

The Sedona Conference Draft Commentary on Discovery of Modern Collaboration and Communications Platforms

The Sedona Conference

Drafting Team Members

Stacey Blaustein	Michelle Briggs
Doug Forrest	Adam Gajadharsingh
Hon. Jane Manning	Benson McGrath
Derek McNally	Jonathan Swerdloff
Cristin Traylor	Beth Wilkins
Jonathan Orent	

Team Leaders

Gareth Evans	Joseph Guglielmo
--------------	------------------

Steering Committee Liaisons

Tara Emory	Sandra Metallo-Barragan
Maria Salacuse	

Copyright 2023. All rights reserved.



This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

I. Introduction

II. Discussion

A. Characteristics of Collaboration and Communication Platforms

Modern collaboration and communication applications can be traced back to the advent of email. These applications were served by on-premises email servers (for example, Microsoft Exchange) or cloud providers (for example, AOL, Yahoo Mail, Hotmail, Gmail, and now Microsoft 365). Emails could be replied to and forwarded, becoming threads and conversations. Importantly, emails were static in nature, a snapshot or recording of a communication. Collaboration was supported within those conversations, both within the emails themselves and by attaching files, such as Word documents with tracked changes and comments, a collaborative revision process which is familiar to any lawyers who have attached and emailed successive drafts of documents amongst themselves.

As technology evolved with the introduction of mobile devices such as smartphones, and the rise of cloud-based social media, the number and variety of communication and collaboration platforms expanded rapidly. Messaging and chat applications have proliferated, first with short messaging service messages (SMS) on mobile devices and then expanding into various text messaging and chat applications, which may include features such as end-to-end encryption, chat groups, and ephemeral messaging. Direct messaging features were also added as a common feature of social media and business networking applications.

Programs designed for group collaboration and communication such as Microsoft Teams and Slack have proliferated in business, as have document-focused collaboration programs such as Google Docs and Microsoft OneDrive. These programs offer communication options that are not static, but rather fluid and dynamic, they include hyperlinks and other pointers to attachments as opposed to attaching a version of a document. There are also many document repository focused platforms for storing and sharing documents, including SharePoint and cloud applications such as DropBox. New circumstances such as the COVID pandemic, which increased work from home and remote work generally, opened the door for the widespread adoption of video conferencing applications. Spaces on X (formerly Twitter) provides a platform for live audio conversations. There will inevitably be many new types of platforms that facilitate collaboration and communication in different ways. Perhaps the highly vaunted metaverse will one day reach the lofty goals of its early promoters.

Modern collaboration and communication platforms generally have a number of common characteristics that can be used to characterize and understand them.

Fluid Nature of Collaboration and Communication Platforms: Information generated is usually “fluid” instead of “static,” which complicates the “typical” eDiscovery processes. The classic notion of a “custodian” may not truly exist when it comes to these data types. As with the fluid nature of the contents, the access, controls, and even ownership of a collaboration document can change over time.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Documents can also be shared with groups and not just individuals. Even the metadata can be different. Different communication applications are springing up every day, and it would be fruitless to try to list them all or address their nuances. Consideration regarding preservation, collection, and review must be given when utilizing these platforms, as options may be limited. It is important to first understand what platforms your client or organization may be using, and then explore the technology options. Thus, it is important to understand as many of the technical nuances as possible. Some of the items that should be considered are discussed here.

Ability to have conversations within documents: Platforms such as Google Docs, SharePoint, and OneDrive allow persons to “tag” other persons, make notes and essentially have a “conversation” within documents (word, ppt, spreadsheets, etc.) with full ability to reply and in some instances respond with emojis. Consider how these “modern comments” are processed and reviewed. Some platforms may be able to put the data in separate fields or show the comments in the viewer or extracted text. It will be important to understand where this data is stored, viewed and produced as that will impact privilege, confidentiality, PI and other review considerations. If the modern comments are present in the native, but not in the production images, or vice versa, that could cause many issues for both the receiving and producing parties. If there is text that needs redaction, producing parties will need to know all the locations where that data resides so it can be redacted properly.

Linked documents: It is important to recognize that it is highly likely that discovery will involve linked documents. These links may be in emails or in collaboration platforms such as Teams or Slack. Parties need to understand whether or not the linked attachments can be collected and “connected” back together with the transmittal source (email, message, etc.) in an automated technology-driven process. If so, then they need to understand which version will be connected - the version that existed at the time of the communication or the version that exists at the time of collection (or some other version). They also need to understand the number of versions that may exist and what the criteria is for a “saved version.” For example, Google Docs currently retains up to 40 versions of a document. Sharepoint and OneDrive have different options for versions. Different metadata may be available for different versions.

Family relationships are not straightforward: The standard definition of a “family” in eDiscovery gets turned on its head when dealing with communication and collaboration platforms.

Collection options can vary: There are different preservation and collection options within platforms depending on the version. For example, the Enterprise version of Slack has more options for preservation and collection than the free version. Microsoft’s Purview Premium has more advanced functionality than its Standard offering. The export formats are also tech-specific so it will be important to understand what is being captured and what format it will be in. Some platforms may be exported in proprietary formats that will then have to be converted to a usable format for review.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Different metadata from various platforms: Available metadata will differ from platform to platform. Even the same metadata may have a different named field. In addition, developers are constantly adding new features and capabilities to communication and collaboration platforms, which can lead to new metadata options.

Deduplication is a challenge: Depending on how different versions of documents are saved, parties may end up with hundreds, or even thousands, of versions of the same document, with just minor differences. In some cases, these subtle differences may be important to the outcome of the case, but in most instances, this will just be additional “noise” expanding the costs and burden of discovery. These versions will not be deduplicated because of the slight differences so this should be considered during collection and review. Short messages are also difficult to deduplicate, depending on the processing platform. For example, if you collect phones from multiple persons that were texting each other, you may end up with the same texts from each person.

Inability to connect different platforms when people “jump” from one to the other: Persons frequently communicate across multiple platforms at one time. They may start by emailing and then jump to Teams or Slack, and then take part of the conversation “offline” to text or WhatsApp. However, each of those platforms is collected, processed and reviewed separately, so it is not simple to understand the full context of these types of cross-platform communications. Although there is technology that may be able to integrate them in some ways, it is currently not standardized.

Mobile collections: Mobile data has lots of intricacies, including the metadata issue referenced above. At the moment, parties are unable to perform targeted collections, which means that the entire phone has to be imaged or collected (there is technology that can do filtering post-collection). This can raise privacy, volume, cost, and other concerns. The mobile data typically needs to be converted into another format for review.

Ephemeral Apps: The applications offer users the ability to set self-destructing messages that automatically disappear from recipients’ conversation histories. Some also have the ability to send encrypted messages. These types of applications pose additional collection challenges. It is important to uncover whether employees are using these types of applications for business purposes.

B. Information Governance of Collaboration and Communication Data

The use of collaboration platforms may create great efficiencies, accessibilities, and opportunities for sophisticated workflow management within an organization—whether it is a company, law firm, charitable organization, or government agency. Collaboration platforms may also enable greater social engagement and information sharing by employees. Work productivity and collaboration can increase tremendously. Nonetheless, the following information governance practices should be considered regarding the use of such platforms.

Before collaboration platforms are rolled out widely, an organization should give careful consideration to whether and how such platforms fit within the organization’s business processes and information systems. This can be accomplished through the creation of a collaborative lead and team that work

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

together to vet the platform and its provider.¹ Once a collaborative platform is selected and a directive is given across the organization to use the platform as a primary method of work collaboration, there should be clearly communicated acceptable use standards, and education across the organization on the use of the platform.²

As platform capabilities change, the governance on their use should keep pace to protect the organization against risks, and to remain compliant with legal, regulatory, and business duties.³ If a collaboration platform is being considered for an organization, the team vetting the platform and tools should understand the functions of data and information retention, preservation and legal hold and ensure capabilities for those functions within the platform or in connection with its use. The “owner” or “custodian” of the platform, as well as the legal eDiscovery team and any other relevant business unit authorized to access the platform data, should understand where the collaboration platform’s data is stored, who has access to it, and how the data may be produced.⁴

Record retention policies should be set for the information and work product created on the platform that align with the organization’s record retention policies. Collaboration platforms should have a reasonable retention period as determined by the relevant records management owner, and should take into consideration legal, business, and regulatory requirements for the retention of documents or categories of documents or communications, consistent with basic principles of information governance.

The platform must have the ability to adhere to legal holds that supersede normal retention periods, and the ability to implement those legal holds.⁵ The stakeholders must create a strategy to create and release legal holds on information and work product in the platform.⁶ The collaboration platform must have the ability to purge data when the retention period is met and when a legal hold is lifted.⁷ To ensure the ability to comply with eDiscovery obligations, data about usage and ability to access data, file sizes, searching capabilities across the collaboration platform or its components, and exportability are data points that should be available and readily communicable as a prerequisite for implementing such platforms.

Before any analysis of individual discovery related issues and best practices for continuous data streams it should be pointed out how important it is that these concepts be a part of a company’s well-crafted information governance plan. It’s vital that a company understand which apps are being used by which parts of their company and in what capacity:

¹ See Principle 1 of the 2019 Sedona Conference Commentary on Information Governance, Second Edition); McKinsey Digital Designing Data Governance that Delivers Value, Ex. 2, by Bryan Petzold, Matthias Ruggendorf, Kayvaun Rowshankish and Christopher Sporleder.

² What is.com Acceptable Use Policy (AUP) by Paul Kirvan.

³ Harvard Business Review. An Agile Approach to Change Management by Sarah Jensen Clayton, January 11, 2021.

⁴ ABA June 20, 2023, Not Your Same Old E-Discovery Reminder- Be Ready for Discovery of Collaborative Tech and Social Media, by C. Jade Davis.

⁵ The Sedona Conference Commentary on Legal Holds, “The Trigger & The Process” Vol. 20, 2019. Second Edition.

⁶ Id.

⁷ Id.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

- a. Retention Policy should anticipate ways to efficiently identify and preserve data e.g., retaining and tracking Channel names and users.
- b. Keep transparent, tracked processes.
- c. Creating policy not enough, companies should monitor if policies are being followed.
- d. Critical to be familiar with archiving and capabilities of chat channel platform.
- e. Know capabilities of version of platform currently being licensed, specifically retention settings for user across their devices (e.g. work desktop, laptop, phone)
- f. Know policies of platform for archiving data not available in current version of platform being licensed.
- g. Stay in contact with software provider to understand any evolving policies regarding access to documents.
- h. The plan should cover all devices and equipment.⁸

Creating a long-term retention strategy appropriate to the value and type of the information involves considering a broad range of factors pertaining to the digital assets and the circumstances of the organization itself. These factors should include business value, regulatory importance, intended retention schedule, legal hold status, file format, continued availability of the technologies required to access and read, the likely failure rate of the storage medium as it is configured, the available budget and resources of the organization, and/or (for third-party services such as cloud storage, software as a service (SaaS), etc.), the contractual agreements between the customer and provider.

C. Preservation of Collaboration and Communication Data

As with any ESI, parties have a duty retain relevant documents and data. When a party does not, it may be subject to sanctions. In re Google Play Store Antitrust Litigation, 2023 U.S. Dist. LEXIS 53218 (2023). In that case, the court found that Google—a sophisticated litigation party—continued with a policy of auto-deleting Chat messages after twenty-four hours even with pending litigation. The Court noted that Google had a strict policy of retaining emails subject to litigation holds. With emails, the employee could not override the preservation of emails. In contrast, Chats were deleted after twenty-four hours, and employees could retain certain Chats at their own discretion. In addition, employees were trained to avoid email communications in favor of Chat messages. While Google could have

⁸*The Sedona Conference Commentary on Information Governance*, Second Edition, 20 Sedona Conf. J. 150 (2019).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

instituted a preservation default on the Chat messages of the 360 employees subject to the litigation hold, it chose not to. The court imposed sanctions of attorney's fees and costs.

Similarly, in Brooks Sports, Inc. v. Anta (China) Co., Ltd., 2018 WL 7488924, at *9, 16-17 (E.D.Va. Nov. 30, 2018), the defendant was sanctioned for failing to produce messages from WeChat, a Chinese communication platform that the defendant used as a primary means of communication. The court rejected the defendant's argument that Chinese law prevented it from forcing its employees to produce the messages, noting, "Anta should not be able to conveniently use Chinese law to shield production of communications responsive to discovery requests when it could have set up Anta-controlled WeChat accounts for its employees' use which would not have the same issues regarding Chinese privacy laws." *Id.* at 13.

Because collaboration tools allow multiple people to edit documents simultaneously, the document collected may lack perfect synchronicity with the dynamic file. *See* Sedona Conference Journal, Vol. 22, at 139-140 (2021). Given that the files are not static and might change frequently, it raises a question about the extent of the preservation obligation. Additionally, the provider of the collaboration tool might place limits on the information that can be collected or preserved. *See id.* at 139-140, n. 146-147 (discussing that a user's ability to export certain information from Slack depends on the specific plan they signed up for). *See also*, Microsoft, Conduct an eDiscovery investigation of content in Microsoft Teams, (Aug. 2, 2023), <https://learn.microsoft.com/en-us/purview/ediscovery-teams-investigation> (last visited Sept. 8, 2023). Each collaboration tool may have its own retention settings. *See Drips Holdings, LLC v. Teledrip, LLC*, 2022 U.S. Dist. LEXIS 178233 (N.D. Ohio Sept. 29, 2022) (defendant found to have spoliated evidence where it changed its Slack retention setting from indefinite to a seven-day retention period and deleted all the Slack data at that point).

A producing party should consider these issues when instituting a litigation hold and when performing collections so that spoliation issues do not arise. *See* Sedona Conference Journal, Vol. 19, No.1, at 118-130 (a producing party is best situated to evaluate its preservation and production), 169-186 (production should be made in the form in which its ordinarily maintained), 193-198 (breach of a duty to preserve information may be sanctionable). *See also*, *Waymo LLC v. Uber Techs., Inc.*, 2018 U.S. Dist. LEXIS 16020, *78 (N.D. Cal. Jan. 29, 2018) (considering a spoliation argument after Uber lost Slack messages); *see also*, *In re Google Play Store Antitrust Litigation*, 2023 U.S. Dist. LEXIS 53218 (N.D. Cal. Mar. 28, 2023) (sanctioning Google for continuing with a policy of auto-deleting chat messages after twenty-four hours, even with pending litigation).

With respect to chat messages, a party should keep in mind that native files might be better to avoid spoliation issues. *See generally*, *Charter Communs. Operating, LLC v. Optymyze, LLC*, 2021 WL 1811627 (Del. Ch. 2021) (party produced Microsoft Teams messages as emails and was later compelled to produce them as native files, which revealed extensive spoliation); *Deal Genius, LLC v. O2Cool, LLC*, 2023 U.S. Dist. LEXIS 35342 (N.D. Ill. 2023) (party produced Microsoft Teams messages as emails raising questions regarding their compliance with their discovery obligations). *See also*, Sedona Conference Journal, Vol. 22, at 174-179 (2021) (discussing the difference between native versus static files and spoliation). Additionally, where there are documents linked to a collaboration tool in emails,

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

it would be prudent to preserve those as well. *See IQVIA, Inc. v. Veeva Sys.*, 2019 U.S. Dist. LEXIS 115894, *16 (D.N.J. July 10, 2019) (ordering producing party to produce Google Drive documents linked via email). *But see, Nichols v. Noom Inc.*, 2021 U.S. Dist. LEXIS 46860, *16 (S.D.N.Y. Mar. 11, 2021) (denying a motion to compel linked documents, finding that the plaintiff failed to establish they were proportional to the needs of the case).

