



Ediscovery for the New Landscape of Data

Considerations and best practices for modern data types



Table of Contents

3	<u>Introduction to the New World of Data</u>
6	<u>Mobile Data, from Emojis to BYOD</u>
8	<u>Slack, Teams, and Workplace Messaging Apps</u>
9	<u>Zoom</u>
10	<u>Social Media Platforms</u>
11	<u>Clubhouse</u>
12	<u>TikTok</u>
14	<u>Twitter</u>
16	<u>YouTube</u>
18	<u>Ephemeral Messaging</u>
20	<u>The Internet of Things (IoT) and the The Internet of Bodies</u>
22	<u>New Data Requires a New Approach</u>
24	<u>Investigating with New Data Sources</u>
25	<u>What Practitioners Need for the Future</u>
26	<u>Sample Custodial Questionnaire</u>



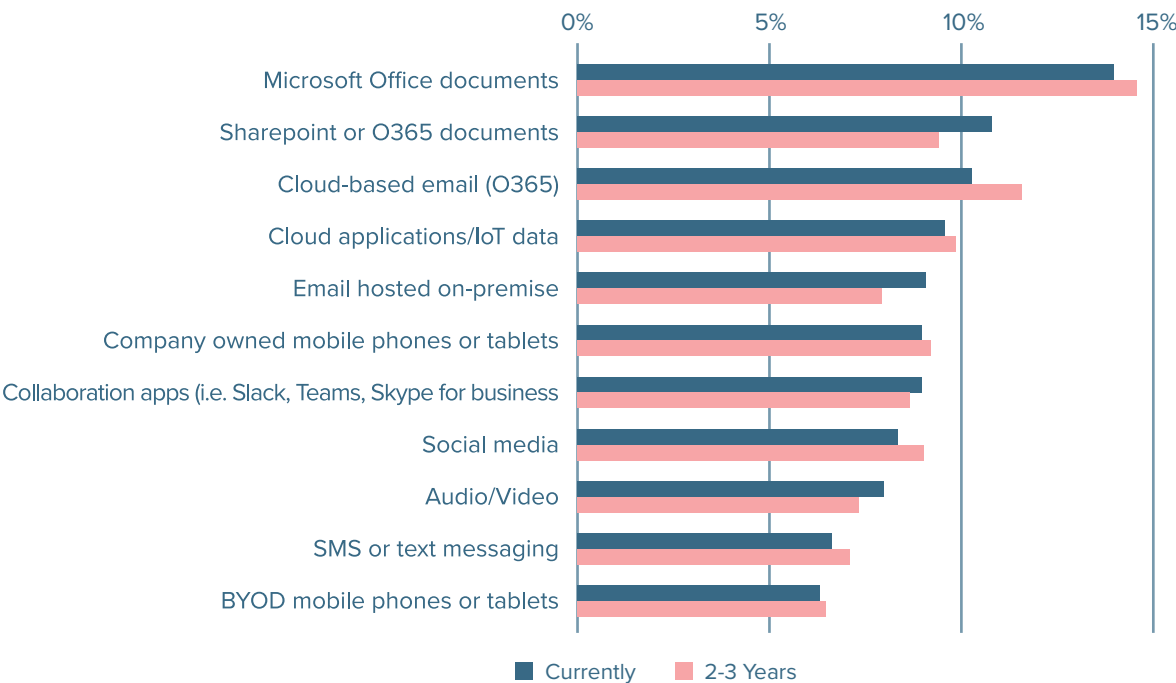
Introduction to the New World of Data

Professionally and personally, we all leave a vast computer-generated footprint in our wake. You might start your day looking at emails on your cell phone, write a tweet, communicate with colleagues internationally on WhatsApp, respond to a request on Slack, and send a quick text to your boss all before crawling out of bed in the morning! This digital trail offers a wealth of information and potential evidence for the savvy legal professional who knows where — and how — to start digging.

The scale and speed of this digital transformation is staggering. Over half of the global population has access to the internet, with [nearly 3.5 billion](#) participating in social media — and new platforms crop up seemingly every year. Additionally, there are new tools and platforms that bring the quick and informal communication style of social tools to businesses. Despite the fact they were meant to reduce unnecessary communication, these tools greatly increase the potentially relevant data for a case team to investigate when conducting discovery efforts.

Atypical data types may not be relevant in every case, but you should level-set early to ensure you are not missing out on highly relevant material. Practitioners must understand where to look, how to extract insight, and the legal or ethical considerations with each new source of evidence.

Of all the data collected during ediscovery/investigations in your organization, what percentage do you currently use and how do you anticipate that percentage will change in 2-3 years?



Source: [IDC's 2020 Legal Tech Buyer Survey](#)

Why does this matter?

People are moving away from email as their default communication tool and are instead relying on a combination of multiple platforms for business and personal communication. To fully understand what was said, to whom and how it affects your case, effectively extracting data across all of these tools is imperative.

Furthermore, communication on these platforms is often more lax — and sometimes more relevant. People do not always realize that what is texted, Slacked, or shared in some other ephemeral manner is still discoverable and dispositive, and a savvy case team can benefit greatly from the resulting candor.

New data poses new challenges

Before diving into the specifics of each data source, there are three important considerations to keep in mind:

1. What constitutes a document has changed
2. Thorough data mapping is more important than ever
3. Not all review platforms are created equal

The New Document

Discovery and production requests were designed with paper in mind. How does a Slack channel with 20 people talking about a specific topic over a two-year period translate to the concept of a document? A Zoom meeting with video, chat, and polls can't possibly be conformed to 8.5"x11" production format. Breaking up a Slack channel discussion just to fit on a page or even determining if portions of a discussion are privileged becomes very challenging when conceiving of the communication like a document.



Tip: While you can always redact non-responsive parts of the conversation, think about whether your case will benefit from additional context (are your key witnesses particularly charming or do they tend toward crass humor?).

Finding the Data

Looking at emails, network shares, and documents stored on a computer is not likely to capture all relevant information nowadays. Savvy legal professionals must dig deeper and inquire about all of the potential ways that custodians are communicating and working together to find all of the potentially relevant information.



Tip: We have included a sample custodial questionnaire in the appendix.

Key information can reside concurrently in multiple locations because modern conversations frequently traverse from email to text to collaboration tools and back again. A single topic thread may traverse internal and external hardware and cloud storage and reside in different platforms and data formats depending on what tools are leveraged. A single subject may have relevant information in a variety of locations and be subject to differing possession, custody, and control.

IT may not even have a full picture of this digital landscape. The use of systems outside of the approval and purview of IT within an organization (aka shadow IT) is a major challenge businesses are seeing today. Employees often begin using newer unapproved tools without engaging with IT (up to 50% of the time according to [one report](#)). [Many employees](#) also use personal mobile devices for work, and as a result, the IT data map or general understanding of the tools employees use in an organization is likely to contain gaps, sometimes substantial ones.

