the Producing Party and request that the particular linked file be extracted and produced or identified. To the extent that the linked file in question is nonprivileged, and is relevant to either Party's claims or defenses and the efforts required to search for it would be proportional to the needs of the case, the Producing Party shall use reasonable and proportionate efforts to collect and produce/identify the document that was linked in the original email. The Parties agree to meet and confer to resolve any disagreements as to what constitutes reasonable and proportional discovery efforts.

Plaintiffs' Proposal - Plaintiffs' Proposed Section 18(a)-(c): Cloud Stored Documents

a) Metadata Preserved. Uber shall preserve the metadata relationship between email messages with links to files on Google Drive. Uber shall preserve and produce (including, if necessary, as custom fields) all metadata collected with respect to all cloud-stored documents. That includes, but is not limited to, all metadata output by Google Vault when exporting a matter. Thus, the metadata exported from Google Vault pertaining to each document shall be preserved and produced as metadata for the same document within the load file of any production containing any such document.

b) Hyperlinked/URL-Referenced Documents. Producing party shall make all reasonable efforts to maintain and preserve the relationship between any message or email and any cloud-hosted document hyperlinked or referenced within the message or email. Thus, for instance, where a collected email links to or references by URL a document on Google Drive (or housed within Google vault,) the metadata for that message or email shall include the URLs and Google Document ID of all hyperlinked documents, and if a hyperlinked document was produced.

c) Contemporaneous Versions of Hyperlinked/URL-Referenced Documents. Uber shall produce the contemporaneous document version, i.e., the document version likely present at the time an email or message was sent, of Google Drive documents referenced by URL or hyperlinks in emails or messages such as Slack. If Uber contends that it is unable to meet this requirement through commercial or vendor software, Google APIs or through other reasonable manual or automated

> means, then Plaintiffs and Defendants shall meet and confer to discuss solutions. This will not exempt Uber from producing the applicable version of any document so referenced by URL or hyperlink.

In PTO No. 9, the Court ordered as follows:

> Uber shall direct an employee with knowledge and expertise regarding Google Vault and Uber's data and information systems to investigate in detail the extent to which Google Vault's API, macro readers, Metaspike's FEC or other programs may be useful to automate, to some extent, the process of collecting the contemporaneous version of the document linked to a Gmail or other communication within Uber's systems, whether the email or communication is stored in Google Vault, or outside. This investigation shall not be limited to documents referenced by URL or hyperlinks in emails or Google documents stored in Google Vault, but shall also include other cloud-based messages such as Slack. Uber's designated employee may consult with Uber's e-discovery experts. Likewise, Plaintiffs shall also more thoroughly investigate these potential solutions.

Dkt. 345 at 21. The Court ordered the parties to meet and confer as to their investigations, and if disagreements remained, and they wanted to submit a discovery letter to resolve their dispute, then the Uber employee designated to conduct its investigation was to "submit a declaration supporting its positions, and Plaintiff's expert(s) and/or e-discovery vendors shall do the same." Id. Uber was also permitted to "submit declarations from its experts and e-discovery vendors." Id. Uber submitted declarations from William Anderson, an Uber employee who specializes in e-discovery; Jamie Brown, a Lighthouse representative; and a supplemental declaration from Philip Favro, Uber's e-discovery expert. Dkt. 499-4, 499-5, 499-6. Plaintiffs submitted a second declaration from Douglas Forrest, their e-discovery expert. Dkt. 499-1.

Uber's position, based on its "exhaustive" investigation, is that "no technical, scalable solution is available" to automate the process of collecting contemporaneous versions of hyperlinked documents. Joint Discovery Letter at 4. Plaintiffs, on the other hand, claim to have identified a process for automation. Plaintiffs submit a Proposed Methodology for Retrieving Google Drive Documents Linked to Within Google Emails (Proposed Methodology). Dtk. 499-2. They have created a Proof-of-Concept program, described in detail in Mr. Forrest's declaration, which supposedly demonstrates that there is a method available to programmatically retrieve contemporaneous versions of linked Google Drive documents. See Forrest Decl. ¶¶ 29–38. The Proposed Methodology provides that Uber is to create such a program based on Plaintiffs' Proof of Concept to produce contemporaneous versions of documents with Google Drive hyperlinks. Dkt. 499-2 at 3.