1. Scope of Preservation

There is no one size fits all solution for preserving data on Collaboration Platforms. Identifying what is appropriate to preserve requires an analysis of what the platform was used for and how it was used. Collaboration platforms are engines rather than documents, and they may have multiple applications or tools contained within or interoperating with them. In order to identify the appropriate data to preserve for litigation, attorneys must understand the platform, the technologies which underlie the platform, and the ways with which platforms are used by the producing party. As a threshold consideration, the first step in identification and preservation of potentially responsive data within a Collaboration Platform is the same as the first step in any eDiscovery project, identifying what, where, and how potentially relevant data is stored. The “what” analysis should include identifying the underlying needs of the case, information gathering about the platform itself as well as its archival and data extraction processes, and any data sources which are integrated with the platform such as secondary applications. For example, Microsoft Teams allows the integration of many web applications directly within an end user’s application in such a way that it may not be clear on the face of the tool where one application ends, and another begins.

Many such platforms store their data disconnected from what appear to users to be “documents.” At time of publication, for example, Slack maintains their data as JSON files which contain user and date-stamped change history of chats and interactions within the platform. Users do not see the files underlying the platform, though Slack is stored as JSON in the ordinary course of business. The application itself is a rendering tool for the stored JSON files. When identifying data to be preserved for litigation, a priority must be to understand the technology, how it is used, and how the data available to it may be extracted.

Collaboration Platforms vary in technology and implementation and as such there is no one-size-fits-all solution for these increasingly complex interoperable systems. Some Collaboration Platforms contain dynamic “documents” which can vary by audience and are saved as datum within a database. A user may not see, know, or appreciate the nature of the pieces of the documents they’re working with. In a corporate context, for instance, certain users could see all information in a document while others with less extensive permissions could see most of a document, but PII might be either redacted or invisible. Special tools may be required to preserve the data and metadata required to appropriately generate documents accordingly.

The method of preservation is both tool and matter specific. Depending on the platform, there may be a significant number of options for facets of documents on which to preserve such as date, conversation, group membership and more.

Unlike email which is a unitary container of information apps such as Slack are a container of many communications similar to chat tools or entire databases. Though there is not as much case law or

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

industry wide consensus on practices, the increased reliance and usage of continuous data streams have created higher expectations of diligence and care in preservation and protocols set up to quickly locate relevant continuous data streams datasets. Databases not documents. Given the accelerated business use of continuous data streams it's clear that are likely to contain relevant information. In order to avoid potential sanctions it's important that the subpoenaed party understand how the tool is being used within their company and develop a plan to preserve this information.

Though continuous data streams are a recent and ever-changing technology space the standard reasonable steps for preservation in FRCP 37 e are applicable. The reasonable steps to preserve standard for continuous data streams will grow with companies and employees greater reliance on these collaborative tools. Retention policies must account for the preservation of chat messages and not rely on such things as email notifications of chat messages. With so much business being conducted in these tools unreasonable policies (e.g. licenses or settings that don't save chats past 24 hours or any unreasonably short amount of time).

Under FRCP 37(e), should a court find that subpoenaed party failed to take these reasonable steps to preserve potentially relevant data from collaboration tools the penalties could include the presumption that the information lost was unfavorable to the subpoenaed party, jury instructions presuming that the information missing information was unfavorable or even dismissal of the action and an entering of a default judgment.

As continuous data stream tools can be found in so many places duplication issues when collecting from multiple devices from a single user. Leveraging tools, such as Brainspace, that can analyze collected extracted text can potentially pare down review sets. Likewise, installing consistent protocols for collecting and processing in same way from same software.

Practice pointer: It is important to know and understand, via retention policies or custodian interviews, how individual employees and channels use these continuous data streams. Included in this evaluation should be individual and group permissions as well as any differences in user capabilities depending on the nature of the license of the app as well as the evolving capabilities of the app over time. (e.g. an analysis of any privacy issues as certain types of apps may be utilized for users for personal reasons or other types of apps, e.g.. Zoom, that offer continuous chat capabilities over time).

It is often likely the case that custodian interviews, lists of custodians, channels and participants and metadata will not completely solve the problems of locating relevant material in an efficient matter. Time and budget should be built into the Review phase to identify and discuss additional potential custodians or channels for processing and loading.

How the core of a business is run day to day is also key particularly when determining if a member of a channel that does not actively participate should still be of interest. The nature of the claim will also help to give color to this discussion. For example, if the cause of action is for the recovery of a bonus that a former employee was denied and the channel in question is for the servicing of a customer than a member of that channel that does not participate is most likely not of interest. However, if the cause

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

of action is insider trading where one participant in a channel is communicating what financial instrument will be manipulated and to what end, merely being in that channel could be grounds for collection.

2. Potential Preservation Issues

a. Chat-type data

Data which records conversations between individuals may pose specific issues depending on platform. How chat-type data will be preserved and on what bases should be established early. Conversations can stop and start at any time and it's possible to respond to something months later.

b. Trigger of the Duty to Preserve

As with all potentially relevant ESI, the trigger for preserving data is when a litigant knows or should know that litigation is likely.⁹ Each collaboration platform has their own retention methodology and many also provide access to third party tools via an API. Establishing what should be preserved as well as how to preserve it should be considered as early as possible in the process. Once a trigger event has occurred, litigation is reasonably anticipated, and a legal hold is established, steps should be taken to ensure that the hold implementation is not altered once it is in place.

In *Drips Holdings v. Teledrip*,¹⁰ defendants Teledrip used Slack as a “typical” means of internal and external communication. The founder of Teledrip, Inc. was found to have adjusted the retention settings of the company’s Slack environment after learning of potential litigation but before receiving a letter with a hold notice from Plaintiffs. In the interim, the defendants changed the retention settings of Teledrip’s Slack from unlimited retention to only retaining seven days of data and deleted a prior Slack export. Defendants argued that the change, despite the timing, was related to their understanding of the requirements of the California Consumer Privacy Act (“CCPA”) and the associated International Standards Organization (“ISO”) implementation. The court was not persuaded. Not only had defendants altered the retention period and destroyed a prior export, they also refused to update their retention policy to accommodate the hold, citing to the CCPA issue. After rejecting the CCPA argument as wrong and characterizing certain assertions by the defendants as “blatantly false,” the court ordered a mandatory adverse-inference instruction.

c. Third-Party Preservation Tools

Preservation of collaboration platform data is a complex set of tasks. Tools are regularly developed to preserve and collect data from new sources. It is important to analyze what third party preservation tools actually preserve. Do they capture all, some, or none of the metadata specific to the collaboration platform? Do they capture in-time snapshots of data on the platform? If the data is stored separately

⁹ The Sedona Conference Commentary on Legal Holds, Second Edition: The Trigger & The Process

¹⁰ *Drips Holdings, LLC v. Teledrip LLC*, 2022 U.S. Dist. LEXIS 153668, 2022 WL 3282676 (N.D. Ohio Apr. 5, 2022) (Carmen E. Henderson, M.J.)

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

from rendered documents, what is saved? How will what is saved be rendered for review and production if necessary?

d. Complex “Documents”

The nature of collaboration platforms may allow for data stored for one purpose by a user in a particular digital location to be viewed, accessed, or modified within what purports to be a document in an entirely different location. The storage of user-generated content in a database which permits creation of what appear to users to be documents, though the underlying data is not stored that way, presents an additional complication. It may be useful to consider data contained in a collaboration platform as ESI even in instances where data is presented facially to a user as a document. The design and implementation of the tool should be understood as should its storage. “Documents” may be generated by a collaboration platform on a permissions-based access restricted basis, rather than stored in that way.

e. Local Collaboration Documents

Due to the nature of collaboration platforms and the ways in which some tools are configured, it is possible that documents which relate to a collaboration platform may be downloaded to a user’s account or personal computer. The version of a document stored locally may not be what would traditionally be seen as a “master” document. Additionally, if the platform allows multiple users to access the same document offline, there may not be a single master document. Where to preserve collaborative documents will determine the tools needed to preserve.

It may not be possible to preserve full in-time copies of documents as collaborators may have changed local copies before they had a chance to sync. As discussed above, data which appears at first viewable by a user may not necessarily be viewable by that same user looking at the same document at a different time or after a sync.

3. Proportionality in Preservation

Preservation costs associated with collaboration and communication platform data will vary depending on a number of factors. These include how prevalent a tool is, how long it has been in service, as well as Platform specific considerations. Prices on preservation, whether in place or via third party tool, will vary depending on Collaboration Platform. As with all forms of ESI, the needs of the matter must be balanced against the cost of preservation.

Each tool, with its unique characteristics, collection tools, and data structure must be well understood in order to appropriately preserve relevant data in a fulsome but not overly burdensome way. An analysis is required to determine whether preserving Collaboration Platform data in whole or in part is proportional to the needs of the case. Federal Rule 26(b) provides guidance on the six factors which must be balanced against the costs associated with preservation and in turn, review and production.

Insofar as a producing party is best positioned to assess their own preservation costs and to understand their own tools, a threshold analysis must be performed early to determine the scope of preservation

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

appropriate for the matter and its implementation on the Collaboration Platform. An important consideration in the sixth test factor is identifying who elected to use a Collaboration Platform. The burden of expense should be viewed through a reasonable lens - if the producing party selected a particularly expensive platform to preserve, they should not necessarily be absolved of their preservation obligation. A producing party required to use a Collaboration Platform by a different entity may be able to demonstrate a higher level of burden than a producing party who elected to use the tool. A producing party that chose their own tools can reasonably be expected to preserve and produce the data from that tool in all but the most exceptional cases.

D. Collection of Collaboration and Communication Data

In collecting ESI from a Collaboration Platform, disclosures should be made by counsel for the parties early in the litigation, about the feasibility of the eDiscovery proposed and the methodology used to select, process, review and produce the ESI.

In collaboration and communication platforms, such as Slack or Microsoft Teams, ongoing and dynamic interactions between individuals and groups enable instantaneous collaboration on documents as well as instantaneous communications and information sharing.

Though the usage and breadth of situations where these tools can be helpful have grown over time, the pandemic caused a massive spike in their prevalence. As of 2021 nearly 80% of worldwide workers used digital collaboration tools (Gartner, Inc. Digital Worker Experience Survey, 2021). With this surge in usage it is vital for companies to understand this new information both in terms of locating relevant documents for litigation but also creating practices and workflows to do that cost effectively.

Generally, social media data can be collected in a screenshot or by using a screen scraping program such as X1 Social Discovery, or “DYI” (Download Your Information)¹¹ functionality provided by some platforms, or by using programs which call on an API (Application Programming Interface) provided by some platforms¹² to enable retrieval of messages/threads/posts, etc.

Screenshots. Screenshots are merely images without textual content or fielded metadata. If OCR’ed, they will have text, but they won’t have fielded metadata such as author or post date and time of the individual items shown in the screenshot. Processing of individual screenshots by an ediscovery processing program will create a single record per screenshot; individual messages, comments, posts, etc. will not become separate records with fielded metadata linked as family groups, and, if the screenshot is OCR’ed, the extracted full text will be a single text .txt file for the whole screenshot.

¹¹ Add cites to, e.g., Facebook, Instagram DYI programs

¹² Add cites to, e.g., Google and MS API’s and programs that use them to collect ESI, e.g., Metasploit Forensic Evidence Collector.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Screen scraping programs. Forensic or ediscovery screen scraping programs capture and use pattern-matching on the underlying HTML code and incorporated components (CSS, JavaScript files, etc.) of a web page that a browser makes visible as the screen that a user sees.

They may be able to extract and collect individual posts with fielded metadata by parsing HTML elements; if an individual post contains elements such as “<author>Jane Eyre</author>,” they can present them in a format similar to their appearance in the original web pages and group them as families.

However, screen scraping programs are fragile and susceptible to failure if/when a platform changes the HTML being scraped and parsed. For example, if a platform changes the author or poster element from “<author>...</author>” to “<Author>...</Author>” or changes the scrolling behavior on their screens or the appearance or location of a “Next” button, a screen scraping program may fail to retrieve post authors or capture data not shown on an initial screen. Additionally, screen scraping programs are at the mercy of a platform which may impose countermeasures against automated collections, such as limiting collection requests from a given computer (IP) address or requiring manual validation for requests from an IP address which had not previously accessed an account.

Screen scraping programs may create load ready files which can be ingested directly into litigation support programs, or not quite load ready files which will require further processing before they can be loaded into a litigation support program.

DYI. A platform’s DYI download may, for privacy reasons, include only content authored by the downloading user himself or herself. For example, while a DYI download of a message thread may include all messages within it, regardless of who authored them, the DYI download of a user’s comments on other users’ posts outside a message thread may include the user’s own comments but not the third-party posts being commented on.

DYI downloads may not be load ready or be in a format that can be readily ingested by an ediscovery processing program¹³, but there are commercial programs such as Message Crawler¹⁴ that can process some DYI downloads into RSMF or review platform ready load files.. Turning the HTML into threads and grouped individual items suitable for loading requires basically the same kind of programming and processing that an eDiscovery/forensic screen scraping program does but based on directly ingesting the DYI static HTML files instead of collecting a live web page. JSON is a format with nested elements containing fielded data, which is relatively easy to parse programmatically to create individual items with fielded metadata. Like a basic text file, a DYI JSON file does not contain formatting code. They are not load ready.

¹³ For example, Facebook DYI downloads come in HTML or JSON format.

¹⁴ <https://www.hashtaglegal.com/message-crawler>

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

API. APIs are interfaces for programs and programmers to access a platform programmatically. For example, Google provides an API to enable its clients and tool makers such as Metaspikes¹⁵ and Onna¹⁶ to interface directly with Google drive documents and folders by creating programs that retrieve files in Google Drive. *Introduction to Google Drive API*, available at <https://developers.google.com/drive/api/v3/about-sdk> (accessed on June 5, 2023). Similarly, Microsoft provides an API to retrieve SharePoint documents using Microsoft's SharePoint REST API. *Get to know the SharePoint REST service*, available at <https://learn.microsoft.com/en-us/sharepoint/dev/sp-add-ins/get-to-know-the-sharepoint-rest-service> (accessed on June 4, 2023). Third-party programs, including commercial programs such as Hanzo¹⁷ as well as custom programs, may use multiple APIs from different platforms, e.g., to collect Slack threads and messages and then to collect the modern attachments to those messages resident on Google Drive/Workspace.

If a party has a heterogeneous environment, e.g., using Microsoft mail but storing linked attachments in both Google Workspace and SharePoint, there may not be commercial programs which can collect all of the information, but using a program to collect whatever portion it can collect is preferable to not collecting at all, e.g., an 80% solution can solve 80% of a problem.

Matters get more complicated when there is no commercial product for all or part of a collection and custom programming must be employed, e.g., to retrieve linked attachments and preserve their family relationship to their parents. In general, such programming will be well within the capabilities of a full-service vendor¹⁸ or a competent programmer.

It can be argued that the lack of a commercial tool does not relieve a party which has chosen to store its ESI in such a way that such custom programming is required to meet its discovery obligation is not thereby excused from such obligations. A producing party may not “shield itself from discovery by utilizing a system of recordkeeping which conceals rather than discloses relevant records, or makes it unduly difficult to identify or locate them, thus rendering the production of documents an excessively burdensome and costly expenditure.” *Wesley v. Muhammad*, No. 05 Civ. 5833 (GEL) (MHD), 2008 WL 4386871, at *5 (S.D.N.Y. 2008) (quotation marks and citation omitted); *see also Kozłowski v. Sears Roebuck & Co.*, 73 F.R.D. 73, 76 (D. Mass. 1976) (same and observing that “[t]o allow a defendant whose business generates massive records to frustrate discovery by creating an inadequate filing system, and then claiming undue burden, would defeat the purposes of the discovery rules.”); *Pom Wonderful LLC v. Coca-Cola Co.*, No. 08 Civ. 86237 (SJO) (FMO), 2009 WL 10655335, at *3 (C.D. Cal. Nov. 30, 2009) (rejecting argument that defendant “has no method to automatically re-link emails

¹⁵ <https://www.metaspikes.com/forensic-email-collector>

¹⁶ <https://onna.com/blog/introducing-onna-onedrive-connector>

¹⁷ <https://www.hanzo.co/blog/hanzo-holds-new-follow-the-link-capability-preserves-full-text-of-linked-documents-in-google-drive>.