Adjust scope and scale accordingly and to use all the tools at your disposal to triangulate and connect the dots across multiple platforms. Getting the big picture of how people are actually communicating is critical and requires a deeper investigative approach than simply requesting a data map from IT.

Reviewing the Data

Legal practitioners face new methods of communication and generation of ESI complete with new data formats, storage locations, and limitations. From .JSON to .PST, .XML to .MSG, legal practitioners need to be versed in the many file types these new communication methods rely upon and understand which partners can best support extracting and parsing the data type they are conducting on.

Ensuring that you have a tool that can optimize these data types for review is key to accelerating review and reducing time to insight. For example, some technology renders Slack channels in a manner that arbitrarily separates threads into separate files that are hard to efficiently review. Ensure that when partnering with a provider you have an ability to review examples of key data types likely to be in-scope.



Tip: Ask your review provider for examples of how different data types will render. We have included examples from DISCO Ediscovery in this ebook.



Mobile Data, from Emojis to BYOD

Although not a new data source, the relationship between people and their phones is constantly evolving — as are discovery protocols. Key information related to business transactions, product development, and nefarious behavior is often found buried in mobile data, and the work-from-home revolution has only amplified this. What does this mean for practitioners?

The screenshot displays the Ediscovery interface for a document titled "0000002.html". The main content area shows a chat conversation with three messages:

- Message 1:** From +4805553399 Wash, dated April 23, 2020 13:32:29. Text: "birthday? How old ar you lucas? By the sounds of your plans, pretty old!". Includes a photo of a birthday cake with candles.
- Message 2:** From Lucas, dated April 23, 2020 13:33:21. Text: "turned 38, and yea birthdays are not as fun as you get older. No bouncy castle or cotton candy maker for me this year".
- Message 3:** From +7805558976 Joel, dated April 23, 2020 13:37:24. Text: "lets celebrate this weekend, we can treat you to some cotton candy flavored ice cream if you want".

The left sidebar contains sections for TAGS, PREDICTIONS, RECENT DECISIONS, FIELDS, and ACTIONS. The right-hand pane shows JUMPER (No hits), RELATED DOCUMENTS (No family members, Not part of a conversation, No similar documents), and METADATA (Doc ID: 37815, Reference Id, Pages: 2, Custodian: Unspecified Custodian, Bates No., Tags, Filenames: 0000002.html, Created/Modified: 1/23/2020 7:17 PM CST, Path: /NATIVES/0001/0000002.html, Custom_DocID: 0000002, Line Number: 5, Chat Number: 2, Identifier: 16930ef4146cd066c22fe37c23ccbad, Participants: Lucas (owner); +7805558976 Joel +3 more..., Number of at...: 8, Source: WeChat, Instant Mess...: 3, Body: Happy birthday lucas).

Considerations and Best Practices for Mobile Data

Getting data

Because there are multiple layers of authentication required to access all potentially relevant information on a mobile device, physical possession alone is not sufficient to access everything. For iPhones in particular, you will want to ensure that you not only have the device passcode and necessary information for two-factor authentication, but also login credentials for iCloud to access material stored off the device.

What you can forensically and defensibly image from a device can depend on the model and software version of the device. It is extremely important to ensure you or your forensic partner are using the right technology to access the data on the specific model and version of device(s) you are faced with.



Making it pretty

While texting is the dominant communication style on cell phones, the data directly exported from collection tools like Celebrite is unfortunately challenging to understand and quickly review. It is important to work with a technology that can render the text messages in an easy-to-understand manner to accelerate time to insight. Not all tech is created equally on this front, so be sure to ask for an example of text data in any review tool you are evaluating for a matter that will be SMS-heavy.

Emoji overload

After becoming an integral part of texting in the 2010s, the small Unicode images have become a [hot topic in courtrooms](#), which has given rise to some challenges.

Different phones depict emojis differently, interpretation is highly subjective, and not all review platforms are equipped to support rendering emojis in anything like the native format. Does the money bag imply bribery or good luck? Is the eggplant a message about dinner or sexual harassment? As with text messaging generally, ask for an emoji exemplar if you anticipate a text message — or collaboration tool-heavy review.

Knowing where to look

There are a variety of data types and locations that may be relevant in mobile data discovery. Data may reside on the physical device, in a cloud-based backup, and in third-party applications. In the event you are dealing with international matters and/or data in a chat app like WhatsApp or WeChat, global data privacy concerns may be implicated depending on where the third-party application hosts the relevant data.

Battling bring your own device (BYOD)

The proliferation of BYOD policies that allow employees to use personal devices as their business smartphone further complicates the mobile data conundrum. From increasing the variety of potential device types and models to commingling of personal and business-related data, these policies add nuance to data preservation and collection. Additionally, BYOD policies may limit an enterprise's control over what applications an employee adds to the device and what remote collection and wiping capability the device has. Dig into what sort of device policy your client has early in the preservation discussion to uncover these issues.



Slack, Teams, and Workplace Messaging Apps

Inboxes are becoming bloated, unwieldy, and more importantly unmanageable. In response, [email is on the decline](#), and short-form communication and real-time collaboration tools like Slack are rapidly replacing email. These collaboration tools offer a wealth of information potentially relevant to a litigation or an investigation but can be a nightmare to manage. It is important for practitioners to find a partner that understands these unique challenges as they incorporate Slack data into their workflow.

What it is: Short-form, real-time collaboration app for businesses

How many users:

- Slack boasts [12 million active users](#) per day across 156,000 companies
- Microsoft Teams has [145 million](#) daily active users across 500,000 companies

Challenges: Unreadable format, rich but complicated data

Collaboration tools contain a wealth of relevant information in a litigation or investigation, possessing a richness in content not normally seen in email. The informal, rapid-fire, short-format nature of Slack communication makes it a treasure trove for identifying potentially highly relevant data, and to understand what actually happened for an investigation or build a narrative, litigators must understand what relevant information tools like Slack may contain.

Considerations and Best Practices for Slack

The major problem with collaboration apps is formatting. Data exported directly from Slack and other tools is nearly undecipherable in its raw format, a file format called JSON, and vendors are not created equal in supporting these new data types.

Extracting all the relevant information from a multi-person stream with links, reactions, graphics, and shared files and then presenting a cohesive picture of the data is complicated. While many providers claim expertise in parsing these files, they are often using a brute force approach that turns one conversation into hundreds or thousands of files that make building continuity difficult. These results are often as unusable as raw .JSON files.

Look for technology that is able to recreate the visual structure of Slack in an intuitive and readily reviewable format. In DISCO, workspaces and full channels, complete with links, usernames, and all active content are recreated in a viewer, providing a holistic picture of the Slack ecosystem for an organization. The reviewer no longer has to toggle through multiple files to recreate context, and can scroll through the communication stream in much the same fashion as the original user. This ease of use greatly reduces time to insight to uncover relevant custodians and topics of interest.