The Court is not persuaded that Plaintiffs' Proposed Methodology is a reasonably available option here. Mr. Forrest's consulting firm created the Proof-of-Concept program based on a post in Stack Overflow, "a well-known and widely used forum for developers," which contains sample scripts purportedly for retrieving "a date-specific revision of a Google Drive Google native document identified by its Document ID." Forrest Decl. ¶ 33. However, as Mr. Anderson points out, the anonymous internet user who posted the script on which the Proof-of-Concept program is based on admitted that it did not work. Anderson Decl. ¶ 62. And even "a functioning version of the script would not address the issues presented here," in part because the "script was designed for a single document using the Google Drive API, restoring a non-Vault document, with owner access. This script would not work for Google Vault." Id. Uber and its experts also object to many other aspects of the Proof-of-Concept program, which they claim "ignores necessary steps in the collection process and disregards numerous points of manual intervention." Joint Discovery Letter at 5.

Perhaps there is a way to work out these kinks, as Plaintiffs suggest (id. at 3), and the Proof-of-Concept program can eventually be used by some developer as a foundation for creating a program to automate the process of collecting from archived Google Vault data the contemporaneous versions of Google Drive hyperlinked documents. But the Court will not order Uber to expend potentially significant time and resources to develop such a program in order to produce discovery in this MDL, as the program's effectiveness is not assured. See Nichols, 2021 WL 948646, at *4 (rejecting a similar proposal from Mr. Forrest that the defendant's e-discovery programmers could create a program to utilize Google's API to "extract links to Google Drive documents from other Google Drive documents, emails, and Slack communications," in part because Mr. Forrest's declaration did not "address the time it would take to apply the program, load, and review the documents."). The Court is satisfied by Uber's showing in the Joint Discovery Letter that it has thoroughly investigated the issue, and that no technological solution is currently readily available to automate the process when it comes to collecting contemporaneous versions of hyperlinked documents.

The Court has reviewed Uber's evidence regarding the manual process to query Google Vault for contemporaneous versions of a hyperlinked document. See e.g., Favro Supplemental Decl. ¶¶ 6–7. The Court is mindful of the burdens to Uber but also recognizes that Uber has chosen Google Vault as its storage

method. See Shenwick v. Twitter, Case No. 4:16-cv-05314, 2018 WL 5735176, at *1 (N.D. Cal. Sept. 17, 2018) (requiring the defendants to conduct manual searches and produce 200 hyperlinked documents despite the undisputed burdens because they chose their storage method and the plaintiffs are entitled to discovery). In Shenwick, the defendants utilized Google Vault for document review and production. See id., Dkt. 190 at 5. Thus, the potential limitations and pitfalls with respect to production of hyperlinked documents from Google Vault have been widely known for many years, yet Uber has elected to transfer and retain its electronic data using this service.

Given all the above, the ESI protocol shall state the following with respect to Cloud Stored Documents:

> a) Metadata Preserved. Uber shall preserve the metadata relationship between email messages with links to files on Google Drive to the extent feasible with existing technology. Uber shall preserve and produce (including, if necessary, as custom fields) all metadata collected with respect to all cloud-stored documents. That includes, but is not limited to, all metadata output by Google Vault when exporting a matter. Thus, the metadata exported from Google Vault pertaining to each document shall be preserved and produced as metadata for the same document within the load file of any production containing any such document.

> b) Hyperlinked/URL-Referenced Documents. Producing party shall make all reasonable efforts to maintain and preserve the relationship between any message or email and any cloud-hosted document hyperlinked or referenced within the message or email. Thus, for instance, where a collected email links to or references by URL a document on Google Drive (or housed within Google vault,) the metadata for that message or email shall include the URLs and Google Document ID of all hyperlinked documents.