¹⁸ For example, Epiq created a Chat Connector for Microsoft Teams which extracts linked attachments from Microsoft Teams records (<https://www.epiqglobal.com/en-us/technologies/legal-solutions/chat-connector> (accessed on June 5, 2023) (referencing “Modern attachment harvesting” as a feature of Epiq Chat Connector tool).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

with their alleged “missing” attachments[]’ and that requiring it to do so would ‘employ a tedious manual process,’” because a defendant “cannot seek to preclude plaintiff from pursuing discovery based on a record-keeping system that is plainly inadequate”); *Lou v. Ma Labs, Inc.*, No. 12 Civ. 5409 (WHA) (NC), 2013 WL 12328278, at *2 (N.D. Cal. Mar. 28, 2013) (“The fact that a corporation has an unwieldy record keeping system which requires it to incur heavy expenditures of time and effort to produce requested documents is an insufficient reason to prevent disclosure of otherwise discoverable information.”) (quotation marks and citation omitted).

D. Search, Review And Production of Collaboration and Communication Data

ESI from collaboration platforms implicates the issue of what is the temporal scope of ESI from the platform- is each communication one message, should there be a time parameter selected to aggregate, review and produce messages or communications, is there a technological limitation by the E-Discovery Review Platforms by which this data, information, and communications must be ingested, reviewed and produced? It is incumbent on all parties to understand the options available to review and process the Collaborative Platform Data/Communications and communicate this to all parties early on in the case in order for an agreement to be reached on processing and producing ESI from Collaborative Platforms.

Parties should also discuss how they plan to address certain issues related to communication and collaboration platforms, if those data sources are a component of the discovery in their matter. Best practice would be to memorialize these discussions as part of an ESI Protocol.

Practice pointer: Though there is rarely a perfect solution it is key to openly discuss and determine the most effective way to unitize collaboration chat data be it in 24 hour period chunks, by breaks (e.g. 4 hours, 8 hours) in a chat or other measures that suit the potentially relevant dataset.

To determine proper unitization it's key to track and compare the home time zones of potential custodians for processing and unitization purposes. For example,

- If all of the potentially relevant custodians/channels are in the same time zone, it may be best to process in that local time zone and unitize at say Midnight of every 24 hour period.
- If the users in potentially relevant channels are across the world, then maybe processing in UTC and unitizing by breaks in chat would be more accurate.

Entire targeted Slack dataset, targeted channels or custodians, must be processed before any information can be extracted. Not possible to know the volume without processing. **Milbeck v. TrueCar Inc.**, 2019 U.S. Dist. LEXIS 165649, 2019 WL 4570017 (C.D. Cal. May 2, 2019) (Alicia Rosenberg, M.J.) SLACK

Given that and similar ways in which continuous data streams cannot leverage threading or deduplication in the same way as email collections can meet and confers are crucial. They can help limit the amount of channels to be collected as well as discuss the scope of users and channels with

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

potentially relevant material. As discussed, whichever unitization path is chosen it may not capture something as subjective as the beginning and end of a conversation perfectly. To assist in this analysis one can:

- Scroll to see where a conversation begins and ends.
- Deploy manual unitization, platforms offer ability to manually break up chat in image mode.
- Discuss with opposing the potential need for contextual, Not Responsive, documents bracketing relevant material to better understand the conversation.
- Possibility of maintaining a master image of full conversation to show context of where redactions for privilege, not responsiveness or privacy have occurred for ease of challenging and tracking.

In the large majority of reviews, it is helpful to either utilize agreed upon search terms or leverage TAR to significantly reduce the review population and bring relevant documents to the forefront earlier. With the review of continuous data stream data it's important to discuss at meet and confers:

- Specific request categories,
- Search methodologies.
- Provide a list of chat channels, including the title and a brief description of each channel, the number of messages in each channel, the users associated with each channel, and other data that will assist the parties in tailoring the review and production.

Search Terms specific to potentially relevant users/channels. Generally, chat messages are not as detailed as emails. Therefore, search terms should be tailored to the nature of both chats in general and the specifics of the parties involved.

Due to the complexity of producing data from continuous data streams, parties must often rely on vendors and/or the platforms to aid in satisfying their discovery obligations. In Calendar Research LLC v. Stubhub Inc., 2019 U.S. Dist. LEXIS 65307 (2019), the plaintiff requested Slack messages from an individual defendant's employer. Because the employer had a free Slack account, certain Slack folders were not retrievable. The defendants paid for an upgraded account, but Slack denied full access because not all the parties on the account had consented. However, Slack provided a "utility tool" to target the channels used by the individual defendants. In Warner Bros. Ent. Inc. v. Random Tuesday, Inc., 2021 U.S. Dist. LEXIS 250597 (2021), the defendant's initial production of Slack messages to plaintiff were "virtually useless." The defendants worked with a vendor to correct the problem and were given an extension to produce the information.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

1. Family Relationships

Linked attachments can pose special challenges. Before committing to producing linked attachments, parties should first research any technological capabilities and limitations their client's systems may have (e.g., Google, Microsoft 365). **[refer to lessons in Stubhub case; cite to *In re Meta Pixel Healthcare Litigation*]** First, parties need to understand whether or not the linked attachments can be collected and “connected” back together with the transmittal source (email, message, etc.) in an automated technology-driven process. If so, then they need to understand which version will be connected - the version that existed at the time of the communication or the version that exists at the time of collection (or some other version). If additional versions have been preserved, parties should discuss whether those should also be produced, even if they are not able to be “connected” to the original transmittal source. Note that various versions may be separately collected as edocs and produced as standalone documents.

Available tools to preserve family relationships:

- (Microsoft Purview (eDiscovery (Premium) edition only)
- Metaspike Forensic Evidence Collector (Google Mail links to Google Drive/Suite/Workspace only)
- Vendor programs (Epiq Chat Connector for Microsoft Teams)
- Vendor or home-brew custom programs)

Commercial program functionality typically lags behind technology usage patterns; some commercial programs are outgrowths of custom programs created to address particular client needs. Commercial programs often will be of limited use in mixed environments, e.g., M 365 email with Google Drive links or Google Mail links to OneDrive or SharePoint. Needed commercial programs functionality may only be available as part of larger enterprise offerings (onna? hanzo?) which may have to have already been in place while others may be suitable to be used if and as needed (Cellebrite). Commercial programs are not per se better than custom solutions. Commercial programs need to address needs (heterogeneous users, user interfaces, support, upgrades, documentation, etc.) that custom programs being run by their creators may not.

Most functionality required to identify links and extract links is fairly basic and common; programs and program libraries to do this have existed almost as long as the internet and email. It is easiest to preserve the relationship of links to their parents at the time that a parent is collected as the link url is part of the parent. The reverse, trying to determine if a standalone document was related to one or more unknown parents, will be difficult and often impossible. Consideration must be given to where link extraction should occur in a discovery process. Links which can be collected from a native form of a parent may be invisible to search and unretrievable further downstream, as some ESI processing programs may not include link urls in extracted text unless the link url is also the display text of the link, i.e., links like “please review this report” where the link url is not visible in the text will not be included in the extracted text

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

2. Social media considerations

DYI programs may show only a user's own interactions with the platform and not the third party content with which a user is interacting. For example, If a Facebook user comments on another user's post, the DYI for that user will include the comment but not the post being commented on. Message threads with multiple parties may, as with Facebook messaging, be included in full with messages from all parties included in the DYI. DYI programs may offer different formats, e.g., html, json, etc., and the metadata available may differ from format to format. Options to capture full interactions may have limitations. PDFs created from screen capture images may not include item level metadata or parsing such item level metadata from OCR'ing screen captures may be difficult and different parsing routines may be required for different types data (comments vs message threads, etc.). Commercial screen-scraping based capture programs may fail if a platform provider changes the underlying html of a page. Platforms may limit or cut off access if requests are made too often, or from different IP addresses or exhibit indicia of automation. Collections may require that account users be available to respond to identity authentication requests.

3. Considerations for searching data post-collection

Search terms are frequently used in standard discovery matters involving emails and edocs to narrow the population for review. Parties should carefully consider whether search terms are a viable option for short message data (e.g. texts, Slack, Teams, What's App, etc.). During the collection process, interview questions should be included that discuss how people are communicating in collaboration platforms. It will be important to understand which acronyms/shorthand are commonly used. This will also be important for privileged communications, as attorneys may just be referred to with first names in short messages. Once data is collected, parties should analyze/sample to get an understanding for how people are communicating within these platforms. Part of this analysis should include evaluating whether different terms may be necessary for the short message data versus emails/edocs. People communicate much more informally in short messaging platforms than they do in emails. Therefore, the same search terms may not yield the intended results. It is important to test the proposed search terms and sample the hits and non-hits to look for additional terms, or to determine whether search terms may or may not be appropriate in the matter. ESI protocols should be drafted to allow for this type of iteration, with analysis and sampling, and parties should not agree to specific search terms prior to testing them.

If search terms are used on short message data, parties may want to discuss whether to include a certain number of messages before and after the hits for review. Before agreeing to this approach, parties should confirm that the platforms they are using for processing and searching have the ability to bring in additional messages for review in this manner. Another option is to review a 24-hour window of messages surrounding the search hits. Parties should clarify how they are interpreting the 24-hour window (Is it a rolling 24 hours surrounding any search hit? What if there are multiple search hits in the same 24-hour window - does that expand it out even further?) A third option would be to

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

review the entire “channel” or “chain,” although that approach may be burdensome and unnecessarily expand the review universe.

Technology is ever-advancing, and there may be other ways to group search term hits for review. AI may be able to assist in grouping topically related messages together, which may be the preferred way to review. This paper recognizes that available technology should be considered as part of the discussions. Parties may use other culling mechanisms either in conjunction with searching or as standalone methods. Parties should discuss whether the review should be limited to certain date ranges, or in the context of phone or other app data, whether it should be limited to certain communicators or phone numbers.

4. Review

As mentioned above, AI or future technologies may be able to group “related” messages together for ease of review. Not all parties will have access to the newest technology so it will be important to have multiple options for review of short messages.

When “smart-grouping” is not available, parties may review messages in 24-hour windows, which appears to be a common standard for review. Although not a perfect solution (you can still have issues where related conversations span multiple days), organizing them in this manner can typically provide a manageable set of messages for review at a time. Microsoft’s Purview Premium creates “transcript” files in 24-hour windows. Messages can be converted to Relativity’s Short Message Format (RSMF) in 24-hour or other time periods. There are also other technologies on the market that will allow parties to create other groupings of messages for review. The manner in which the messages are grouped should be left to the producing party pursuant to Principle 6. Parties should not be required to purchase additional technology in order to review short messages.

When creating the grouping of messages for review, one aspect to consider is cross-platform communications. People will frequently “jump” from one communication platform to another. For example, they may send an email and from there move the conversation to Teams for easier collaboration, and then they may take the discussion “offline” to text messaging. It is important to note that it is currently not “simple” to combine these different platforms so that the whole discussion is together. Although there is technology that purports to do that, parties should not be required to purchase additional technology to try to combine data from separate platforms. This weighs heavily in favor of a burden argument and should be considered in any proportionality considerations. Parties may undertake such an analysis in order to understand their data and the communications, but that is more akin to “work product” than standard e-Discovery..

TAR/Active Learning may be applied to short message data, although parties should carefully consider whether the technology is providing reasonable results based on the nature of this data. There may not be enough text for machine learning to accurately identify documents as likely to be relevant/not relevant based on the type of matter. Technology will continue to advance so this option should be available to parties, with the same caveats on validation and process as standard documents.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Parties may use search terms in conjunction with TAR, or use one or the other in isolation. Producing parties are best situated to analyze their data and make reasonable decisions on how to filter and review documents for production. The parties should apply the same analysis as they do on standard documents as to when to use certain technology in the review process.

5. Production

a. Content of Production

When it comes to determining what information from a collaboration platform should be produced, a number of considerations must be given. Parties should discuss and attempt to agree upon the content of production in an ESI protocol to avoid potential disputes. For example, as discussed above, there is no clear definition of “family” when it comes to short messages, thus messages can be “standalone” or they can be part of a related conversation. The idea of “family” in short messages would be the message plus any attachment (linked or otherwise) or reaction (ex. Gif, emoji) that is part of the individual message. The part that is not analogous to an email and attachment is whether messages surrounding the responsive message should also be produced.

Context is subjective and therefore it will be difficult to create hard and fast rules. Parties reviewing the messages will be best suited to determine which messages provide context to the responsive message. Parties should be able to apply rules to determine that a certain number of messages before and after the responsive message should be produced, which should reduce disputes over this issue. The downside to this approach is twofold. First, there may be messages outside the before/after messages that would also provide helpful context, and those messages would not be produced if the parties are following the protocol. Second, totally irrelevant messages may be produced, which could contain personal and sensitive information. Communication within short messaging applications tends to jump from one subject to another in quick succession, and then back to earlier subjects. Many of the “interspersed” messages may be entirely personal. Since messages are not “documents,” we cannot impose the same boundaries on them, and need to think of them as individual communications, some of which may be totally irrelevant to the matter at hand.

The issue of what messages to produce for context also depends somewhat on the platform. On some platforms, parties are limited to the message grouping they are using for review. For example, if they create a 24-hour window of messages, then they are reviewing all those messages as one record. They would need to redact out irrelevant messages, which could be burdensome and time consuming. It could also lead to further discovery disputes if the other party challenges the redacted messages. Other platforms allow for “slicing” or creating a new record with just the responsive messages. This is less time consuming and more efficient than redacting out 100s or 1000s of irrelevant messages. The potential downside is that the receiving party doesn’t know where the messages that were not produced appeared in the original conversation. However, they are likely not entitled to that information, just as they are not entitled to “sample” the documents that weren’t produced in standard discovery.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Depending on the parameters of the review and the platform, parties can either redact the privileged messages, or create separate “slices” of the privileged and non-privileged messages. Identifying privilege may be more complicated when dealing with short messages. Attorneys may be referred to as first names only or as initials, or some other shorthand, which will be easy for reviewers to miss. Another complication is that people may be emailing with an attorney, but then jump platforms to Slack/Teams and start discussing the legal advice provided in the email, but not include the attorney in the short message communication. It may be difficult for reviewers to recognize that privileged content without the benefit of the context of the other platforms. Parties should keep in mind that reasonableness is the standard and not perfection. If cross-platform communications are later determined to be privileged from context not apparent from standalone review, parties should allow leniency for clawbacks.

b. Form of Production

Parties should discuss and memorialize the form of production in an ESI protocol, just like they do with other discovery requests. Note that this may also be technology dependent, and may change as the technology advances. Standard tiffs/pdfs, plus text, metadata and load files should be sufficient for short message data. There currently is no “native format” production option for short message data. In the future, parties may be able to exchange production in RSMF so that they can review it in a way that simulates the native application. Other collaboration platforms may be able to be produced in native format, if needed. Parties should evaluate the technological options, keeping in mind how the data will be used in depositions and at trial. A static format may be preferred for ease of use.