To further drive cost and time savings, forensic experts can help refine scope, determining which Slack data is presented and ultimately pushed on for review.



Zoom

When you consider that some professionals are now spending six to eight hours of their work day in some form of video conference, the potential scope of the ESI generated is daunting. Although Zoom is perhaps the most common conferencing app, some organizations may use Teams, Skype, or other platforms.

What it is: Video conferencing app

How many users: Zoom has over [300 million](#) daily active meeting attendees

Challenges: Data spans communication channels, scoping is overwhelming

Considerations Best Practices for Zoom

Video conferencing tools add the complexity of audio and video as well as real-time chat functionality to the mix. Ensure that whatever technology you use has the ability to effectively parse the chat portions of Zoom or Skype as well as the video and audio portions. To surface evidence effectively, practitioners need to refine the scope of their discovery to include these data sources and to adopt methods to quickly identify key information and reduce the data to a manageable size.

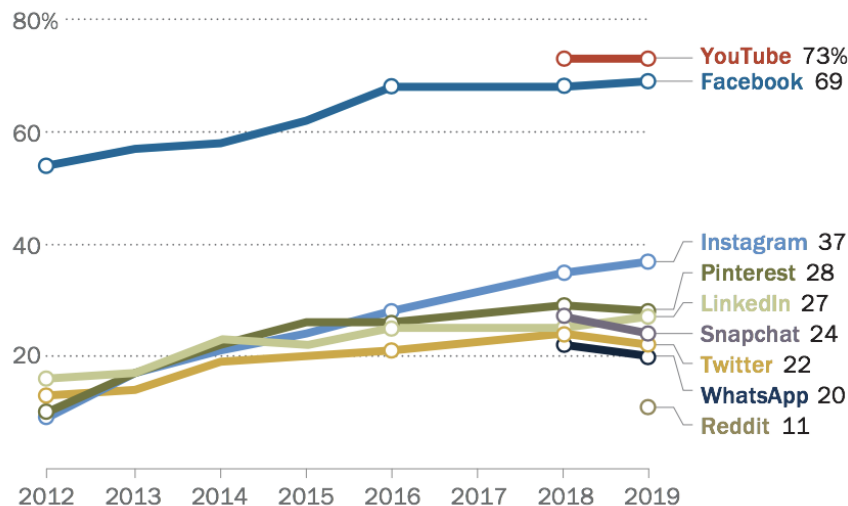


Social Media Platforms

Today, social media use extends far beyond sharing cat memes. A [2019 Pew Research Center survey](#) found that the most commonly used social media platforms by adults in the U.S. are YouTube (used by 73% of adults) and Facebook (69%). Platforms like Reddit and Twitter are essential vehicles for breaking news; Instagram and Facebook are advertising behemoths; and LinkedIn has become crucial for networking. The ways people engage with social media platforms and their messaging functions creates a wealth of potentially relevant ESI that practitioners increasingly rely on.

Facebook, YouTube continue to be the most widely used online platforms among U.S. adults

% of U.S. adults who say they ever use the following online platforms or messaging apps online or on their cellphone



Note: Pre-2018 telephone poll data is not available for YouTube, Snapchat and WhatsApp.

Comparable trend data is not available for Reddit.

Source: Survey conducted Jan. 8-Feb. 7, 2019.

PEW RESEARCH CENTER

Evidence goes viral

The vast web of digital data generated through the use of social media platforms falls squarely within the scope of ESI that is potentially discoverable under the Federal Rules of Civil Procedure. The official Advisory Committee notes accompanying the amended [37\(e\)](#) in the federal rules even go so far as to call out this data source explicitly: “It is important that counsel become familiar with their clients’ information systems and digital data — including social media — to address these issues.” Still, there is confusion as to which aspects of social media data are discoverable and what the most defensible process is for each platform.



Clubhouse

Inboxes are becoming bloated, unwieldy, and more importantly unmanageable. In response, [email is on the decline](#), and short-form communication and real-time collaboration tools like Slack are rapidly replacing email. These collaboration tools offer a wealth of information potentially relevant to a litigation or an investigation but can be a nightmare to manage. It is important for practitioners to find a partner that understands these unique challenges as they incorporate Slack data into their workflow.

What it is: A platform hosting a variety of voice chat “club rooms”

How many users: Clubhouse has [2 million](#) active weekly users

Challenges: No official recordings, can lead to false claims

Reminiscent of old-school internet chat rooms where a group of various strangers could join an open room and talk about whatever subject tickles their fancy, Clubhouse is a series of voice chat “club rooms” that anyone can join.

The app gained notoriety because it was initially invite-only and populated by celebrities and tech moguls like [Elon Musk and Mark Zuckerberg](#). Think of it as a mix between a live podcast, a Reddit board, and talkback radio where all the chats disappear once the conversation is over.

Clubhouse in Court

Clubhouse has not made an appearance in court yet, but it certainly has the potential. For example, it’s ripe for defamation claims — in one instance, New York Times reporter put a tech mogul on blast for using [a slur in a club room](#). In this case, the claim was inaccurate and immediately rebuffed by other members of the room, but had the other members not stepped forward, the accused could face substantial personal or economic backlash. High-profile and not-so-high profile people alike have little resource if a false claim is levied and they do not have other members or audio recording to refute a claim.

Considerations and Best Practices for Clubhouse

At first glance, Clubhouse seems like a pretty low-risk platform, especially, given that it is not archiving or recording the audio interactions. But, just because the platform does not facilitate retaining audio files does not mean that the participants cannot do so themselves, or worse yet make a claim about what was shared without proof to back it up.

There is ample precedent for audio evidence being in scope and even courts sanctioning for [failure to preserve audio evidence](#). The application does not currently record the rooms in the normal course of doing business, which alleviates some of the burden under [37\(e\)](#). Yet, there is nothing to prevent any attendee or participant from recording. In the event a recording does exist on any medium, it is potentially discoverable.

Anybody can listen and record

People in the chatrooms often feel as comfortable as if they are talking with a close group of friends, but the attendees often span the globe and can number in the hundreds or thousands in a single room. Any person can use their laptop or mobile phone to capture audio or share what is said in the room. Depending on the topic this could easily become digital evidence in a case.

If you have an audio recording from a platform like Slack, Webex, or Clubhouse, it is imperative that you have a platform that can support search and transcription to facilitate reducing time to evidence. Next-gen tools like DISCO [support the file formats audio files are most often stored in](#) and do not charge extra for audio and video transcription, while some legacy tools may require working with a third-party application that charges a premium for this service.



What it is: Social media platform with short videos

How many users: TikTok has over [50 million](#) daily active users in the U.S.

Challenges: Legal woes, extracting all metadata, holistic viewing within platform

TikTok is one of the newer social media applications, arriving in the U.S. in 2017 and gaining popularity in recent years. Its users are largely [under 30](#), and videos range from [dance challenges](#) to [insider tips on home inspections](#).