> c) Contemporaneous Versions of Hyperlinked/URL-Referenced Documents. Uber shall produce, to the extent feasible on an automated, scalable basis with existing technology, the contemporaneous document version, i.e., the document version likely present at the time an email or message was sent, of Google Drive documents referenced by URL or hyperlinks therein. For hyperlinked Google Workspace data archived using Google Vault, Uber is not required to produce the contemporaneous document version at the time the email or message was sent, as this is not possible through an automated process with existing technology. However, Plaintiffs may identify up to 200 hyperlinks for which they seek the contemporaneous referenced document even though the email or message has been archived with Google Vault. Uber shall identify and produce the likely contemporaneous versions that Plaintiffs have requested. The scope of this production does not exempt Uber from any obligation that it preserve historic versions or revision history of any document referenced by URL or hyperlink.

The parties may seek to modify the protocol with respect to the production of contemporaneous versions of hyperlinked documents based on the need for relevant discovery and what is proportional to the needs of the case, or what electronic data is not reasonably accessible because of undue burden or cost.[4] By stipulation of the parties, or through Court order, Plaintiffs may request additional contemporaneous documents, and Uber may seek relief from the production of certain versions or other obligations under the ESI protocol based on undue burden or costs, overbreadth or disproportionality.

II. METADATA FIELDS

PTO No. 9 required the parties to meet and confer as to certain disputed metadata fields, but proposed a resolution of the disputes if they were unable to come to an agreement. Dkt. 345 at 25. The parties now indicate that they reached an agreement as to three of the disputed metadata fields. Joint Discovery Letter at 6. The fields still in dispute are LINKGOOGLEDRIVEURLS and the Account field. Id. In PTO No. 9, the Court proposed that the LINKGOOGLEDRIVEURLS and Account fields be excluded. Dkt. 345 at 25.

The Court first addresses the disputed LINKGOOGLEDRIVEURLS metadata field. As Plaintiffs' expert Mr. Forrest explains it, this is a custom created metadata field designed to identify which Google Drive documents, if any for whatever reason, have not been produced. Forrest Decl. ¶ 95. Uber contends that Plaintiffs did not promptly follow up in regard to this metadata field, and only made related inquiries the day before the parties' joint discovery letter was due. Joint Discovery Letter at 6. Uber also responds that the custom creation of the LINKGOOGLEDRIVEURLS would impose a substantial burden. Id.

Plaintiffs now propose an alternate approach instead of creating the LINKGOOGLEDRIVEURLS metadata field. The proposed alternate approach involves adding a metadata field, "Missing Google Drive Attachments," and a second metadata field, "NonContemporaneous." Forrest Decl. ¶ 96. Uber did not present any evidence indicating that the alternate proposal would impose a substantial burden. Nor did Uber state that it was unable to respond to the alternate proposal because Plaintiffs did not raise it in sufficient time. The first new metadata field proposed by Plaintiffs, "Missing Google Drive Attachments," identifies the Google

Drive documents that were not produced by providing links to all linked Google Drive documents that could not be retrieved. Id. The second new metadata field proposed by Plaintiffs identifies the produced Google Drive documents that were not the versions in existence at the time the document was hyperlinked, i.e., the "non-contemporaneous" version of the document that was produced. Id

As noted previously, the Court directed the parties to meet and confer regarding disputed metadata fields, which included the LINKGOOGLEDRIVEURLS field. The Court did not intend to broaden negotiations regarding what metadata fields should be included in the ESI protocol. The Court recognizes, however, that early ESI decisions may need to be modified when new information is learned. "[P]roportionality considerations may need to be re-balanced at later points in the litigation, and…discovery plans may be modified when new information is learned." Nichols, 2021 WL 948646 at *3. In this Order, the Court has declined to require Uber to produce contemporaneous versions of every hyperlinked document.