In certain situations, it may be acceptable to produce screenshots of relevant texts or messages. If there are no questions as to authenticity and only a limited number of messages are relevant, it may be more cost effective and proportional to allow parties to produce in this manner. For example, a custodian may have a few relevant text messages on their phone, and it would be unreasonable to require a full forensic image just to produce those few messages (unless there was some dispute over metadata or when the messages were sent or whether the messages had been altered). Parties should consider the burden to the producing party and the value to the receiving party when determining the format of production from data sources that are not easily/cost-efficiently collected and produced.

Parties should discuss what metadata is readily available for the short message data. For example, if parties are reviewing in 24-hour periods, they may only have the metadata from the first message, or they may be able to provide metadata for all the individual messages. Or, they may be able to provide all the participants and a date range, but not an entry for each message. It is important to understand the technical limitations prior to agreeing to a format that is not readily available. Requesting parties should not seek metadata in a format that is not readily available or is burdensome to produce.

If individual messages are produced as separate records, parties should try to produce a field to indicate family relationships (messages and attachments). If possible, parties should try to produce a field that indicates context groups.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

If reviewing and producing multiple messages together as one pdf or tif, they may redact individual privilege messages. If creating slices, the parties may agree to produce a privileged placeholder for the slice of messages that was withheld (in addition to logging). Parties should discuss treatment of unavailable/unproduced attachments. These can either be by image placeholders or withheld (non-responsive attachments).

Collaboration and communication platform data must be produced in a way that is useable. It is axiomatic that disclosed information is useful to the opposing party only when it can be organized and accessed logically. A producing party may be tempted to release large amounts of data and place the burden on the requesting party to find what they requested. Even if the producing party is not purposefully “hiding a needle in a haystack,” document dumps are disfavored by the courts. Further, parties should consider what value information may be garnered from *how* collaborative platforms are used, independent of the communications and documents contained therein, particularly with respect to how individuals or documents are organized within the platform, and be prepared to request or produce documents in a way that preserves this information.

In Media v. Spain, 2020 U.S. Dist. LEXIS 260540 (2020), the plaintiff produced Slack messages in response to the defendants’ requests for production. The defendants could not open the file and subsequently complained that the messages were not Bates-numbered. The plaintiff then produced over 139,000 Slack messages as individual documents. The plaintiff did not provide any reference as to which messages responded to which request for production. The plaintiff merely supplied a generic response to each request: “documents already produced in this case.” The defendants complained that many of the produced Slack messages were not relevant “to the action at all.”

The U.S. District Court for the Southern District of California ruled that the plaintiff needed to amend its responses “to identify which documents are responsive to which requests.” Noting that discovery rules prohibit production of “a mass of undifferentiated documents for the responding party to inspect,” the court ruled that the 139,000 individual-documents-dump did not constitute a “reasonably usable format” within the meaning of FRCP 34(b)(2)E(ii). According to the court, the document dump effectively shifted the burden of review for responsiveness from the plaintiff to the defendants.

In Podium Corp. v. Chekit Geolocation Servs., 2022 U.S. Dist. LEXIS 98197 (2022), the U.S. District Court for the District of Utah, Central Division analyzed a discovery dispute similarly to the court in Media v. Spain. The plaintiff argued that it produced the documents, including Slack messages, “as they are kept in the ordinary course of business.” That may be, said the court, but discovery rules exist to “facilitate the production of records in a useful manner to minimize discovery costs.” The court ruled the production of documents must be organized in such a way that the requesting party may determine with reasonable effort which documents are responsive to its requests.” Noting “wholesale dumping of documents” does not satisfy a producing party’s obligations when responding to interrogatories or requests for production, the court compelled the plaintiff to Bates-number the documents and identify which documents were responsive to particular requests for production.

In contrast to judicial concerns about dumping documents and flooding the opposing party with too much information, some courts found that the responding parties did not provide enough information

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

in their discovery disclosures. Information contained in continuous message streams may need context and even irrelevant, nonprivileged communications must be disclosed to provide such context. In Bidprime, LLC v. Smartprocure, Inc., 2018 U.S. Dist. LEXIS 222868 (2018), the U.S. District Court for the Western District of Texas, Austin Division ruled that the responding party is not within its rights to redact irrelevant information and full chat logs must be produced. In Lubrizol Corp. v. IBM Corp., (get cite), the U.S. District Court in the Northern District Ohio, Eastern Division found that production of entire Slack threads containing only one responsive message overly burdensome. It ruled that, if the Slack conversation has fewer than twenty messages, the entire channel must be produced. Again, context is key. If the channel has more than twenty messages, the ten prior messages and ten subsequent messages must be produced.

Collaboration platforms may have their own value as compilations, independent of whether documents contained therein are duplicative of custodial productions. In securities litigation involving Pfizer, it was discovered that Pfizer used a platform called “eRooms,” a collaborative application for employees to share documents, conduct discussions/instant message, and conduct informal polls. In re Pfizer Inc. Securities Litigation, 288 F.R.D. 297 (S.D.N.Y. 2013). The court recognized that “[a]lthough the eRooms contain documents that may be largely duplicative of the custodial productions, they have a value in of themselves as compilations. The manner in which Pfizer and its employees internally organized documents is relevant because it allows Plaintiffs to draw connections and understand the narrative of events in a way not necessarily afforded by a custodial production.” *Id.* at 317.

E. Common Issues Arising from Collaboration and Communication Platforms

1. Who is a Custodian?

Identifying, preserving, and collecting relevant documents for litigation by “custodian” predates the advent of electronically stored data (“ESI”). After all, when confronted with a large amount of data, it makes sense to first identify who is likely to have created, stored, sent, and received documents and communications relevant to the issues being litigated and to review that data first. Organizing discovery by custodian continues to be the predominant way to identify, preserve, and collect data now that ESI is the primary source of evidence used in litigation.¹⁹

In the world of collaboration platforms, however, identifying the custodian of shared documents and communications may be difficult because a custodian may collaborate with many other people on a single document; or may communicate with colleagues in private and public chats as well as in email.

¹⁹ See, e.g., The Sedona Conference, *Commentary on Preservation, Management and Identification of Sources of Information that are Not Reasonably Accessible*, 10 SEDONA CONF. J. 281, 287 (2009) (“Once a party identifies the information that would be relevant to a pending or imminent dispute, it must determine where that information likely resides. This task typically involves identifying (1) the custodians who created or controls relevant information and (2) the data sources where the custodians’ information resides.”).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

In some cases, strict adherence to “custodian” as a governing approach to discovery may result in either over- or under-collection of data.

Consider, for example, a high-level employee identified as a “key custodian” who is a member of many private and public communication spaces (i.e., chats or channels). In a large company, these chats or channels might number into the hundreds. While a fair argument may be made that all those channels should be preserved; that does not necessarily correlate to collecting, producing, and reviewing all of that ESI. However, a responding party would need to know how much the “key custodian” participates in each channel; the purpose why such a custodian is communicated with, who the other collaborators are; and what is being discussed on those channels in order to determine whether that particular collaboration platform should be searched for that custodian for relevant ESI.

While a custodian might create a document, it does not necessarily follow that the same custodian contributed to the relevant substance of the document. Knowing how several collaborators worked to create a final product may lead to the individual with the knowledge most relevant to the issues at stake.

Certainly, organizing electronic discovery around custodians is not going away. Yet, the challenge presented by collaboration tools and modern communication is knowing how and with whom custodians are collaborating without allowing such knowledge to create disproportionate discovery. Defining “Custodian.”

At its most basic level, the term custodian as it is used in eDiscovery refers to an individual who it is reasonably believed possesses information relevant to the issuing being litigated. For example, a custodian may be “employees and representatives who have information relevant to the asserted claims and potential defenses.”²⁰ Or custodians may be described as “those inside the organization who create, receive, and store their own information (i.e., individual custodians).”²¹

Other eDiscovery resources have defined custodian similarly. The Electronic Discovery Reference Model (“EDRM”) defines custodian as a “[p]erson having administrative control of a document or electronic file; for example, the data custodian of an email is the owner of the mailbox which contains the message.”²²

²⁰ The Sedona Conference, *Primer on Managing Electronic Discovery in Small Cases*, 24 SEDONA CONF. J. 93, 111 (2023).

²¹ The Sedona Conference, *The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production*, 19 Sedona Conf. J. 1, 120 (2018). [Consider whether the Sedona Conference should update its definition of custodian in its glossary.]

²² EDRM does not seem to have its glossary online. This definition was found here: <https://percipient.co/what-is-an-esi-or-data-custodian/> (last visited June 15, 2023). See also Madhava, Rakesh, *Ediscovery Glossary: 30 Terms Every Litigator Should Know*, Nextpoint https://www.nextpoint.com/ediscovery-blog/ediscovery-terminology-key-terms-glossary/?utm_source=jdsupra&utm_medium=referral&utm_campaign=blog&utm_content=ediscovery_glossary (last visited June 15, 2023); Michael I Quartaro, *Project Management in Electronic Discovery* 356 (2d ed., eDiscoveryPM.com LLC 2021) (defining “Custodian” as “[a]n individual who is likely to create, store, or have documents and ESI relevant to a litigation or

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Apart from the dictionary definition of collaborator as “someone who works with another person or group,”²³ a definition of collaborator for eDiscovery purposes has not been located.²⁴ While any definition of collaborator may depend on the particular collaboration platform at issue, a working definition may be that a “collaborator” is an individual with permission to participate in communication or to access ESI but who may or may not have possession, control, or custody over the ESI.²⁵

Courts have held that ESI created or stored in collaboration tools is discoverable, although there may be issues regarding access and proportionality that may not justify the collection and production of such ESI in each case. *See e.g., Lubrizol Corp. v. IBM*, 2023 U.S. Dist. LEXIS 84927, *7 (N.D. Ohio May 15, 2023) (collecting cases where courts order the preservation or production of Slack ESI); *Benebone LLC v. Pet Qwerks, Inc.*, 2021 U.S. Dist. LEXIS 43449, *6 (C.D. Cal. Feb. 18, 2021) (“Here, because Benebone uses Slack as part of its internal business communications, there is no real dispute that Benebone's Slack messages are likely to contain relevant information.”). Therefore, it is incumbent on the eDiscovery practitioner to understand the particular collaboration platforms in use, how they are being used, and who is using them.

At the preservation stage, parties should cast their net widely to avoid permanent loss of relevant information. *See, e.g., The Sedona Conference, Commentary on Proportionality in Electronic Discovery*, 18 SEDONA CONF. J. 141, 150 (2017) (“In particular, at the preservation stage parties should be wary of applying too narrow a definition of what constitutes relevant ESI.”); *see also In re Google Play Store Antitrust Litig.*, 2023 U.S. Dist. LEXIS 53218, *24 (N.D. Cal. March 28, 2023) (“Approximately 360 individuals are subject to the legal hold for this case, about 40 of whom have been designated as custodians.”); *Twitter, Inc. v Musk*, 2022 Del.Ch. LEXIS 219 (Del. Ch. September 7, 2022) (parties disputed whether the term “Messaging Platform Custodians” as used in agreed upon ESI protocol included Slack for purposes of resolving dispute over how many individual Slack accounts would be searched). The may require the preservation of much more ESI than will eventually be collected, reviewed, and produced.

Collaboration platforms with often provide some combination of records, logs, or metadata to track access and editing of documents or communications, making it easier to discover previously unidentified custodians. This information can provide insight into who other possible custodians may be. These tools can be particularly helpful in the early stages of litigation when the parties may lack a full picture of the scope of discovery. For this reason, the parties may want to proceed in phases to first get an idea of the scope of ESI contained within collaboration platforms (maybe with limited number of custodians or some kind of sampling). *See The Sedona Conference, Commentary on*

investigation in their possession, custody, or control. Often referred to an ‘user’ by IT personnel; sometimes referred to as ‘source’ by legal review teams.”).

²³ “Collaborator.” Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/collaborator>. Accessed 7 Jul. 2023.

²⁴ Neither The Sedona Conference nor EDRM define collaborator.

²⁵ This is my attempt at a definition.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Proportionality in Electronic Discovery, 18 SEDONA CONF. J. 141, 156-65 (2017) (noting that “[i]n the early stages of litigation, application of the proportionality factors may be complicated by the parties’ and the court’s lack of information” and discussing phased discovery and sampling).

The records, logs, and metadata provided by collaboration tools may offer insight about the ESI itself, particularly if this information reveals redlines and comments made by collaborators; when and who made edits to the ESI or deleted ESI; what individuals said about relevant documents in chats; and possibly custom organizational tags that may have been attached to the ESI by the creator. *See, e.g.*, The Sedona Conference, *Commentary on ESI Evidence & Admissibility, Second Edition*, 22 SEDONA CONF. J. 83, 139-140 (2021) (discussing admissibility of ESI collected from collaboration platforms, including metadata available that can authenticate content and potential issues with collecting data through an API); *In re Google Digit. Adver. Antitrust Litig.*, 2023 U.S. Dist. LEXIS 49016 (SDNY March 17, 2023) (including in the decision a list of metadata attached as appendix to the decision includes “Drive Collaborators” which is defined as “Users with access to edit a Google Drive document.” Metadata may also play role in connecting ESI to its source for authentication purposes. Without the metadata, authentication may pose a problem where many collaborators contribute to a document or participate in a chat.

Thus, identifying all the relevant custodians and collaborators may require an iterative approach. Parties may want to adopt flexible ESI protocols that allow for identification of additional relevant custodians after ESI is produced. *See, e.g.*, The Sedona Conference, *Commentary on Proportionality in Electronic Discovery*, 18 SEDONA CONF. J. 141, 155 (2017) (“Weighing the accessibility and associated expense and burden of discovering relevant information, as well as the discovery needed in a given case, requires a nuanced and often iterative approach.”). This may be particularly true with regard to collaboration platforms and modern communications. Additional custodians may be identified by reviewing who collaborated on a key document or who is communicating in relevant chats and channels as well as the substance of those communications.

It may be useful to understand how the collaboration tool is being used. For example, most collaboration tools allow users to collaborate on documents and to communicate in groups or one-on-one (i.e., Teams, Google), some tools are primarily communication tools (i.e., Discord)²⁶, while still others are communication and file sharing tools that integrate other cloud-based document management systems (i.e., Slack). Knowing at the outset what the capabilities are of a given platform will help to define the scope of discovery.²⁷

²⁶ *See, e.g., Carty v. Steem Monsters Corp.*, 2022 U.S. Dist. LEXIS 209305 (E.D. Pa. Nov. 18, 2022) (Discord channels recognized as a communication channel requiring preservation).

²⁷ Google Chat options include one-on-one chats; group chats for 3 or more people, and “topic- or project-based” rooms. *See In re Google Play Store Antitrust Litig.*, 2023 U.S. Dist. LEXIS 53218, *22 (N.D. Cal. March 28, 2023). Slack options include one-on-one direct messages, group messages, private channels, and public channels. Teams has meeting chats, shared channels, and messages stored in personal mailboxes or mailboxes created for particular channels.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

The scope of discovery from collaboration platforms may also be guided by the type of license the enterprise has acquired for use of the collaboration platform. For example, certain licenses may not provide a party with possession, custody, or control of ESI. *See, e.g., Laub v. Horbaczewski*, 2020 U.S. Dist. LEXIS 247102, *11-13 (C.D. Cal. Nov. 17, 2020) (concluding that defendant did not have possession, custody, or control over private Slack channels under the free version and standard version of Slack). Other licenses may have limited retention periods or a default retention period. *See In re Google Play Store Antitrust Litig.*, 2023 U.S. Dist. LEXIS 53218, *22 (N.D. Cal. March 28, 2023) (noting that Google Chat’s standard retention period for one-on-one chats with “history” turned off is 24 hours. In contrast, the retention period for one-on-one chats with “history” option turned on is 30 days while group chats with the “history” option turned on is 18 months).