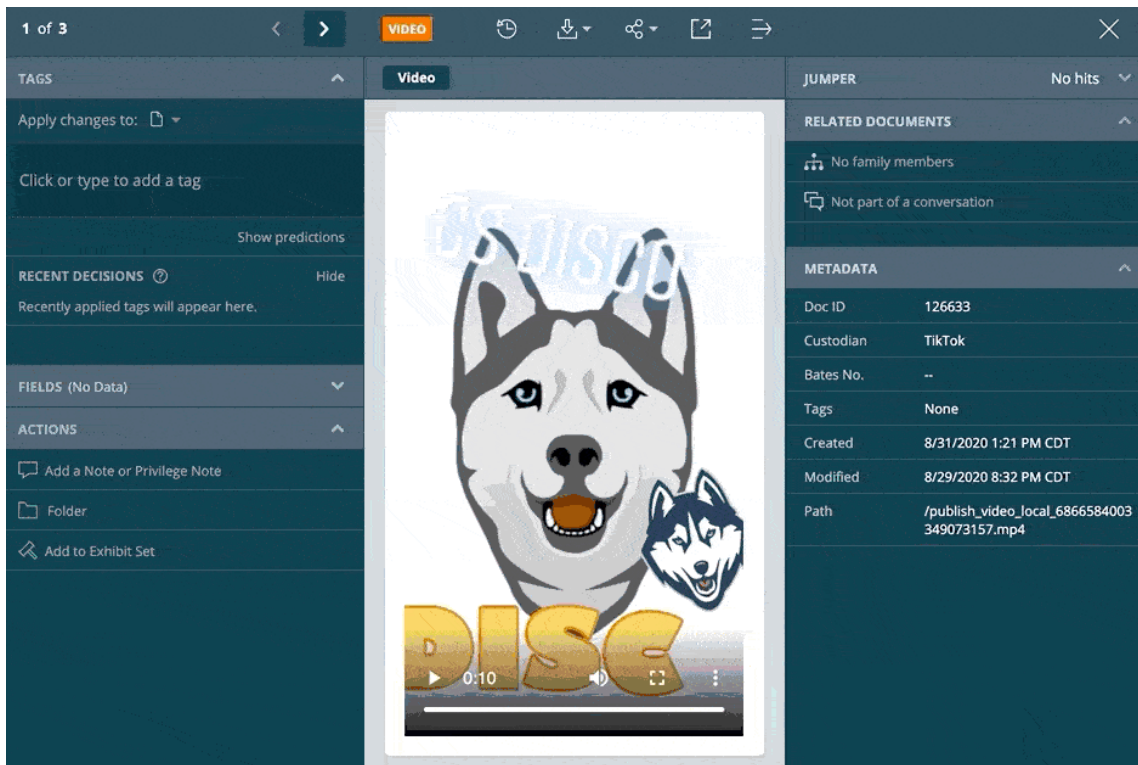
TikTok in Court

To date, most litigation and scrutiny around TikTok has been about [the platform itself](#) (it notably [drew the ire](#) of former President Trump, although litigation was [dropped](#) by President Biden). However, user-generated content will almost certainly start to show up in court.

For legal practitioners, causes of actions involving TikTok could range from social media marketing liability to large-scale copyright infringement concerns as the platform becomes increasingly commercial. As users and corporate entities alike begin to monetize the platform and convert it into an advertising engine, concerns about individual likeness, song sampling, and unfair advertising practices could all birth regulatory scrutiny or large-scale litigation. Perhaps a plaintiff in a class action claiming grievous injury will post a TikTok video dancing their heart out, or a TikTok will be repurposed for advertising without the consent of the user who made it.

Considerations and Best Practices for TikTok

TikTok, like many next-gen social media platforms and communication applications, is not a doc in any traditional sense. So the workflow and methods used to gain insights must be refined to reflect this structure difference. A few of the ways TikTok differs from a document in the traditional sense include: video and audio content, dynamic user interaction data including likes and comments, and metadata about the user and the posting. Key metadata fields including file name, path, date created, date modified, and user upload can facilitate search.



The screenshot displays the Ediscovery interface for a video document. The central pane shows a TikTok video featuring a husky's face with the text "DISC" overlaid. The video player includes a progress bar and a timestamp of 0:10. The left sidebar contains navigation options: TAGS, RECENT DECISIONS, FIELDS (No Data), and ACTIONS. The right sidebar shows document details: JUMPER (No hits), RELATED DOCUMENTS, and METADATA. The metadata table lists the following information:

METADATA	
Doc ID	126633
Custodian	TikTok
Bates No.	--
Tags	None
Created	8/31/2020 1:21 PM CDT
Modified	8/29/2020 8:32 PM CDT
Path	/publish_video_local_6866584003349073157.mp4



Twitter

Over the last several years, Twitter has regained prominence as a major avenue of social, political, and inter-personal discussion. Perhaps more than any other social media application, the potential for legally actionable content is only growing on this platform.

What it is: Social media platform limited to messages of 240 characters

How many users: Twitter has [186 million](#) daily active users

Challenges: Short lifecycle, specific requests required

Twitter in Court

Twitter has a lengthy history in courts, from a [stalking case](#) to [libel and slander cases](#) to [inciting the London riots in 2011](#).

In one [high-profile case](#) related to Wikileaks and the 2016 presidential election, Twitter sought to subvert the Rule 45 subpoena based upon First Amendment rights to anonymous speech. In this case, the court ruled against Twitter because of the narrowness of the request, which excluded personal communication and demonstrated material relevance of the user's identity, and the fact that only Twitter itself could directly provide the information.

Considerations and Best Practices for Twitter

Private info requires a subpoena or court order

While some material is publicly available, much will require either the cooperation of the account holder or more challengingly Twitter itself. Information not readily accessible to the public includes the following:

- Password
- Email address
- Cell phone or address book (helps Twitter suggest users you know)
- Location information (where you're tweeting from)
- System log data (mobile carrier, device and application IDs, IP address, browser, the referring domain, pages visited, and search terms)
- Specific tweets set as private



Per the [company's FAQ on legal requests](#):

"Obtaining non-public information, such as an email address used to sign-up for an account or IP login information, requires valid legal process like a subpoena, court order, or other local legal process, depending on the country that issues the request.

Requests for the contents of communications (e.g., Tweets, Direct Messages, media) require a valid search warrant or equivalent to be properly served on the correct Twitter corporate entity. Law enforcement or government agents must demonstrate a higher burden of proof before a judge will authorize such a request.

Twitter may seek to narrow requests that are overly broad, request additional context if the nature of the investigation is not clear, or push back on the request for other reasons."

Time is of the essence

The most recent [3,200 tweets](#) are visible in a feed and Twitter's advanced search function can drill down even deeper based on timing, user, and subject matter. If a user deletes an incriminating tweet, the window of time to recover it is [merely 30 days](#).