Plaintiffs' new metadata proposals appear helpful to streamline review of Uber's productions because the metadata will identify which hyperlinked documents are missing from the production and which documents produced are the non-contemporaneous versions. Streamlining review of the production of hyperlinked documents will advance the speedy and less expensive determination of this action. Fed. R. Civ. P. 1. Furthermore, Uber, as a responding party, is bound by its obligations under Rule 26(g) of the Federal Rules of Civil Procedure. An attorney's signature pursuant to Rule 26(g) "certifies that the lawyer has made a reasonable effort to assure that the client has provided all of the information and documents available to him that are responsive to the discovery demand." Fed. R. Civ. P. 26, Notes of Advisory Committee on Rules— 1983 Amendment. Conversely, parties should not demand metadata fields for which they have no practical use, or which do not materially aid in the discovery process. See The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production, 19 Sedona Conf. J. 1, 173 (2018). "[T]he parties should follow the Federal Rules and seek to reach agreement for the production of ESI as 'ordinarily maintained or in a reasonably usable form or forms' that provide the requesting party a functionally adequate ability to access, cull, analyze, search, and display the ESI, as may be appropriate given its nature and the proportional needs of the case." Id. at 175. Given the challenges associated with the production of hyperlinked documents from Google Vault, Plaintiffs' request to add the metadata field "Missing Google Drive Attachments," and a second metadata field, "NonContemporaneous," is granted.

Plaintiffs' request for the Account field without the limitation proposed by Uber (Forrest Decl. ¶ 92) is denied, as Plaintiffs have not explained why knowing the particular email address associated with a custodian's responsive documents will materially advance the development of their case. Plaintiffs will already be informed of the custodian's first and last name, and their primary email address, and may inquire through other discovery methods whether other email addresses were used. Plaintiffs' request is denied without prejudice to a future, more limited metadata request where Plaintiffs demonstrate that such granular data is material to the discovery process.

III. DEFINITIONS

The parties were unable to resolve their dispute over the definition of "attachment." Dkt. 499 at 7. The proposed language for the ESI protocol is as follows:

> Uber's Proposal - Uber's Proposed Definition of "Attachment":
>
> > An "Attachment" is typically a file associated with another file for the purpose of storage, transfer, processing, production, or review. There may be multiple attachments associated with a single "parent" or "master" file. In many records and information management systems, or in a litigation context, the attachments and associated record(s) may be managed and processed as a single unit. In common use, this term most often refers to a file (or files) associated with an individual email or other message type. For the avoidance of doubt, a hyperlinked document, such as a cloud-based document in Google Drive, is not an "attachment."
>
> Plaintiffs' Proposal - Plaintiffs' Proposed Definition of "Attachment"
>
> > "Attachment(s)" shall be interpreted broadly and includes, e.g., traditional email attachments and documents embedded in other documents (e.g., Excel files embedded in PowerPoint files) as well as modern attachments, pointers, internal or non-public documents linked, hyperlinked, stubbed or otherwise pointed to within or as part of other ESI (including but not limited to email, messages,

comments or posts, or other documents).

Plaintiffs' proposal is adopted with the following modification:

> "Attachment(s)" shall be interpreted broadly and includes, e.g., traditional email attachments and documents embedded in other documents (e.g., Excel files embedded in PowerPoint files) as well as modern attachments, pointers, internal or non-public documents linked, hyperlinked, stubbed or otherwise pointed to within or as part of other ESI (including but not limited to email, messages, comments or posts, or other documents). This definition does not obligate Uber to produce the contemporaneous version of Google Drive documents referenced by URL or hyperlinks if no existing technology makes it feasible to do so.

IV. CONCLUSION

Having resolved all the remaining disputes as to the terms of the ESI protocol, the parties are hereby ordered to incorporate the Court's rulings from PTO No. 9 and from this Order into a final Stipulated and [Proposed] ESI Protocol, to be submitted to the Court by April 30, 2024.

IT IS SO ORDERED.

Footnotes

[1]

Unless specified otherwise, the Court refers to the PDF page number generated by the Court's efiling system.

[2]

In this Order, the Court refers to Defendants collectively as "Uber," as the Court and the parties have throughout this litigation.

[3]

The steps in the manual process are described in extensive detail, but there is no information concerning cost or length of time to conduct the process.

[4]

Plaintiffs provide data estimates to flag the large volume of hyperlinked documents that will be produced in non-contemporaneous form. Joint Discovery Letter at 2. The Court is unable to assess how meaningful these estimates are, given that there may be duplication among the hyperlinked documents, and that multiple emails may refer to the same document that is rarely edited over a period of time. The Court has concerns regarding the difficulty of securing the version of a document that is contemporaneous with the email that hyperlinked it. However, the full scope of the problem is not yet defined, and typically only a fraction of documents produced in discovery are material to the litigation.

End of Document.