Collaboration tools often provide for different collaborators to have different levels of access or permission, which can be useful to establish whether a custodian has the ability to create, edit, and/or delete a document.²⁸ Some collaboration platforms (such as Slack and Teams) allow users to delete or edit messages after they are sent (in some cases even if the system is set to retain messages). For these reason, understanding the information governance protocols of the enterprise will assist in what backup or redundancy is in place to preserve ESI. “Preservation in Place” technology automatically preserves data even if a custodian deletes it or turns off auto-preserve features.

Even when a collaborator is granted very robust permissions, that individual may only be passively participating in a collaborative environment. For example, an individual could have the right to make edits to a document but hasn’t in fact made any edits. Likewise, an individual could be the member of a group chat channel but not actually communicate on the channel. On the other hand, if an issue in the case relates to the custodian’s state of mind or knowledge, having even passive access to certain documents or communications may weigh in favor of preserving, collecting, reviewing, and producing the ESI (assuming the other proportionality factors similarly weigh in favor of production of the ESI).

Finally, in identifying who is a custodian, the following should be considered:

- Determine if participants in chats are using “handles” instead of their true names that may hide or obscure their identities.
- Carefully review proposed ESI Protocol to make sure that the definition of “custodian” or “custodial sources” used in the ESI protocol does not create unintended custodians and consider narrowing the definition of custodian by level of access or permission and custodial sources to identified repositories.
- Determine if legal counsel is participating in any of the chats with the custodian, or contributing to a shared document, and, if so, if any privilege has been waived.
- Determine if collaborative ESI has been shared with or edited by third parties.

²⁸ Some enterprises employ the “principle of least privilege,” which is a computer security concept that limits users’ access rights to only what is required to do their jobs.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

- Determine what happens to ESI from collaborative platforms when an employee leaves the company.

2. Hyperlinks in Collaboration and Communication Platforms

A “hyperlink” is a link in an email or document that points to another location. Hyperlinks can be similar to attachments but they also have material characteristics that are substantively different, which necessitates a different analysis with respect to many e-discovery issues. Hyperlinks can also be embedded objects but not all embedded objects are hyperlinks. Thus, understanding the e-discovery issues related to hyperlinks requires first an explanation of what they are and how they compare and contrast with attachments and embedded objects. This *Commentary* then considers some of the issues and practice pointers that practitioners should consider when addressing hyperlinks throughout the e-discovery process, including relevance and proportionality.

There is some overlap conceptually and even technologically between hyperlinks, embedded objects, and attachments. This can cause confusion, however, in the context of e-discovery if a precise understanding of their differences is lacking. Accordingly, each of these terms is examined further below. *The Sedona Conference Glossary: eDiscovery and Digital Information Management*, Fifth Edition, 21 SEDONA CONF. J. 263 (2020) (the “Sedona Glossary”), provides the following definitions:

- **Attachment:** A record or file *associated with* another record *for the purpose of* retention, transfer, processing, review, production, and routine records management. There may be multiple attachments associated with a single “parent” or “master” record. In many records and information management programs 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 often refers to a file (or files) *associated with* an email *for retention and storage* as a single message unit. See Document (or Document Family); Message Unit; and Unitization.²⁹ (Emphasis added.)
- **Embedded Object:** A file or piece of a file that is copied into another file, often retaining the utility of the original file’s application; for example, a part of a spreadsheet embedded into a word processing document that still allows for editing and calculations after being Embedded.³⁰
- **Hyperlink:** A *pointer* in a hypertext document—usually appearing as an underlined or highlighted word or picture—that, upon selection, sends a user to another location either within the current document or to another location accessible on the network or internet.³¹ (Emphasis added.)

²⁹ Sedona Glossary at 270-271.

³⁰ *Id.* at 305.

³¹ *Id.* at 319.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Based on the above definitions, an attachment is a record that meets two basic requirements – it must be “associated with” another record and that association must be for a substantive purpose (e.g., retention and storage). Sedona defines “attachment” separately from “hyperlink,” which is characterized as a “pointer” to another data location. The similarities and differences between attachments and hyperlinks are discussed in detail in the next section.

Then there are embedded objects, which “often” but not always retain the utility of the original file’s application. Whether embedded objects should be extracted and treated like separate documents and/or attachments is a point typically negotiated by parties in their ESI protocols and often depends on the perceived substance of the object. For example, a party may request that spreadsheets³² embedded in other documents be extracted as separate documents and treated like attachments to the documents, in part so that the spreadsheets can be manipulated in their native application. However, parties often agree that other embedded objects like logos or an email sender’s “vcard” need not be extracted and may be produced either within the documents themselves or not at all. Embedded objects may or may not involve the types of files and data that are the subject of this *Commentary*, which is focused on modern collaboration tools and communication platforms.³³ Accordingly, embedded objects as a general topic is not discussed further but the specific issue of hyperlinks is.

a. Comparing Hyperlinks with Attachments.

Attachments are associated with their parent document for a substantive purpose. They typically are considered part of the same “message unit” as their parent emails.³⁴ Attachments and their parents are also considered to be part of the same “document” or “document family.”³⁵ As one court has noted, attachments are a “necessary” part of the overall communication or document.³⁶ Technologically, attachments are also generally part of the same container or storage file as their parents, which makes their preservation and collection during the discovery process relatively easy because entire document families can be handled together. This can be on a singular level, where an email attachment is part of

³² *See, generally*, United States ex rel. Martin v. Life Care Ctrs. of Am., Inc., Case No. 1:08-cv-251; Case No. 1:12-cv-64, 2015 WL 10987073, at *9 (E.D. Tenn. Aug. 31, 2015).

³³ *See* Nichols v. Noom Inc., 20-CV-3677 (LGS) (KHP), 2021 WL 948646, at *__ (S.D.N.Y. March 11, 2021) (“While the protocol does reference ‘files with extracted embedded OLE documents,’ the Court understands this to refer to embedded, displayed documents such as a graph or a chart within a Word document or email—not hyperlinked documents.”).

³⁴ *See* Sedona Glossary at 337 (defining “message unit” as an “email and any attachments associated with it”).

³⁵ *Id.* at 299 (defining a “document” and “document family” to be a “collection of pages or files produced manually or by a software application, constituting a logical single communication of information, but consisting of more than a single stand-alone record”). The Sedona Glossary also defines an “email message” to be a “file created or received via an electronic mail system” but that any “attachments that may be transmitted with the email message are not part of the email message but are part of the Message Unit and Document Family.” *Id.* at 304.

³⁶ *See* Nichols, 2021 WL 948646, at *__ (“When a person creates a document or email with attachments, the person is providing the attachment as a necessary part of the communication.”)

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

the same .msg file as its parent, or on a broader level. For example, collecting a custodian's PST file will include all of their available emails and attachments.³⁷ Thus, from both a sender's intent and a technological perspective, an attachment is part and parcel of its parent.³⁸

In contrast, hyperlinks are not necessarily intended by the sender to be associated with, or a necessary part, of the overall communication or document. As explained in *Nichols v. Noom Inc.*, 20-CV-3677 (LGS) (KHP), 2021 WL 948646, *__ (S.D.N.Y. March 11, 2021):

When a person creates a document or email with a hyperlink, the hyperlinked document/information may or may not be necessary to the communication. For example, a legal memorandum might have hyperlinks to cases cited therein. The Court does not consider the hyperlinked cases to be attachments. A document also may contain a hyperlink to another portion of the same document. That also is not an attachment. A document might have a hyperlink shortcut to a SharePoint folder. The whole folder would not be an attachment. These are just examples. An email might have hyperlinks to a phone number, a tracking site for tracking a mailing/shipment, a facebook page, a terms of use document, a legal disclaimer, etc. The list goes on and on.

A primary reason people use hyperlinks, as opposed to attachments, is to facilitate work collaboration.³⁹ As explained in *Shenwick v. Twitter*, No. 16-CV-05314 (JST)(SK), 2018 U.S. Dist. LEXIS 189263, at *__ (N.D. Cal. Sept. 17, 2018):

Defendant Twitter, Inc. ("Twitter") uses a Google Suite ("GSuite") environment in which individuals can work simultaneously on the same documents lodged in Twitter's system. In different electronic messages, individuals referred to documents stored in the GSuite environment, via a hyperlink. After receipt of the electronic mail message, a recipient accesses the referenced document via the hyperlink, and sender and

³⁷ See, generally, *Karsch v. Blink Health Ltd.*, 17-CV-3880 (VM) (BCM), 2019 WL 2708125, at *__ (S.D.N.Y. June 20, 2019) ("A PST file, or personal storage table (.pst) file, is a Microsoft Outlook Data File that stores a user's Outlook data for POP3, IMAP and web-based mail accounts, including all mail folders and the items within the folders, such as emails, email attachments, to do items and appointments, contacts and more." (citing Personal Storage Table (PST), Wikipedia, https://www.wikipedia.com/TERM/P/personal_storage_table_pst.html (last visited June 20, 2019))); *ComRent Int'l, LLC v. Thomson*, Civil Action No. RDB-20-3757, 2021 WL 1733471, at *__ (D. Md. May 3, 2021) ("A '.pst' file is an open proprietary file form used to store copies of messages, calendar events, and other items within Microsoft software such as Microsoft Outlook. These files allow a user to create a locally stored copy of the emails on a laptop or USB device. This specific '.pst' file contained more than 900 emails with numerous attachments, including files and spreadsheets . . .").

³⁸ See, generally, *Judge Rotenberg Educ. Ctr., Inc. v. United States FDA*, 376 F. Supp. 3d 47, __ (D.D.C. 2019) ("While emails and their attachments are not per se a single record, at a minimum 'attachments should reasonably be considered part and parcel of the email by which they were sent' when the email 'make[s] explicit reference to, or include[s] discussion of, the missing attachments.'").

³⁹ Another reason for using hyperlinks include avoiding email size limits that may be triggered with large attachments.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

recipient can then modify the referenced document, which is stored centrally so that more than one person can access it.

If a document is sent as an attachment then the recipient now “owns” that attachment - if any edits are made to it, no one else will see those edits unless and until the person proactively sends the edited document to someone else. If multiple people need to review and edit the document, it can result in the laborious process of having to carefully name each version and circulate them and people have to take turns inputting their changes. The use of hyperlinks, especially when coupled with a document platform like Google Docs, which allows multiple people to edit a document at the same time, avoids these issues.

From a technical perspective, hyperlinks are not stored directly with their parents - they merely point to other sources of data.⁴⁰ These sources can include documents where ongoing collaboration and editing is taking place or a Wiki⁴¹ site where visitors can add or edit content. In many respects then, hyperlinks are “active” or “evolving” data sources, whereas attachments are “static.”⁴² This fact has led several cases to note that courts will not accept hyperlinks as attachments for purposes of court filings.⁴³

⁴⁰ See May 9, 2023 Declaration of Russel Brown, Senior Forensic Consultant at Epiq eDiscovery Services, Inc., at Para. 8, from the In re Stubhub Refund Litig., Case No. 20-md-02951-HSG (TSH), 2023 U.S. Dist. LEXIS 74007 (N.D. Cal. April 25, 2023) (Proceeding Number 202-4): “Hyperlinked documents and data are not e-mail attachments. A document attached to an e-mail is contained within the e-mail data itself, and the attachments are collected with the e-mail in the standard course of an eDiscovery e-mail collection. Hyperlinked documents, however, are simply website addresses stored within the e-mail body. The actual document is not contained within the e-mail itself. The hyperlink may reference a document, a form, or network storage location (folder) containing numerous other documents and subfolders. In the case of a network storage location, the location itself, with potentially thousands or hundreds of thousands of files (or more) could never be physically attached to an e-mail.”

⁴¹ The Sedona Glossary defines “Wiki” as, “A collaborative website that allows visitors to add, remove, and edit content.” Sedona Glossary at 389.

⁴² See, generally, Shenwick at *___ (“Defendants claim that the hyperlinks are not like attachments to electronic mail messages and argue that producing an attachment — a static document — is relatively simple, but producing a document referenced in a hyperlink — an evolving document — requires a multi-step process by a human being.”).

⁴³ See, e.g., Ramos v. Taylor, 1:20-CV-1256-RP, 2022 U.S. Dist. LEXIS 227519 (W.D. Tex. Dec. 18, 2022) (“A hyperlink to a webpage does not qualify as an attached document, so there is no indication that Taylor's videos are attached to the complaint.” (citations omitted)); Sturgeon v. PharMerica Corp., 438 F. Supp. 3d 246, 258 n. 72 (E.D. Pa. 2020) (“It is obviously problematic to take judicial notice of materials found online based only on hyperlinks. [] The concern is not so much that the documents themselves might change, but that their location on the internet might change—as it apparently has here—leaving unclear to future readers what was in the record before the court. The best practice would be for a party seeking judicial notice to both attach copies as exhibits to the motion for judicial notice and hyperlink to the website where the material can be found.”); Am. Coastal Ins. Co. v. Electrolux Home Prods., Case No.: 2:19-cv-180-FtM-38MRM, 2019 U.S. Dist. LEXIS 175286, n. 1 (M.D. Fla. Oct. 9, 2019) (“Disclaimer: Documents filed in CM/ECF may contain hyperlinks to other documents or websites. These hyperlinks are provided only for users' convenience. [] The Court accepts no responsibility for the availability or functionality of any hyperlink.”).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Thus, when comparing attachments to hyperlinks, from both a substantive and technological perspective, hyperlinks are less connected to their parents.⁴⁴ Accordingly, referring to hyperlinks as “modern attachments” is a misnomer.⁴⁵ To the extent an alternative name is useful for referring to hyperlinks, this *Commentary* recommends adoption of the term “pointers,” which has been used by Sedona, commentators, and technical experts. [*Note: I think there is a demand from practitioners for clarification of terms and this Commentary is an opportunity to move the law forward here.*]

Differentiating attachments from hyperlinks does not mean, however, that hyperlinks cannot be relevant or that proportionality factors necessarily disfavor the preservation, collection, review, and production of them. To the contrary, hyperlinks may point to important data directly relevant to key claims and defenses in a case. The differences between attachments and hyperlinks simply means that the relevance and proportionality analysis afforded to attachments should be modified with respect to hyperlinks to reflect the differences. Whether an attachment or hyperlink is intended by the sender to be associated with, or a necessary part, of the communication or document goes to relevance, while the technological ability to preserve, collect, and produce attachments or hyperlinks goes to proportionality. Each of these issues is addressed further below.

b. Hyperlinks and ESI Protocols

Courts have not reached a consensus on whether documents contained at hyperlinks or imbedded links must be produced, therefore parties should consider whether to address production of such information in an ESI protocol.

Even within the same district, different conclusions. In the Southern District of New York, embedded metadata including internally linked files and hyperlinks (at least when related to spreadsheets) has been deemed “generally discoverable” and “should be produced as a matter of course” because this type of metadata is often “crucial to understanding an electronic document.” Aguilar v. Immigration and Customs Enforcement Div. of U.S. Dept. of Homeland Sec., 255 F.R.D. 350 (S.D.N.Y. 2008); However, in a more recent decision from the same district involving discovery of Google Drive documents, the court did not consider hyperlinks to be an attachment because many hyperlinks are not a necessary part of the communication (for example, links to Facebook pages, phone numbers,

⁴⁴ See Nichols, 2021 WL 948646, at *__ (“[T]he Court does not agree that a hyperlinked document is an attachment. While the Court appreciates that hyperlinked internal documents could be akin to attachments, this is not necessarily so.”).