Be specific

Requests for data from Twitter must be sufficiently narrow and specific for the social media behemoth to comply. Twitter is not afraid to fight back if they feel one or both of these factors are not met.

In general the best practice with regard to Twitter requests should be to ensure your request is limited to material that is clearly relevant to the case, time-bound, and not readily accessible from any other data source. It is also important to include the following data points in any request:

- Username
- URL of the Twitter profile
- Date range(s) of the requested information
- Details about the specific information being requested and relevance to the case
- Valid email address for Twitter to acknowledge receipt of the legal request



Every minute, over [500 hours](#) of user- and enterprise-generated content is uploaded to YouTube, which generates nearly \$7 billion in ad revenue a year. YouTube videos and real-time livestreams span everything from the mundane to the catastrophic, and the site's traffic [has skyrocketed](#) during the COVID-19 pandemic.

What it is: Video hosting platform

How many users: YouTube is the second most visited website in the world, behind parent company Google

Challenges: Obtaining metadata, establishing authenticity

Unlike other streaming platforms, YouTube relies on users to create content and then searches for policy violations after the fact. This loose approach to regulation of content has certainly exposed the organization to criticism and misuse of the platform. It also means that people by the millions are uploading potentially relevant video content each and every day. Former barriers to including video evidence in a matter (cost and complexity of managing the video data in review) have been greatly reduced, and this wealth of potentially relevant data is increasingly prominent as a result.

YouTube in Court

YouTube has faced myriad critiques ranging from [copyright infringement](#) and [peddling conspiracy theories](#) to darker things like [violence](#) and sexual exploitation of [adults and minors](#). Legions of content moderators are bombarded by questionable material every day and strive to pull down violators, and some [former moderators have sued for emotional distress](#).

Considerations and Best Practices for YouTube

User-generated content on platforms like YouTube, Vimeo, and others can be instrumental in drawing attention to human rights abuses, used as evidence of a crime, or can itself pose concerns about copyright infringement or defamation. Including this evidence source in your ESI scoping is critically important and as with any other alternative media there are some key steps to take to include these videos as an evidence source in your case.

Whether you are looking to use video evidence as character evidence or direct evidence of a crime, the content must meet the threshold of admissibility for relevance and authenticity mentioned earlier. Once the video has been admitted as potential evidence there are some additional considerations in inclusion in your digital evidence analysis:

Gaining access

Even in the event the video is still actively being hosted on the video sharing platform, use appropriate forensic collection technology to ensure that all relevant account metadata is preserved along with the video itself. From an authentication standpoint, information like date and time of upload, account information, and even IP address may all be impactful to a case. In the event the video in question was deleted recently, you may be able to have counsel directly request a copy from YouTube, but time is of the essence as these deleted files are unrecoverable after a period of a few weeks.



Time and cost to review

Historically, it was often time- and cost-prohibitive into video evidence because the cost of converting the media to a reviewable format was high and then the amount of billable time it would take to review tens or thousands of hours of video rendered the effort untenable. Luckily, with today's AI-powered tools like DISCO, every frame of audio or video content is transcribed and converted into a format that can be searched, categorized, and analyzed for words, phrases, and voice identification. AI can make connections across thousands of hours of video that would have previously been impossible.

Deep fakes

Deep fakes are AI-generated content where real people do and say fictional things, and their quality is nearly indistinguishable on the face from real authentic video. Thankfully, digital forensic experts can identify certain things that are a dead giveaway that a video has been tampered with including:

- Lens distortion
- Color filter array (CFA) artifacts
- Noise level and pattern anomalies
- Compression artifacts
- Editing artifacts



Ephemeral Messaging

Once squarely in the domain of Hollywood spies, businesses and individuals now have access to a whole host of applications that disappear or automatically encrypt upon reading.

The screenshot shows a DISCO interface for reviewing ephemeral messaging. The main window displays a chat conversation with a system message and a user message. The right sidebar shows related documents and metadata.

Display names	Local user	Number of messages	First message sent date/time	Last message sent date/time	Case time zone
17372552689@s.whatsapp.net Android	17372552693@s.whatsapp.net CSDATA (owner)	30	January 22, 2020 16:31:17	January 22, 2020 19:53:11	(UTC-6)

System Message System Message January 22, 2020 16:31:17

Messages to this chat and calls are now secured with end-to-end encryption. Tap for more info.

17372552689@s.whatsapp.net Android January 22, 2020 16:31:17

Hello

17372552693@s.whatsapp.net CSDATA January 22, 2020 19:39:01

How are you?

RELATED DOCUMENTS

Family (1 of 4)

0000006.pdf

289ad8c2-82fa-4d96-a2ee-56...

9ef20287-545d-4902-97d2-42...

60ab2430-a0e7-4906-8216-11...

Not part of a conversation

METADATA

Doc ID 25240

Pages 7

Custodian Unspecified Custodian

Bates No. --

Tags None

From System Message System Message

Ephemeral messaging review in DISCO

People from all walks of life — including engineering departments for [billion-dollar Fortune 500 companies](#) and [protesters](#) — are relying on self-destructing messaging applications (also known as ephemeral messaging) to communicate without leaving a digital trail. Larger technology players like [Facebook](#), [Google](#), and [WhatsApp](#) have also incorporated ephemeral settings.

What it is: Messaging platforms that don't leave a digital trace (Snapchat, Dust, Wickr, etc.)

How many users: Snapchat has [280 million](#) daily users

Telegram reports 500 million monthly users

Challenges: Messages don't leave behind data...or do they?

Generally, ephemeral messaging apps are short-form communication on mobile devices that disappears from the recipient's screen after the message has been viewed. This self-destruct function is deployed in several ways, including: programmatic self-destruct function, specific trigger event (e.g. opening or closing a message), or upon the expiration of a pre-defined time frame. Deletion happens concurrently on the receiver's device, the sender's device, and on the system servers. Any lasting records are eradicated.

The applications can further obfuscate data with functions that [preclude screen shots, limit distribution, auto-encrypt, removal of messages from recipient devices, and untraceable messaging](#).



Considerations for Ephemeral Messaging

While there are actually [many legitimate business reasons](#) to deploy ephemeral messaging in a business context (including lower storage costs, better compliance with data privacy regulations, confidential and encrypted communications), the risks remain substantial. Use of the applications complicates meeting discovery and preservation obligations substantially and may open up an organization to large spoliation sanctions or, as was the case in [Uber v. Waymo](#), the court may determine that the use of such applications alone is grounds for an adverse inference.

The context of ephemeral messaging use as well as the timing of implementation both play a role in whether sanctions are levied. In the case of [WeRide Corp. v. Huang](#), a directive by executive leadership to use ephemeral messaging following the litigation filing to mitigate discoverable data led to sanctions.