⁴⁵ The term “modern attachments” originated with Microsoft but it is not an accurate description of hyperlinks based on Sedona’s definitions of “attachments” and “hyperlinks” and how courts and technical experts have interpreted them. See, e.g., May 9, 2023 Declaration of Julie Lewis, CEO and Founder of Digital Mountain, Inc., at Para. 6, from the In re Stubhub Refund Litig., Case No. 20-md-02951-HSG (TSH), 2023 U.S. Dist. LEXIS 74007 (N.D. Cal. April 25, 2023) (Proceeding Number 202-5): (“The term modern attachment is misleading. The existence of linked files within collaboration software is not new. Lotus Notes had used the “link” concept prior to newer systems. This link is also not, however, an attachment. Rather, it is a pointer to somewhere else.”). A February 20, 2023 Microsoft article on its “Learn” website refers to hyperlinks as “cloud attachments” but it is unclear if this is how Microsoft solely intends to refer to them going forward. See <https://learn.microsoft.com/en-us/microsoft-365/compliance/ediscovery-cloud-attachments?view=o365-worldwide> (last visited June 29, 2023).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

etc.), whereas attachments are a necessary part of the communication. Nichols v. Noom, Inc., 2021 WL 948646 (S.D.N.Y. Mar. 11, 2021).

The Northern District of California took a “split the baby” approach, allowing the plaintiffs to identify up to 200 hyperlinked documents they wanted produced from an original request for 725 hyperlinks that were referred to in Google Suite messages via the hyperlink. Shenwick v. Twitter, Inc., 2018 WL 5735176, at *1 (N.D. Cal. Sept. 17, 2018). In contrast, the District Court for the District of New Jersey ordered a defendant to produce 2,200 Google Drive documents that were referenced in emails by hyperlink, even though the linked documents were not stored within the emails in the ordinary course of business. IQVIA, INC. v. Veeva Systems, Inc., 2019 WL 3069203, at *6 (D.N.J. Jul. 11, 2019). There, the court warned the defendant that, if after searching for the documents and attempting to re-link them, the defendant intended to assert that the documents were deleted in the ordinary course of business, the defendant would need to explain when the documents were deleted and how it determined this information. *Id.*

Given the uncertainties as to how courts will treat discovery of linked documents in collaborative platforms, some parties have proactively addressed this issue in their ESI agreements. For example, where StubHub claimed it ran search terms on Google Drive and produced responsive documents but failed to preserve the parent-child relationship as required by its ESI Protocol, it was ordered to produce the documents as required by the ESI protocol and if it was unable, to produce a corporate representative with full knowledge of everything StubHub and its vendors did in an attempt to produce linked documents as attachments.” In re StubHub Refund Litigation, 2023 WL 3092972, at *2 (N.D. Cal. Apr. 25, 2023).

The court admonished, “Let's get back to basics: Litigants should figure out what they are able to do *before* they enter into an agreement to do something. Litigants should live up to their agreements, especially when they are embodied in court orders, as the ESI Protocol is here. And if for some reason, a party learns that a so-ordered discovery agreement has become impossible to comply with, the party should promptly move for relief, with a good showing that despite its best efforts, compliance is impossible.” *Id.* An ESI order has also protected a defendant where the plaintiff sought production of hyperlinked documents referenced in automated emails from software applications. Deibler v. SanMedica International, LLC, 2021 WL 6198062, at *14 (D.N.J. Dec. 30, 2021). There, the court determined that the defendants’ failure to include hyperlinked documents was not inconsistent with the effective ESI protocol. *Id.*

c. Relevance and Hyperlinks

There are two primary Federal Rules affecting the analysis of relevance as to attachments and hyperlinks. First, Federal Rule of Civil Procedure 26(b)(1) provides that “Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case” Fed. R. Civ. P. 26(b)(1). Second, Federal Rule of Evidence 106, often called

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

“the rule of completeness,”⁴⁶ provides: “If a party introduces all or part of a writing or recorded statement, an adverse party may require the introduction, at that time, of any other part — or any other writing or recorded statement — that in fairness ought to be considered at the same time.” Fed. R. Evid. 106. As explained in *Abu Dhabi Commer. Bank v. Morgan Stanley & Co.*, 2011 U.S. Dist. LEXIS 95912, *19-23 (S.D.N.Y. Aug. 18, 2011), *adopted without objection*, 2011 U.S. Dist. LEXIS 94727, 2011 WL 3734236 (S.D.N.Y. Aug. 24, 2011):

Conceptually, there is a good basis for considering each item (each e-mail and each attachment) separately. Relevance is the sine qua non of discovery. *See* Fed. R. Civ. P. 26(b)(1). If information is not relevant, it is not discoverable under plain text of the Rule. Thus, if an e-mail attaches three disparate items in one communication package, that does not mean that all three items relate to the same thing or would be equally relevant to a discovery request.

At the same time, the “completeness” standard of Fed. R. Evid. 106 states that when part of a document is introduced into evidence, the entire document may be required if it “ought in fairness” be considered contemporaneously. The logic of this evidentiary rule extends backwards to discovery which often leads to a conclusion (or at least a presumption) that if something was attached to a relevant e-mail, it is likely also relevant to the context of the communication. In addition, harkening back to the days of paper discovery, communications and documents that were attached contemporaneously (as with a staple through all pages) were often treated as a single object for relevance assessments.⁴⁷

Thus, although certain attachments may not be relevant on their face, there often is a presumption they are relevant because they were intended to be included for a substantive reason and, therefore, are likely to provide helpful context regarding the parent. This presumption does not exist for hyperlinks in general because they are often not intended to be a necessary part of the communication or document.

d. Identification And Preservation and Hyperlinks

There are several Sedona Principles relevant to identification and preservation of hyperlinks. An important first step is to take reasonable steps to preserve relevant evidence once the duty to preserve

⁴⁶ *See* Fed. R. Evid. 106, Notes of Advisory Committee on Proposed Rules (“The rule is an expression of the rule of completeness.”); **ADD CITE**

⁴⁷ The rule of completeness does have limits, however. “The rule does not . . . require introduction of portions of a statement that are neither explanatory of nor relevant to the admitted passages.” *United States v. Tahlil Mohamed*, 18-cr-603 (ARR), 2022 WL 15493545, *__ (E.D.N.Y. Oct. 26, 2022) (internal quotations and citation omitted).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

is triggered.⁴⁸ Then, as soon as practicable in a case, “parties should confer and seek to reach agreement regarding the preservation and production of electronically stored information.”⁴⁹ These “early discussions should include procedural issues relating to form of production”⁵⁰ as well as an assessment of the scope of a preservation duty by considering not just persons likely to have relevant ESI, but also non-custodial sources⁵¹ and shared data/areas like public folders and network folders.⁵² But “[p]reservation efforts need not be heroic or unduly burdensome”⁵³ and “[r]esponding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.”⁵⁴

An important distinction when it comes to the preservation of hyperlinks is, assuming a preservation duty exists, whether a party has a duty to preserve relevant targets of a hyperlink at their source versus preserving the link and its target together and as they existed at the time the hyperlink was first used. Barring proportionality concerns, possession, custody, or control issues, or other special circumstances, a party’s obligation as to the former is clear. It is less so with respect to the latter.

It generally is more difficult to identify, preserve, and collect hyperlinks, as compared to attachments.⁵⁵ Defendants in *Shenwick* explained their multi-step process:

The steps required are locating the document containing the link, clicking through the link to the source file, determining the file owner's identity if a passcode is required and obtaining that passcode, manually identifying the date-stamped version of a linked GSuite document that corresponds to the referring electronic mail message, capturing the data in a manner that minimally affects the metadata, exporting the data to the vendor for processing, and producing the data in a manner that matches it to the referring electronic mail message.⁵⁶

⁴⁸ Sedona Principle 1; The Sedona Conference, Commentary on Legal Holds, Second Edition: The Trigger & The Process, 20 SEDONA CONF. J. 341 (2019).

⁴⁹ Sedona Principle 3; Fed. R. Civ. P. 26(f)(2) and (3); Fed. R. Civ. P. 16(b)(3)(B)(iii).

⁵⁰ Comment 3.c. to Sedona Principle 3.

⁵¹ Comment 5.c. to Sedona Principle 5.

⁵² Comment 5.i. to Sedona Principle 5.

⁵³ Comment 5.e. to Sedona Principle 5.

⁵⁴ Sedona Principle 6.

⁵⁵ See *Porter v. Equinox Holdings, Inc.*, Case No. RG19009052, 2022 WL 887242, at *__ (Cal. Super. Mar. 17, 2022) (“[L]inked documents can present unique challenges that make them different from email attachments.”).

⁵⁶ *Shenwick* at *__.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

The declaration of a forensic consultant in another case noted the following challenges based on the technology available that party today:

- The lack of any pre-existing tools that can identify all possible hyperlinks in a collection and connect them with the linked source at the time of collection or after;
- The inability to programmatically distinguish between links to public sources versus internal documents; and
- After collection, the need for an e-discovery vendor to determine a method in which it could first identify any hyperlinks in the text of an email, then what they point to, and finally how to associate the target of the hyperlink back to the source email.⁵⁷

There can be other challenges with hyperlinks as well:

- Links may be broken or non-functional⁵⁸;
- Targets of links may no longer exist or access to them may have been revoked⁵⁹;
- Links may point to external sources outside of an organization's possession, custody, or control;
- A link may point to an entire shared folder with voluminous⁶⁰ and/or non-responsive data; and
- If an organization has archived their email in a third-party application that does not allow for the preservation or collection of hyperlinks.

⁵⁷ See May 9, 2023 Declaration of Russel Brown, Senior Forensic Consultant at Epiq eDiscovery Services, Inc., at Para. 9, from the *In re Stubhub Refund Litig.*, Case No. 20-md-02951-HSG (TSH), 2023 U.S. Dist. LEXIS 74007 (N.D. Cal. April 25, 2023) (Proceeding Number 202-4).

⁵⁸ See Declaration of Julie Lewis at Para. 6.

⁵⁹ See *Shumway v. Wright*, No. 4:19-CV-00058-DN-PK, 2020 WL 1037773, at *__ (D. Utah Jan. 13, 2020), *report and recommendation adopted*, 2020 WL 1038152 (D. Utah Jan. 29, 2020) (“The Court's technical experts were stymied in their ability to identify most of the Google Drive linked documents referenced in the aforementioned 419 emails because those documents were inaccessible. For the majority of the inaccessible documents, the Court's technical experts received an ‘Access Denied’ message from Google. As the Special Master understands it, the ‘Access Denied’ message indicates the Google Drive linked documents are no longer shared with Defendant Slavens. In addition, there was a small number of Google Drive linked documents that either no longer exist or are located in the ‘trash’ folder of the user who shared the linked document with Defendant Slavens. By placing a Google Drive linked document in the ‘trash’ folder, the user made that document inaccessible to Defendant Slavens’ Google account.”); *see also* Porter, 2022 WL 887242, at *__ (identifying a challenge with hyperlinks as when users “have made linked documents inaccessible by revoking access to files, thus disabling the responding party’s ability to collect that information.”) (citing *Shumway*, 2020 WL 1037773, at *2)).

⁶⁰ *Id.* (“Also, these links or pointers can reference an entire voluminous drive versus just a file (e.g., if there is a reference to a 1 Terabyte drive, this can be equivalent to approximately 38,000 banker boxes of documents in just a single link).”).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

A particularly problematic issue related to the identification and preservation of links is versioning, which is when an electronic record changes in some way to create a new version of it.⁶¹ These changes can include modifications to file format, metadata, or content.⁶² Collaboration platforms and the use of hyperlinks can result in a target document that has undergone numerous changes or edits from the time the hyperlink is originally used to when ESI is being preserved or collected for a case. Having to preserve every version of every document can create a burden for the producing party. In addition, for “active” documents undergoing changes, “matching the version at the time of the email can be difficult or impossible to resolve depending on how the data is stored.”⁶³ Furthermore, it may not even be possible to preserve the version of a document as it existed at the time the hyperlink was first used because that version may have been deleted prior to when a party’s duty to preserve was triggered. Thus, one of the very reasons organizations utilize hyperlinks, to foster collaboration on documents, also presents one of the greatest challenges in discovery.

Whether ESI is reasonably accessible is not based solely on “its source or type of storage media. Inaccessibility is based on the burden and expense of recovering and producing the ESI and the relative need for the data. Whether data are not reasonably accessible due to undue burden or cost will depend on the facts of the case.”⁶⁴ However, parties should not “object to the discovery of ESI on the basis that it is not reasonably accessible unless the objection has been stated with particularity, and not in conclusory or boilerplate language.”⁶⁵ To avoid disputes later, parties should exchange information on relevant data sources, including those not being searched, and identify forms of production early.⁶⁶ To this end, consider discussing the topic at the Rule 26(f) conference. Based on the identification and preservation challenges hyperlinks may present, it behooves parties to discuss hyperlinks and attempt to reach agreement as soon as practicable in a case.

e. Search, Collection, Processing and Hyperlinks

(1) Complexity and Technology of Hyperlinks

The process of searching for, collecting, and processing hyperlink data is more complicated than it is for attachments. The level of complexity can vary, however, depending on the type of platform at issue and its capabilities. Another variable is that these capabilities may depend on the particular license

⁶¹ See Sedona Glossary at 387 (defining “version, record version”).

⁶² *Id.*

⁶³ See Declaration of Julie Lewis at Para. 4.

⁶⁴ Rules of the Commercial Division of the Supreme Court of New York, Appendix A, Guidelines for Discovery of Electronically Stored Information.

⁶⁵ *Id.*

⁶⁶ The Sedona Conference Cooperation Proclamation at 2.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

an organization has obtained. For example, “[a]round August 2022, Microsoft introduced the capability to gather links if the proper licensing exists in M365 with its Purview offering. However, there are limitations with Purview that are similar to Google due to permission issues.”⁶⁷ In addition, “document repositories such as Box, Dropbox, Google Drive, and others, to the extent they exist, cannot be accessed using Purview without separate authentication which is a manual and burdensome effort.”⁶⁸

The extent to which an organization may export data out of Slack for discovery purposes depends on the type of plan it has. Different plans, at different costs, allow access to certain channels and/or conversations.⁶⁹ Regarding links in particular, all of the plans allow a user to export them from public channels “but not the files themselves.”⁷⁰ [Note: I’m no expert in Slack so is it correct that the targets of hyperlinks cannot be exported automatically out of Slack, regardless of the plan?]

Technology is evolving to provide options for preserving and collecting hyperlinks but the licenses and plans needed to be able to do so come at a cost.

(2) Proportionality and Hyperlinks

“The responding party generally selects the technology to identify relevant information” but a if it “refuses to consider the use of an appropriate technology to reduce e-discovery burdens, even when it is reasonably available and within that party’s resources, [it] will have a difficult time making any later claim based on disproportionality or undue burden caused by that refusal.”⁷¹ Parties should discuss early in a case the tools and technology available to search, collect, and process hyperlinks. Courts are reluctant to force a producing party to upgrade their system or adopt an entirely new one

⁶⁷ Declaration of Julie Lewis at Para. 6; *see also* Microsoft Learn website, Overview of Microsoft Purview eDiscovery (Premium) (“Collects cloud-based content shared with users by use of links or modern attachments in email message and Teams chats.”): <https://learn.microsoft.com/en-us/microsoft-365/compliance/ediscovery-overview?view=o365-worldwide> (last visited June 29, 2023). But this capability is only included with Microsoft Purview, which requires a particular license like Microsoft 365 E5. *See* <https://learn.microsoft.com/en-us/office365/servicedescriptions/microsoft-365-service-descriptions/microsoft-365-tenantlevel-services-licensing-guidance/microsoft-365-security-compliance-licensing-guidance#microsoft-purview-ediscovery> (last visited June 29, 2023).