The DOJ offered further guidance on mitigating risk when using ephemeral messaging as an enterprise in its update to [Evaluation Of Corporate Compliance Programs](#). Specifically, if a company adopts a short retention period or adopts an ephemeral messaging tool, it should generate a thoughtful, [advanced business justification statement](#) that:

1. shows how the company weighed the risks when setting its policy and
2. serves as a record that could later be used to explain why the company took this approach

From the bench to regulators, the message is clear: employ ephemeral messaging in a business context at your own risk. And the onus is on you to demonstrate a business need, appropriate preservation protocols, and spoliation mitigation efforts on penalty of sanctions, adverse inference, and loss of remediation credits.

Best Practices to Manage Ephemeral Data

- Ensure your team understands the risks and limitations of each tool covered in the policy and that employees are trained on this
- Construct a clear business justification for any standard policy of using ephemeral communication and short retention periods and ensure it is employed consistently
- Clearly outline permissible use, authorization, and communication protocols for any ephemeral application
- Clearly outline any legal prohibitions on using ephemeral communication
- Once on reasonable notice of potential litigation, disable the automatic deletion of ephemeral communications and institute a “litigation hold” to preserve relevant documents and evidence
- Audit the ephemeral applications for data security and use across your organization
- Penalize bad actors misusing the technology

Depending on how ephemeral messaging apps are used in an organization and what the preservation protocol is, you may have to move preservation of ephemeral data to a top priority to avoid spoliation. Like mobile and Slack, the data format is challenging to render in a readily reviewable way and this can impact review speeds significantly.



The Internet of Things (IoT)

What it is: The network of physical objects (things) that are embedded with sensors and software to connect and exchange data with other devices or systems via the internet.

How many users: The IoT is an ecosystem of over [30 billion](#) web-enabled devices from smart refrigerators to doorbell cameras

Challenges: Balancing privacy with the need for data

From smart cars or TVs to Fitbits and Amazon Alexa, there is a proliferation of IoT in our daily lives. As these IoT devices have become commonplace, they have created a vast digital fingerprint that savvy legal practitioners are beginning to mine for ESI.

The Internet of Bodies

There is an entire subgenre of the internet of things called the [internet of bodies](#) that is composed of smart devices on and sometimes inside of the human body. Think of devices like smart pace-makers, Fitbits, and insulin pumps controlled from a mobile device. There were nearly [400 million wearable devices shipped in 2020](#).

These smart devices are revolutionizing healthcare and often greatly improving quality of life, but there are legal and ethical issues practitioners should be aware of. This wealth of data poses fundamental concerns about the individual right to privacy and autonomy.

IoT and IoB in Court

Several cases have been decided based upon information harvested from IoT devices and that number is likely to grow significantly in the coming years. Here are a few examples:

The comprehensive biometric tracking of devices like FitBits has been used as evidence to [disprove personal injury claims](#) when the claimant was exercising despite claiming grievous injuries.

Large logistics and shipping organizations have begun employing web-enabled trackers across their fleet of vehicles over the last few years. Data from these trackers has found its way into [personal injury](#) and [property damage cases](#) and the lack of defensible preservation of this data has resulted in claims of spoliation.

Ever-listening devices like Amazon Echo and Dot and Google Home are increasingly turned to as a source of evidence of crime. In the [most well-known case](#), a combination of Amazon Alexa and a smart water meter were hotly contested sources of evidence in a murder case that hinged on the 3 a.m. draining and refilling of a hot tub.

Most cars today collect, record, and transmit potentially relevant ESI. GPS sends and receives information about location and speed.



Considerations and Best Practices for IoT Devices

Remember to include them!

For many legal practitioners, IoT data is not the first thing that is considered when scoping for ediscovery and data requests. Expand your custodial interview and ESI scoping questions to include potentially relevant sources of ESI, and make sure your request is as narrow as possible to avoid being overwhelmed by data. Most corporate legal departments have yet to include IoT data in their data governance or litigation response plan, so it is important to raise the issue and reevaluate your data governance to include it. Additionally, specificity is key in making requests for IoT data, especially because many respondents may be unfamiliar with it.

Data privacy

IoT and IoB devices have raised privacy concerns in terms of the data that is collected and shared by the devices and various litigation and regulations exist to face that challenge. As privacy considerations relate to ediscovery, the [Stored Communications Act](#) (SCA) generally prevents providers of electronic communication services from divulging private communications. That being said, the SCA does not preclude the court from requesting the data from the person in physical possession of a smart device ([Flagg v. City of Detroit](#)).

IoB devices have myriad privacy concerns because they track, record, and store things like users' whereabouts, bodily functions, and what they see, and hear. Can a health insurance company deny coverage based on information from a wearable or embedded device? Determining who can access, collect, or interact with this sort of personal information and personal health information is a key consideration with any IoB device.

Spoliation

IoT devices have an increased susceptibility to data modification and destruction. Something as simple as restarting the device, interrupting its power supply, disabling its internet connection, or perhaps even just moving it out of range of its known internet access point so that it cannot sync its data to the cloud can all impact the data integrity of some smart devices. Additionally many IoT devices have very short windows of time before data is overwritten or automatically erased, so determining a preservation approach quickly should be a top priority.

Gaining access

Large enterprises with IoT products are not always forthcoming with providing the ESI from their smart devices. In a murder case involving Alexa data, Amazon [fought for years](#) before ultimately turning the data over. It is also important to consider the device owner's actual "possession, custody, or control" under Rule 34 — this question is complicated due to the web-enabled nature of the devices.

Data headaches

Data generated by IoT devices does not always play nice with traditional ediscovery platforms. It is important to understand the data format you are engaging with, any potential limitations, and what level of usability the technology you rely on will provide the data for you.



New Data Requires a New Approach

Understanding how your organization communicates and conducts business will enable you to construct a plan to manage the burgeoning data volumes. A linear approach of throwing as many attorneys in a room and going page by page through the data is no longer going to cut it.

Where do you look?

The key to determining which platform to investigate is to understand how relevant custodians are communicating, and on which platforms. Ephemeral messaging applications like Signal, Wickr, and Telegram may be relevant if an engineering team frequently uses them, WhatsApp or WeChat may be relevant in a case involving international organizations, and Twitter, Facebook, or Instagram may be relevant in a case involving unfair marketing practices. Understanding the nature of a case and if and how custodians are leveraging social media help determine the priority of social media discovery.

What do you look for?

Each social media platform contains a potentially voluminous amount of disparate data dating back to the inception of a user's account. It is important to understand what your technology partner will include in their capture of a social media profile and what metadata will or will not be included.

User-generated ESI may include:

- A user's posts, likes, and comments
- Direct messages
- Chat logs
- Friends or connections
- Profile
- Log-on and posting times
- GPS data from photographs (just ask John McAfee)[89]
- Some deleted materials

System-generated ESI may include:

- Proprietary unique identifier
- Item type
- Parent item/thread
- Recipients
- Author/poster
- Linked media
- IP addresses



How do you get it?

While it may be enticing to print out a screen capture of a public social media site or even have the account owner press the “download your information” button several platforms have, it is important to remember that this will not necessarily include all the data you are looking for and may be limited to only public posts. Additionally, some social platforms limit what you are able to export based upon the type of account a user possesses. Working with a forensic collection technology that specializes in social media collection will ensure you are able to gain access to the full scope of potentially relevant information. It is also important to work with a technology that can render the social media data in an easily reviewable manner.

When can you use it?

While there is ample precedent and case law to support the inclusion of social media ESI (even data that is private) in a discovery request, the requesting party still has an obligation to meet the requirements of [FRCP 26\(b\)](#) and demonstrate relevance to the case. And the bar for relevance, [Federal Rule of Evidence 401](#) is far from high. Evidence is relevant if “it has any tendency to make a fact more or less probable than it would be without the evidence” and “the fact is of consequence in determining the action.” To ensure that this does not become a fishing expedition, the court will often limit subject matter and duration of admissible ESI.

An additional area of concern with social ESI is authentication that the account and material posted to it were generated by the custodian or named account owner. As with relevance, the bar is not terribly high. [Federal Rule of Evidence 901](#) states to establish authenticity, “the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it” and this done via presenting the “distinctive characteristics” of an account according to 901(b)(4). These characteristics may include account name, photos of the account owner, nicknames, IP address, specific topics, or slang.

Best practices for scoping today:

- Ask key custodians and members of their team what tools are used to communicate. Prioritize key custodians’ data sources based on frequency and type of communication
- Determine situations when certain tools may be used (work-related texts or Slack messages are often sent after working hours)
- Inquire where various data sources reside (on-premise, in iCloud, on backup servers)
- Determine how employees are using new data types (social communication vs work product)
- Research what kind of licenses the enterprise has for each tool (some licenses are limited in what data is exportable)
- Find out if any of the custodians deal with structured data sources (IoT, databases like Salesforce, or trading platforms)
- Use insights gained from each data source to refine the scope of data reviewed in new sources. For example, key dates and communication maps found in Slack can help refine what emails you review



Investigating with New Data Sources

Prioritize and triangulate

No one has a limitless budget or timetable, so as with all aspects of investigation and litigation, it remains important to evaluate the potential value of a data source before investing time or money in reviewing it. Prioritize data sources based upon the frequency in which they are relied upon by key custodians, and understand that email may not be the first data source investigated. Use insights from each successive investigated data source to triangulate in on key periods of time, concepts, and data ranges to ensure that you are continually reducing the burden of reviewing new data sources and taking a precise approach to mining for relevant information.

Determine how hard it is to get the data

Not all communications tools are equal in terms of effort to extract their data. Once a new data type is identified, it is important to understand the type of license (if any) the enterprise has, because certain levels of licensing provide only a limited data export. A second key consideration for paralegals dealing with atypical data is determining which technology or technology service provider can best handle the data type employed by a given tool. A final consideration, where physically the servers hosting a given app reside, different regions have differing data privacy and jurisdiction considerations.

Rely on partners with specific expertise for the data source

Tools like Slack and ephemeral messaging applications have a different data format than email. Certain methods of extraction, processing tools, and/or document review platforms have limitations that can make a review of atypical data substantially more challenging and costly. When in doubt, demand that a provider show examples of what the data will look like in the review tool before engaging. Ensuring that your technology partner can not only handle the data, but also provide it in an easily reviewable format can save a case team substantial time, effort and of course client money.



What Practitioners Need for the Future

Once you have access to these new data types, making sense of the formats, linguistic nuances, and fluidity of topics is the next hurdle to face.

Need for (AI) Speed

Because key concepts in this post-document universe often traverse a myriad of data types in a given conversation, AI becomes an increasingly important resource. Advanced algorithms can make correlations across disparate data types in a way the human mind is not equipped to. In next-gen tools, AI surfaces cross-data-type concepts or identifies “similar” pieces of ESI regardless of data type. If a custodian discussed a key topic in text, then moved to Slack, sent an email about it, and then typed a short Slack message, AI can locate and correlate the similar concepts despite the differing formats.

Social network analysis

Leverage social network analysis to prioritize custodians based upon the frequency of communication with other key custodians across data sources. As you have more and more sources of data to review, leveraging tools to effectively prioritize and triangulate limited resources is extremely important.

Find your data guide

The universe of data is evolving so rapidly that it is challenging to keep pace, but luckily there are companies and individuals whose entire job is to do just that. Ensure that you, your client, or the service provider has a well-versed expert to help guide the process, to avoid pitfalls specific to the data type, and to ask the right questions as you navigate through this process. Understanding the risks and benefits of atypical data at the outset can avoid costly mistakes and decisions that may render the data unreviewable.

What got you here won't carry you into the brave new digital world

Relying on the tools, methods, and people that got us across the finish line in the era of email is not enough to navigate a rapidly evolving world where self-destructing text messages, emojis, GIFs, and Slack messages may contain the smoking gun. Practitioners should have a basic level of technological fluency, understand enough to spot issues with technical repercussions, and have a network of resources to address deeper technical questions. In this ever-changing digital ecosystem, no one has to go it alone.



Sample Custodial Questionnaire

Custodian Details

1. Custodian name, email, office location, IM/Slack username
 - a. Note any former names, email addresses or nicknames
2. Custodian job questions
 - a. Current title & role/responsibilities?
 - i. What other roles have you held at this company?
 - ii. Confirm start date and role changes with HR in advance of discussion if possible
 - b. Time in current role/role at relevant time for the matter?
 - i. Information on predecessor
 - ii. Did your predecessor provide you with any hard copy or digital files upon leaving? If so, where are those files currently stored?
 - c. Who do you report to?
 - d. What business unit are you a part of?
 - e. Do you have direct reports, and if so how many?
 - f. What was your title and role as it related during the matter?
 - i. Did your role change at any point during the relevant time period of this matter?

Case Specific Questions

1. During what periods of time did you work on/engage with the subject matter of this case?
2. What was your involvement in the subject matter of this case?
3. Who were you engaging with on the subject matter of this case?
4. Did you have any conversations with in-house counsel (or outside counsel previously)?
5. Are you currently under legal hold? Since when? Do you understand your obligations under legal hold?
6. Are there any specific channels or email distribution lists that were used to discuss the subject matter of this case?



Physical Devices

To the extent possible, work with IT to get a list of known work related assets in advance of custodian interviews. Recommend reviewing the company's device, privacy policies, litigation hold notices prior to any custodian interviews.