⁶⁸ *Id.*; *see also* In re Meta Pixel Healthcare Litig., 2023 U.S. Dist. LEXIS 118262, *5 (N.D. Cal. June 2, 2023) (finding limited utility of Purview to collect links “would disrupt Meta’s standardized workflow for ESI-related discovery processing”).

⁶⁹ *See, generally*, Laub v. Horbaczewski, , 2020 WL 7978227, at *__ (C.D. Cal. Nov. 17, 2020) (accepting Defendants explanation regarding the inability to access certain messages because although they upgraded from a free to paid account, the tool allowing for the export of private and direct messages “is only available to accounts on Slack’s ‘Plus’ tier or above, which Defendants do not have.”).

⁷⁰ *See* Slack Help Center page: https://slack.com/help/articles/204897248-Guide-to-Slack-import-and-export-tools#h_01EJ96AV9MF56A2RHKE5JCWE7 (last visited June 29, 2023).

⁷¹ The Sedona Conference, *Commentary on Proportionality in Electronic Discovery*, 18 SEDONA CONF. J. 141, 173-74 (2017).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

to make e-discovery easier without good cause supported by evidence.⁷² Thus, if a producing party has a system that presents challenges in associating hyperlinks with their targets, proportionality factors will require greater attention.

One factor relevant to hyperlinks is “the importance of the discovery in resolving the issues.”⁷³ A “court may limit discovery if the information sought, while relevant, is not sufficiently important to warrant its production. This issue often arises when discovery requests seek information that is duplicative, cumulative, or not reasonably accessible.”⁷⁴ The Court in *Nichols* applied this factor to hyperlinks:

Noom has argued persuasively that the redundancies of pulling hyperlinked documents would be burdensome. Indeed, one email thread may contain multiple hyperlinks to the same document that already was flagged for production. The same underlying hyperlinked document may be pulled tens if not hundreds of times in some cases. This additional collection would certainly increase the review population and, as Noom's expert explained, complicate de-duplication, delay production, and impose additional costs. [] Nor do Plaintiffs explain why a recollection of hyperlinked documents, many of which may be of no real value in the case and are redundant of the documents already collected, is proportional to the needs of this case.⁷⁵

Parties should discuss whether and to what extent the targets of hyperlinks will be produced independently (e.g., through collection of a custodian's document folders) because if they are, proportionality concerns may favor the requesting party undertaking some or all of the process of associating hyperlinks back to their parents. Similarly, if hyperlinks are pointing to public sources like news articles then the receiving party is equally situated to retrieve them. When the parties' relative

⁷² See, e.g., *Nichols* at *__ (“After fully hearing the parties' arguments, the Court held that Noom could use its preferred software to collect email documents, finding that method reasonable and deferring to the principle that a producing party is best situated to determine its own search and collection methods so long as they are reasonable. [] The Court also took into account the relative costs and delays attendant to utilizing FEC.”) (citing Sedona Principle 6); see also The Sedona Conference, *Commentary on Proportionality in Electronic Discovery*, 18 SEDONA CONF. J. 141, 174-75 (2017) (“[C]ourts should leave the choice of technological methods to the responding party so long as the methods are reasonable and appropriate to meet the needs of the case.”).

⁷³ Fed. R. Civ. P. 26(b)(1).

⁷⁴ The Sedona Conference, *Commentary on Proportionality in Electronic Discovery*, 18 SEDONA CONF. J. 141, 162 (2017).

⁷⁵ *Nichols* at *__; see also *Shenwick* at *__ (“Defendants also challenge the usefulness of this exercise, given that Defendants separately searched documents in the GSuite environment and produced relevant ones; the search will identify documents that Defendants have already produced.”). But see, *In re Google RTB Consumer Privacy Litigation*, 21-cv-02155 *YGR(N.D. Cal. April 21, 2023)(ordering production of 338 selected hyperlinks in 51 documents finding the request was reasonable and that an adequate justification was provided for those hyperlinked documents.)

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

access to the information is approximately the same, the costs to force a producing party to associate all hyperlinks with their targets likely is not justified.⁷⁶

f. Review, Production, Privilege and Hyperlinks

(1) Review and Hyperlinks

The challenges with preserving and collecting hyperlinks will spill over into the review phase of discovery.⁷⁷ For example, if a link is broken then the linked information will not be immediately (or perhaps ever) available and, therefore, cannot be reviewed. Also, if every hyperlink is collected from every custodian, the same linked document may exist hundreds of times in the review universe.⁷⁸ De-duplication is a technological means of minimizing the amount of data for review and production but it involves comparing electronic records based on their characteristics and removing exact duplicates.⁷⁹ This is typically achieved, however, by calculating a record's hash value using a mathematical algorithm, which is similar to a digital fingerprint.⁸⁰ Any differences between documents will prevent de-duplication and so this process has limitations. To the extent parties reach agreement on the production of hyperlinks, the producing party should be sure to anticipate any challenges with actually reviewing them.

(2) Production and Hyperlinks

Assuming the standards for relevance and proportionality have been satisfied, the format of production for documents is governed by Fed. R. Civ. P. 34 (b)(2)(E), which says:

- (i) A party must produce documents as they are kept in the usual course of business or must organize and label them to correspond to the categories in the request;
- (ii) If a request does not specify a form for producing electronically stored information, a party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms; and

⁷⁶ See Nichols at *__ (citing Sedona Principle 8 to say “it is appropriate to limit collection and review to non-duplicative, relevant information”).

⁷⁷ See Porter, 2022 WL 887242, at *__ (“Finally, linked documents can also create review challenges and inefficiencies given the complexity of connecting those documents to the communications which reference them.”) (citing Nichols, 2021 WL 948646, *4).

⁷⁸ See Nichols at *__ (“Indeed, one email thread may contain multiple hyperlinks to the same document that already was flagged for production. The same underlying hyperlinked document may be pulled tens if not hundreds of times in some cases. This additional collection would certainly increase the review population”).

⁷⁹ Sedona Glossary at 293 (defining “de-duplication”).

⁸⁰ *Id.* at 293, 317.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

(iii) A party need not produce the same electronically stored information in more than one form.⁸¹

For attachments, many courts have interpreted this to mean that document families consisting of parent emails and attachments should be produced together because “emails are ordinarily stored with their attachments and since an e-mail typically needs to be paired with its attachment in order for either to be fully comprehensible, those courts that have addressed this issue have held that e-mails and their attachments must be produced together.”⁸² In *Symettrica Entm't, Ltd. v. Umg Recordings, Inc.*, CV 19-1192-CJC (KS), 2020 U.S. Dist. LEXIS 263577 (C.D. Cal. July 17, 2020), the Court ordered that where the party produced responsive emails, it was also required to produce any attachments to those emails. In reaching this conclusion, the Court cited Federal Rule of Civil Procedure 34(b)(2)(E), Federal Rule of Evidence 106, and noted that “Courts in this Circuit have long recognized that an email and its attachment comprise ‘one document or message unit’ and consistently require a producing party to ‘re-link the emails with the attachments or re-produce the emails with their attachments.’”⁸³ The Court also stated that “attachments can only be fully understood when read in the context of the emails to which they are attached.”⁸⁴

The two elements favoring production of attachments with their parents – that attachments are necessary to understand their parents and the fact they are stored together – are grounded in relevance and proportionality. For hyperlinks, these factors require a different analysis because they may not be necessary and are not stored with their targets.

In *IQVIA, Inc. v. Veeva Sys.*, Case No.: 2:17-CV-00177-CCC-MF, 2019 U.S. Dist. LEXIS 115894 (D.N.J. July 11, 2019), the producing party agreed at the outset “to produce all Google Drive documents that any custodian authored, edited, reviewed, accessed, or otherwise had access to.”⁸⁵ Accordingly, the Court found there was no dispute as to relevance. The Court also stated that although “the linked documents are not stored with emails in the ordinary course of business, IQVIA has no way to link the documents, only Veeva is capable of linking the emails to the Google Drive

⁸¹ Fed. R. Civ. P. 34 (b)(2)(E); *see also* The Sedona Principles, Third Edition, Principle 12 (“The production of electronically stored information should be made in the form or forms in which it is ordinarily maintained or that is reasonably usable given the nature of the electronically stored information and the proportional needs of the case.”).

⁸² *Nguyen v. Roth & Rau AG*, Case No. CCB-06-1290, 2009 WL 10682036, at *2 (D. Md. Jul. 28, 2009); *see also* *Pom Wonderful LLC v. Coca-Cola Co.*, NO. CV 08-6237 SJO (FMOx), 2009 WL 10655335, at *___ (C.D. Cal. Nov. 30, 2009) (granting motion to compel production of missing email attachments and citing Second Edition of Sedona’s Glossary for definitions of “message unit” and “document family” to find that party receiving discovery “must have the ability to identify which attachments belong to which emails”).

⁸³ *Symettrica* at *___ (quoting *Pom Wonderful*, 2009 WL 10655335, at *3).

⁸⁴ *Symettrica* at *___.

⁸⁵ *IQVIA* at *___.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

documents.”⁸⁶ After addressing proportionality factors, the Court concluded that the producing party should, “to the extent possible,” relink the documents at issue with their parent emails.⁸⁷ Thus, a key factor was comparing each party’s ability to associate relevant hyperlinks back to their parents.

In *Shenwick*, the Court acknowledged the burden demonstrated by Defendants if forced to search for and produce documents in 725 hyperlinks identified by Plaintiffs but also noted “that Plaintiffs have a right to determine if an electronic message refers to a document” and, if so, “then Plaintiffs should be able to access that document.”⁸⁸ The Court ultimately ordered Defendants to produce documents referenced in a hyperlink for 200 documents of Plaintiffs choosing.⁸⁹ In *Nichols*, the Court noted an earlier ruling it made, whereby “if there were certain documents discovered in the production containing hyperlinks for which the corresponding hyperlinked document could not be located or identified, Plaintiffs could raise the issue with Noom and Noom would be required to provide the document or Bates number.”⁹⁰

In *Porter*, the Court acknowledged a party’s general obligation to produce ESI within the scope of discovery but that hyperlinks are “distinct” from this general obligation.⁹¹ It found *IQVIA* to be “inapposite” in addressing the issue of hyperlinks because they are not directly comparable to attachments.⁹² The Court declined to order Equinox to produce all linked documents because “such an order would be inappropriate as Plaintiffs have not advanced fact specific good cause to substantiate the production of *all* linked documents in family relationships.”⁹³ Instead, the Court opted to follow the procedure outlined in *Nichols* and held that “if there are key linked documents referenced in Equinox-produced communications that Plaintiffs cannot find, Plaintiffs shall notify Equinox accordingly and identify the particular communications referencing the linked documents in question. In response, Equinox shall identify the linked document at issue by Bates number or, if it has not been produced, Equinox shall produce the linked document without delay.”⁹⁴

⁸⁶ *Id.*

⁸⁷ *Id.*

⁸⁸ *Shenwick* at *___.

⁸⁹ *See also* U.S. v. Google, No. 1:20-CV-03010 (LCAPM) (Dist. D.C. June 17, 2022) (Docket Entry No. 361) (Minute Order: “Pursuant to the status conference held on June 17, 2022, the court orders the following with respect to further proceedings in this matter: [] (5) Google shall produce up to 200 additional linked-to documents that Plaintiffs may identify on a case-by-case basis”).

⁹⁰ *Nichols* at *___.

⁹¹ *Porter*, 2022 WL 887242, at *___.

⁹² *Id.*

⁹³ *Id.* (emphasis in original) (citing *Nichols*).

⁹⁴ *Id.* The Court also said that if disputes arise, the parties shall submit the dispute to the “Expert Advisor.”

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Finally, in *In re Meta Pixel*, the Court stated that “the ESI protocol should make clear that hyperlinked documents are not treated as conventional attachments for purposes of preserving a ‘family’ relationship in production.”⁹⁵ The Court also said, however, that for some documents it may be important “to collect (or attempt to collect) hyperlinked documents and associate them with the underlying ESI in which the links appear.” If this happens, the parties should consider “reasonable requests” on a case-by-case basis but that such requests “should not be made as a matter of routine.”⁹⁶

There are differences between how attachments and hyperlinks are “kept in the usual course of business” and/or are “ordinarily maintained.” These differences can affect whether and how they should be produced. Sedona Principle 12, Comment 12.b. cautions: “Parties should not demand forms of production, including native files and metadata fields, for which they have no practical use or that do not materially aid in the discovery process.”⁹⁷ Parties should consider whether they will be able to associate hyperlinks back to their parents and if such association is necessary. Parties should also consider whether they can reach agreement on a reasonable number of hyperlinks that the producing party will endeavor to locate and associate upon request.

(3) *Privilege and Hyperlinks*

For a producing party, duplicative or “versioned” hyperlinks can present difficulties when attempting to analyze and protect privileged information. As noted above, if every hyperlink is collected from every custodian, the same linked document may exist hundreds of times in the review universe. In addition, a single hyperlinked document may have undergone dozens of changes in content or otherwise, resulting in dozens of versions to be considered. If duplicates and/or multiple versions of a document must be analyzed for privilege, the time and cost to complete the review will increase.

3. Metadata Issues with Collaboration and Communication Platforms

Metadata, or data about data, in collaboration platforms can be more extensive than traditional stand-alone documents. The nature of collaboration platforms and the design of same makes them all different. Understanding the metadata of a Microsoft Word document framework will not be sufficient in this case. To preserve collaboration related metadata appropriately, it must first be understood on a case-by-case basis.

It may be helpful to think of Collaboration Platforms as databases which store data in component pieces that leverage Metadata to assemble what appears to an end user as a “document.” Though the Metadata specifics vary as much as they do between any databases, within any specific environment, one can expect consistency from record to record. Unlike traditional documents, however,

⁹⁵ *In re Meta Pixel* at *__.

⁹⁶ *Id.* [or add accurate page cite].

⁹⁷ See also Nichols at *__ (citing Sedona Principle 12).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Collaboration Platforms may be built in ways which allow full or partial sharing of datum between “documents” or as enabled by access rights. The Federal Rules of Civil Procedure apply to Electronically Stored Information including “data and data compilations,”⁹⁸ which definition Collaboration Platforms fit neatly. There is not necessarily a “document” analogue to the actively dynamic data that comprises what is stored in a Collaboration Platform. A Collaboration Platform may have access restrictions which provide different users different experiences of the same document.

There are many types of metadata - as metadata is generally implemented to support technical processes rather than user generated, the scope, quantity, and quality of fields will vary between tools. Much like a database, data within a Collaboration Platforms may be synced with local copies of documents and Collaboration Platform data may be constantly updating meaning one “document” may be in multiple locations with no “master” document.

Some of the varieties of metadata to consider include:

- **Descriptive metadata**, such as author or creation date, which will be featured in most collaboration tools and are familiar to practitioners who have worked with eDiscovery processes and the metadata associated with a Microsoft Word document.
- **Security and Permissions oriented metadata**, identifying access controls within platforms. Different tiers of software license and different user tiers may have different access to data. In-scope and out-of-scope Custodians in chats and collaborative documents can, in some instances on some platforms, be identified based on permissions. Identifying joiners and leavers from chats and documents may be appropriate and warranted depending on the facts of the case.
- **Administrative or Structural Metadata**, or metadata which is primarily focused on organizing data such as file structure, change history, or organization within a sequence.
- **Rights Metadata** which contains licensing and copyright information.
- **Technical metadata** related to the data objects within the platform.

The relevance of each type of metadata as well as its availability within a platform varies based upon the needs of the case.