1. Do you have a company-issued desktop computer, and what is the machine name? (Mac or PC, model type)
2. Do you have a company-issued laptop computer and what is the machine name? (Mac or PC, model type)
3. Do you conduct business using a personal desktop or laptop and what is the machine name? (Mac or PC, model type) Where is that machine currently located?
4. Do you have a company-issued mobile device? (If so, what type? iPhone, Blackberry, Android, etc).
5. Does your company have a BYOD device policy and if so, are you conducting business on a personal device? (If so, what type? iPhone, Blackberry, Android, etc).
6. Do you have any other company-issued or personal devices you conduct business on? If so what type (iPad, tablet, etc.)? And where are they located?
7. Do you have any company-issued external storage devices? (External hard drive, thumb drive, jump drive, etc.) If so, is it encrypted and can you provide the encryption key?
8. Do you have any personal external storage devices? (External hard drive, thumb drive, jump drive, etc.) If so is it encrypted and can you provide the encryption key?
9. Do you have any company data stored on physical media like CDs/DVDs? If so is there any encryption or password protection?

Traditional Data Sources

1. **Paper** - Please identify all locations you keep paper documents.
(Filing cabinet in office or home, with secretary/admin, central filing, offsite)
2. **Electronic Documents** - Please identify any electronically stored document types other than email you possess
(Documents, Spreadsheets, PowerPoints, PDFs, Internal Wikis, Intranet etc)
 - a. Do you store these documents locally on your personal laptop, desktop or mobile device? If so please identify the full file path location for these documents.
 - b. Do you store any of these documents on an external hard drive, thumb drive, or other digital device? (If so, provide device and folder naming convention)
 - c. Do you have any of these documents stored on physical media like CD or DVDs?
 - d. Do you store electronic documents in a document management system (iManage, Interwoven, etc.) or in a shared drive?
 - e. Who in your organization is responsible for these document management systems(DMS)?



- 1. Email** - Do you have email that may be relevant to this matter?
 - a. What email system does your organization employ (O365, Gmail, Lotus Notes, etc.)
 - b. Where is the email you believe to be relevant stored? (folder name or inbox)
 - c. Do you review your email locally in Outlook, Thunderbird, Apple Mail? If so, do you have a local Personal Storage Table (.pst) file archive or similar email archive on your hard drive?
 - d. Do you have any other email aliases that you have used to send or receive company email or ESI? (Address and provider)
 - i. Do you believe any relevant information is in your personal account(s)? If so, which and where in the accounts?
 - e. How is email stored in your organization? Is there a backup, retention policy, or deletion protocol I should be aware of?
 - f. Do you archive or backup your email to any other location?
- 2. Calendar** - Do you have a digital or physical calendar where you track appointments and/or take notes?
 - a. Do you use a digital calendar - if so which kind?
 - b. Do you have a digital calendar or assistant outside of the one provided by your organization?
 - c. Do you have a physical calendar or planner you use to track appointments and/or take notes?
 - d. Does your assistant manage a digital or hard copy calendar on your behalf (if so where)?
- 3. Shared Resources** - Do you use network shares and data management tools and/or SharePoint?
 - a. Does your organization have shared network drives or a document management system in place, and if so which?
 - b. Do you have a personal folder on a shared drive, and if so what is the name?
 - c. Is material on the shared drive accessible or saved locally as well?
 - d. What are the folder names or file paths where you keep the information? Provide screenshots if possible.
 - e. Who in your organization is responsible for these DMS?
 - f. Are files backed up on-premise or in the cloud?
 - g. Are there any other relevant company databases to include?
 - i. How are they accessed? Who administers the database?
- 4. Data Repository** - Do you personally or does your company employ cloud-based data repositories?
 - a. Which cloud-based data repositories do you use? (Google Drive, Box, Dropbox, OneDrive, etc.)
 - b. Is there any material in the repository that is not present locally on your device(s) or in network shares?



- 1. Passwords and Encryption** - Do any relevant pieces of ESI or data sources (physical and application or database) have passwords, encryption or login requirements?
 - a. Please provide all relevant passwords (iCloud, GSuite, physical device logins, password-protected documents)?
 - b. Please identify any devices or media that are encrypted and provide the encryption key(s)

Atypical Data Sources

1. Mobile

- a. Device type, passwords, and whether personal and professional (or BYOD) devices are all used for work communication or tasks
- b. How do you use your personal and professional cell phones? (Text, email, voice, voicemail, apps, photos?)
- c. Do you have any relevant electronically stored information on your mobile device?
- d. Are there top contacts we should look at within text, messaging apps, or call logs/voicemail?
- e. When did you last backup your device? Do you employ iCloud? Have you recently replaced or updated your device?

2. Collaboration tools - Does your organization have an enterprise license for any collaboration tools? If not, are you still employing any collaboration tools outside of company approval?

- a. Which collaboration tools are you using? (Slack, Teams, CRM etc.)
- b. How do you/does your group use the collaboration tool(s)?
- c. Did you discuss substantive business on the collaboration tool? Which channels might contain relevant information during which periods of time?
- d. Do other groups use this tool or other ones? How so?

3. Instant messaging - Do you personally or does your team use any instant messaging platforms in the course of doing your job? Does your organization have an enterprise license for any instant messaging tools? If not, are you still employing any collaboration tools outside of company approval?

- a. Which tools does your organization use and/or which other tools do you personally or your team use (Jabber, Yammer, Messenger)?
- b. Do you have your IM tools set to archive or do you have screenshots? If so where?
- c. How did you or your team use the IM function? Was substantive business discussed on the IM tool? Were documents shared on the IM platform?



- 1. Ephemeral messaging** - Does your organization have an enterprise license for any instant messaging tools? If not, are you still employing any collaboration tools outside of company approval?
 - a. Which ephemeral messaging applications do you or your team use (auto-encrypting or self-destructing short form messaging) (Signal, Wickr, Telegram, etc.)?
 - b. How were you or your team using ephemeral messaging? Was substantive business or information relating to this matter shared via an ephemeral messaging app?
 - c. Does your company have a policy about ephemeral messaging and/or a data preservation document that relates to ephemeral messaging?
 - d. Does your app have a login?
- 2. Social Media** - Did you in the context of your job or personally use social media to discuss matters relating to this case or conduct business related to this case?
 - a. Which Social media platforms were you using and which may have material relevant to this matter? (Twitter, LinkedIn, Facebook, Instagram, Snap)
- 3. Short-form communication via application** - Did you, in the context of your job or personally, use messaging apps to discuss matters relating to this case or conduct business related to this case?
 - a. Which messaging applications do you or your team use to discuss business matters or information relevant to this case? (WeChat, WhatsApp, Line, Kakao, Skype, Viber, etc.)
 - b. Does your company have a policy about messaging and/or a data preservation document that relates to messaging apps?
 - c. Does your app have a login?

Closing Questions

- 1.** Is there anyone in your organization that you feel may have relevant information to this matter?
 - a. Current or former colleagues
 - b. Assistant
 - c. People outside of the organization
- 2.** Do you have any additional information that you believe is relevant to this matter or the identification of material relevant to it?
- 3.** Are there any acronyms, terms, or abbreviations that are used to refer to information relevant in this matter?