Any and all metadata associated with relevant data should be identified, preserved if needed, and raised with the requesting party early in negotiations. The analysis of which metadata to preserve should be done at the earliest possible moment as some metadata may be overwritten over time, after which appropriate steps to preserve would be impossible.

⁹⁸ FRCP 34(a)

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Example: As a Collaboration Platform, Slack allows custom Metadata to be used to trigger events.⁹⁹ This metadata can, and likely is, customized to the connected application.¹⁰⁰ This “trigger event” Metadata is stored and if relevant, can be preserved. An understanding of the ways in which a Collaboration Platform was used as well as the implementation of their Metadata structure is needed to appropriately preserve when litigation is reasonably anticipated.

As Collaboration Platforms have grown and become standardized, a secondary market in eDiscovery tools has grown to support the preservation and collection of data from them. Several of the available tools have been designed to make preservation of most varieties of potentially relevant metadata possible.

Different communication and collaboration platforms may provide access to metadata at the thread level or at the level of messages within the thread or both. Individual emails will always provide fielded internal metadata at the individual email level; in an individual email was collected from a container such as a .pst, then external metadata such the mailbox folder containing the email will also be available. Earlier emails incorporated as a thread in the body of another email are just text without fielded data; but the original individual source emails of those incorporated emails will, if they still exist, contain full fielded internal metadata.

A standard eDiscovery processing program, such as *Reveal*, *Relativity* or *Nuix*, will create separate records for an individual email and each physical attachment¹⁰¹, with extracted fielded metadata for each separate record, and link the parent email with its attachments as a family group. The level of metadata available for other platforms such as social media will vary depending on the collection method.

4. Proportionality and Collaboration and Communication Platforms

Principles of proportionality must be considered in recognizing data, documents, information, communications from a Collaboration Platform as a potential source of ESI.¹⁰² For purposes of being mindful of proportionality in finding relevant information, it is important to understand the distinction between how individual employees, potential custodians, use chat/collaboration tools in comparison to the participants within Channels within these tools.

⁹⁹ <https://api.slack.com/automation/metadata-events>

¹⁰⁰ “Metadata Events are structured data payloads that contain information about events occurring in your Slack-connected application, in the form of a custom event_payload as part of a message's metadata property”.
<https://api.slack.com/reference/metadata>

¹⁰¹ The process for hyperlinked attachments is more complicated, but hyperlinked attachments collected along with their parent emails can also be linked together with their parents as family groups in litigation support platforms such as *Relativity* or *Reveal*.

¹⁰² See The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production, The Sedona Conference Journal, Vol. 19, Nov. 1, 2018, Comment 2 and 2a.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

One rule will not apply as it's likely that collecting all chats by a single custodian, all messages within a single channel or consider all participants in a potentially key channel as custodians will be overbroad, costly and inefficient. Beyond the analysis of custodian vs channel parties in a litigation can examine metadata such as date ranges, specific interactions and document comment history that can ultimately reduce the amount of data collected and reviewed. Working with IT to determine the permissions granted to each user within the chat platform and within each potentially relevant chat channel can help understand both the specific metadata and the data set in general. As important is properly analyzing the relationships between custodians and channels is having complete knowledge of potential handles, nicknames of employees as well as the changing members of each channel over time.

As with any discovery dispute, relevance and proportionality are at the center of discovery involving collaboration platforms, including continuous data streams and communications with linked or embedded documents. If messages are relevant, then the courts must weigh proportionality concerns. When assessing proportionality, the courts must rely on the parties' experts and vendors statements as to the burden imposed to produce ESI. In Benebone LLC v. PetQwerks, Inc., 2021 U.S. Dist. LEXIS 165649 (2021), the U.S. District Court for the Central District of California was faced competing estimates as to the burden of reviewing Slack messages. The plaintiff claimed the cost of producing 30,000 messages would cost \$110,000 to \$255,000. The defendants estimated the cost to be \$22,000. The difference? The hourly rate of the reviewing attorneys. The plaintiff wanted to pay its reviewing attorneys the hourly rates of partner attorneys. The defendant suggested that contract attorneys at the rate of \$40 an hour could do the job. The court sided with the lower estimate and ordered the Slack messages to be produced.

Even when continuous data streams are relevant to the case, logistical—not monetary— concerns may prevent the information from being produced. In Milbeck v. Truecar, Inc., 2019 U.S. Dist. LEXIS 65307 (2019), the court found that Slack messages were relevant and that the defendants could produce them. However, the plaintiff had asked for an expedited trial. The production of 1.67 gigabytes of information would make its way to the plaintiff shortly before trial with no time to digest the information. The court denied the production with the caveat that, if the trial were postponed, the defendants would be required to produce the Slack messages.

5. Evidentiary and Privacy Issues

New communication and collaboration platforms present new evidentiary and privacy considerations that need to be evaluated as part of the Rule 26(f) conference and should be included into the parties' ESI protocols.

While collaboration tools, such as instant messaging, social platforms (e.g., Slack), and team workspaces (e.g., Microsoft Teams) have undoubtedly helped businesses increase the opportunity for communication and teamwork among their employees, these tools pose unique challenges when they become the subject of discovery. Unlike other, more traditional forms of ESI, documents and communications exchanged and stored on collaboration and communication platforms present new issues without analogues.

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

When word processing documents are shared and simultaneously edited by multiple employees in a workspace that overwrites the document with every edit, admissibility concerns may arise. For example, the author of the document may be difficult to identify making authentication and hearsay trickier. If multiple people work on a document, can either person authenticate the document or lay a foundation for hearsay? Where employees are communicating via a chat platform (or within a collaboration tool), there may be issues pertaining to privilege and/or privacy. For example, if a document lives in the cloud and people in multiple locations are accessing it, which jurisdiction's privacy laws apply? If a privilege applies to one of these people and not the others, does the privilege apply wholesale?

Considering the thorny, largely unanswered questions raised by this relatively new technology, it seems prudent to address these items directly early on in litigation during meet and confers, a Rule 26(f) conference, an ESI protocol, and a protective order. *See* Sedona Conference Journal, Vol. 22, 173-174, 179 (parties should think about ESI evidentiary issues early in the case and ensure they have defensible preservation and collection protocols).

a. Admissibility Concerns

(1) *Authenticity*

A traditional method of authenticating documents often involves a witness with personal knowledge testifying that the document is what it is claimed to be. *See* F.R.E. 901(b)(1). With a document stored in a platform that allows multiple people to view and edit the document simultaneously, identifying a single author may prove difficult. *See* Sedona Conference Journal, Vol. 22, at 156 (2021). Thus, it may be beneficial to consider an agreement on this issue early on in the litigation, such as during stipulations regarding ESI and related protocols. *See id.* at 179. *See also*, Sedona Conference Journal, Vol. 19, No.1, at 71-86 (advising that parties confer to reach agreements regarding ESI early in the litigation) Addressing whether an individual who saw or edited a document can authenticate it in such a protocol may help avoid admissibility issues regarding the same in the future.

Otherwise, it will be necessary to rely upon other authentication rules. *See* Sedona Conference Journal, Vol. 22, at 139-140 (discussing collaboration tools and the various ways to authenticate them), 194-209, 218 (identifying methods of authentication for various types of evidence and providing supporting case citations). One other avenue is to rely on a certification by a qualified person via Rule 902(14), which provides authentication of electronic data by digital identification. *See* F.R.E. 902(14). *See also*, Sedona Conference Journal, Vol. 22, at 98-104 (describing the two new subsections to Rule 902 and their application), 156-168 (describing digital identification methods). However, the information necessary to support this certification would need to be gathered before collections occur and revealing this information to opposing counsel early on may not be desirable. *See* Sedona Conference Journal, Vol. 22, at 149-151 (2021) (discussing the challenges of Rule 902(14)).

(2) *Hearsay*

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

As with authentication, testimony is often used to lay a foundation for a hearsay exception—for example, that an out-of-court statement is a statement by a party opponent or that it meets the criteria for a business record. *See generally* F.R.E. 803. Where multiple people are collaborating on a document, however, the author of the statement might be ambiguous making compliance with Rules 801(d)(2) or Rule 803(6) more difficult. Additionally, employees may use chat platforms, like Slack, to discuss personal matters and there might be a question whether specific chat records are properly considered to be business records.

To the extent that a Rule 902 certification is being used, a producing party might consider whether it can include language to lay the foundation for business records. *See* Sedona Conference Journal, Vol. 22, at 167 (2021) (proposing a certification that includes the requirements to satisfy both the authentication and hearsay rules). Where it is difficult to establish that either Rules 801(d)(2) or Rule 803(6) applies, one might consider whether Rule 807 could apply. *See id.* at 152-155 (discussing the changes to Rule 807).

b. Privacy Concerns

(1) Attorney Work Product

Because collaboration tools may make it more difficult to assess the owner of a document or the person who made changes to a dynamic document, it may be difficult to determine whether the work product doctrine applies to shield information from discovery. For example, if an attorney is one of many non-attorney contributors to a particular document, should that document receive work product protection? In such a scenario, it becomes more difficult to assess whether the document was prepared in the ordinary course of business or whether it was prepared in anticipation of litigation. Understanding the existence of this issue on the front end, will allow participants of the document collaboration to document the particular basis for the “anticipation” of litigation and label the document as Attorney Work Product and limit access to the document from its initial inception forward.

(2) Attorney-Client Privilege

Attorney-client privilege generally applies if the communication is for the purpose of legal advice. Accordingly, understanding the context is vitally important. However, communication tools like Slack have a long stream of different conversations over time, which can make separating out discrete conversations (and their purpose) more difficult. To make matters worse, these ongoing chats might stray into less formal conversations and as such, there may be many mixed-purpose communications in one long chain of text. This can make it more difficult to establish that the primary purpose of the communication is for legal advice. Additionally, where a message is sent to a group that includes a lawyer, it can also be difficult to establish the application of the privilege. *See generally Exch. St. Hotel LLC v. Tocci Bldg. Corp.*, 2021 R.I. Super. LEXIS 93, *48 (R.I. Super. Ct. Dec. 20, 2021) (“In a large number of the communications where an attorney was listed as a party to the communication, this

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

Court found that the attorney neither participated in the communication nor was asked for or provided legal advice”).

Additionally, collaboration tools raise questions with respect to the application of the attorney-client privilege with respect to documents—namely when it is waived. For example, if third parties have **access** to a document stored within a collaboration tool, is the attorney-client privilege waived? Does the answer turn on whether the third parties *in fact* accessed the document? Query how difficult it would be to determine the answer to the latter question for each document housed within a collaboration tool.

(3) *Within the borders of the United States:*

i. U.S. patchwork approach

While the United States has a mix of federal laws to protect specific types of data, “[it] has no overarching and preemptive national ‘privacy law’ or ‘data security law’ in place.” *See* Sedona Conference Journal, Vol. 22, at 497 (2021). To the extent there is an all-embracing privacy law, it is state law and there is a lack of consistency across states. *See generally, id.* at 497-505. Communication and collaboration tools function over the internet and the various users may be located in different states. Courts have found that the law governing the claim applies with respect to privileges. *See e.g., Lieberman v. Unum Group*, 2021 U.S. Dist. LEXIS 200941, *5 (C.D. Cal. Oct. 14, 2021) (rejecting arguments that other state’s privacy laws apply, deciding that the law underlying the claims and defenses governs). However, when there are users based in different jurisdictions, considerations must be given to whether each user’s privacy laws will govern that communication.

ii. Stored Communications Act

The Stored Communications Act (18 U.S.C. §§2701, *et seq.*)¹⁰³ prohibits disclosure of the contents of communications from electronic communications services or electronic storage—but the prohibition applies only to those who provide an Electronic Communication Service¹⁰⁴ or a Remote Computing Service.¹⁰⁵ 18 U.S.C. §2702(a). Arguably, this may apply to the content of communications on collaboration tools (including chat messages), but only if the request is to a person engaged in such services.

Because this only applies to those providing specified services, this may not apply where the party does not provide such services and the information is within the party’s control. *See e.g., Flagg v. City of Detroit*, 252 F.R.D. 346, 354-55 (E.D. Mich. 2008) (in considering a motion to quash a subpoena to

¹⁰³ The Stored Communications Act is Title II of the Electronic Communications Privacy Act.

¹⁰⁴ An Electronic Communication Service has been defined as a service that enables one to send or receive wire or electronic communications. 18 U.S.C. § 2510(15).

¹⁰⁵ A Remote Computing Service has been defined as a computer storage or processing services that uses an electronic communications system. 18 U.S.C. § 2711(2).

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

a third-party electronic communication service with whom the Defendant had a contract for text messaging, the court determined that the Defendant had control over the text messages such that they should be produced in accordance with Rule 34). But it may apply to a third-party subpoena seeking information that is outside the control of the party and/or outside the scope of the party's business. *See Shenwick v. Twitter, Inc.*, 2018 U.S. Dist. LEXIS 22676, *6 (N.D. Cal. 2018) (denying plaintiff's request to search the Twitter direct messages of Twitter's custodians because "Twitter did not require its employees to use direct messages for communication").

(4) Electronic Communications Privacy Act

Title I of the Electronic Communications Privacy Act (18 U.S.C §§ 2510, *et seq.*) prohibits the intentional actual or attempted interception, use, disclosure, or "procure[ment] [of] any other person to intercept or endeavor to intercept any wire, oral, or electronic communication." 18 U.S.C § 2511. However, there are exceptions for operators and service providers for uses "in the normal course of his employment while engaged in any activity which is a necessary incident to the rendition of his service" and for "persons authorized by law to intercept wire, oral, or electronic communications or to conduct electronic surveillance, as defined in section 101 of the Foreign Intelligence Surveillance Act of 1978." 18 U.S.C. § 2511.

(5) Health Insurance Portability and Accountability Act
("HIPAA")

Information collected by collaboration platforms may raise additional privacy concerns, especially when information may be considered protected health information. While HIPAA protects the disclosure of personal identifiable information, the HIPAA privacy rule only applies to covered entities—namely health plans, health plan clearing houses, and health care providers who electronically transmit any health information in connection with transactions for which HHS has adopted standards. *See* 45 CFR § 160.102. To the extent a covered entity houses personal identifiable information within a communication or collaboration tool and such information is sought in litigation, then HIPAA may apply. *See e.g., Lillard v. Univ. of Louisville*, 2014 U.S. Dist. LEXIS 201024, *52 (W.D. Ky. Apr. 4, 2014) (denying request for slack because, *inter alia*, it would be a potential HIPAA violation if it contains patient information). In such a scenario, it would be wise for the parties to agree on a protective order that complies with the requirements of 45 C.F.R. § 164.512.

(6) *Outside the United States – General Data Protection Regulation*

The use of collaboration platforms may also raise specific foreign data privacy laws and consideration should be given to the identification of the location of users. In contrast to the United States, Europe has a comprehensive privacy law, General Data Protection Regulation ("GDPR"), which governs how data is processed and controlled within the European Union. The GDPR applies to protect certain information, such as personal identifying information, from any documents, data, or information collected or processed from the European Union. Notably, it can apply to businesses based in the United States if it has an establishment in the European Union or if the business targets individuals in

This document was created for discussion purposes only for the 2023 Annual Meeting of The Sedona Conference Working Group on Electronic Document Retention and Production (WG1). It is not intended for distribution beyond members of the Sedona Working Group Series. Comments are welcome and may be sent by email to comments@sedonaconference.org.

the European Union for offering goods and services or monitoring their activities. *See generally*, Sedona Conference Journal, Vol. 22, at 284-343 (2021). Where a party has employees located in the European Union or does business in the European Union, there may be certain information stored within collaboration tools that is subject to the GDPR.

Appendices

- A. Glossary
- B. Other Sedona Publications Addressing Collaboration and Communication Platforms

DRAFT