DIGITAL MOUNTAIN®

WINTER 2022 E-NEWSLETTER

At Digital Mountain, we assist our clients with their electronic discovery, digital forensics, cybersecurity, and data analytics needs. For this E-Newsletter, we discuss the collaboration software market and various eDiscovery nuances such as file links within communications.

Approaches to Resolving Google File Links in Email During eDiscovery

Many organizations and attorneys are grappling with the issue of how to deal with links to external files within email and message communications. These are circumstances whereby files are not attachments, but rather links to storage locations in cloud drives. For this article, we specifically focus on Google mail ("Gmail") and embedded links in the body of Gmail messages, as well as different approaches that can be taken in discovery.

Google Workspace is a collection of cloud computing, productivity and collaboration tools, software, and products developed and marketed by Google. First launched in 2006 as Google Apps, rebranded as G Suite in 2016, and renamed Google Workspace in late 2020, Google Workspace consists of Gmail, Contacts, Calendar, Meet, and Chat for communication; Currents for employee engagement; Drive for storage; and the Google Docs Editors suite for content creation, including Docs, Sheets, Slides, and Forms. As part of the Google Docs Editors suite, a link may be referenced in an email, but in eDiscovery these links historically have not been reconstituted. Electronic discovery evolves over time, and collecting particular information may be burdensome and inefficient until tools develop to improve automation and assist in minimizing costs. As eDiscovery practitioners, we use the latest and greatest commercially available tools. This article addresses approaches using built-in tools by Google versus third-party application providers for data collection of document links along with strengths and drawbacks.

Google Built-in Data Collection Tools

Google Takeout provides the ability to export data for an individual account. Presently, forty-six data types and records can be preserved. Because Takeout is part of the Data & Privacy options of the individual user's Google account settings, it cannot be done without the user's knowledge and cooperation to log-in with the username and password. This method will collect all emails and all files that are stored in Google Drive. The emails will include any files that are actual attachments, but will not include linked files from Google Drive. The collection provides the name of the file, but the file may have changed since the email was actually sent or received. Under the user’s profile, it is possible to grab all versions of a file with dates and times, but you have to manually correlate what is the closest file to the email data provided (definitely, not ideal). Also, the output for mail is in the mbox format that may need to be converted to .pst or .msg to be further processed by most eDiscovery tools.

Takeout will not automatically collect files that are stored on Google Drive that do not belong to the subject user, i.e., it will not grab data from someone else's storage. This is true even if the document was shared with full edit permissions. Also, when there are links to drives outside of Google, these documents may be difficult to collect, and if permission is still available, would have to be downloaded manually. The file may not reflect the document as it appeared in the ordinary course of business due to subsequent edits that may have occurred from the time of the email.

Google Vault, if enabled for the Google Workspace edition, provides enterprise search functionality for a single user account or an entire organization domain. Most smaller organizations typically do not pay this extra cost. The results are much like Google Takeout except that by utilizing a super admin account with the right permissions set to collect data from all users in the domain, discovery occurs without alerting the end user. Beyond mbox format, you can also export as a .pst file which does not require further conversion. Just like Google Takeout, the emails will contain files that are the actual attachments, but not the underlying file stored in Google Drive that correlate to links. For collecting within an email, one will only get the filename of the document. This is not enough to provide the emails and the shared files as a single record and must manually be associated as outlined above in the Google Takeout section of this article.

Third-Party Data Collection Tools

There are many third-party tools and approaches for preserving Gmail, but as of this writing there is only one that we have found that can batch collect files that are linked in an email.

Metaspike Forensic Email Collector (FEC) can collect email across a single user’s account with his or her cooperation, or if using an enterprise-level Google account, FEC makes it possible to use a single set of administrator-level credentials to access end-user mailboxes. Based on the configuration for preservation and the account type, Google may send an email to the target mailbox notifying the user that FEC was granted access. FEC relies on Google’s API and will collect emails and files that are linked from Google Drive. This coupling is provided as a zip file with each email and corresponding files from the link. This also includes files that are shared from other users’ Google Drives. FEC has more robust search parameters and has an option to provide every revision history of a Google document provided by the custodian being preserved. For emails that have a linked file from another user’s Google Drive, only the most current version of the file will be provided, and you will not be able to see the revisions unless you are able to collect the emails and files from the other user’s account. What are some drawbacks?

- Metaspike requires an additional purchase of a software license that is not included as part of the built-in Google Workspace.
- The link being collected may not be the version of the file viewed in the email by the custodian.
- If choosing to collect all versions, it is possible to get flooded with files that were not relevant to the email being reviewed. Also, you may run into Drive attachments whose revisions you cannot acquire due to lack of authorization or marked as Confidential.
- The preservation can take longer if grabbing all the linked files.

- If an email contains a string where the file or a folder is referenced numerous times, it will be preserved every time there is a link which can be voluminous.
- If a link occurred to a drive folder, an enormous amount of data may be collected inadvertently.
- Depending on the review tool and how it displays documents in .zip files, the relationship between the email and files from the link may not be apparent.
- If there is an exception due to permissions, confidential label, or other reasons, the exception is documented as a number. The number cannot be traced to the filename stored in Google drive, so these can be difficult to resolve.

Many corporations have asked about collecting links in Google mail more effectively, and FEC is certainly a step in the right direction for automation. However, case teams need to approach discovery with an understanding of the strengths and drawbacks as there are no best practices yet. If only a handful of file links may be relevant, a more manual approach may be the best solution. If an enormous amount of file links may be relevant, a more automated approach may be appropriate with an understanding of the caveats. The file links issue should be part of effective case planning when dealing with Google mail, and it may even be considered an issue for the "meet and confer" requirement imposed by the Federal Rules of Civil Procedure (or, more specifically, FRCP 26(f)).

Please direct questions and inquiries about electronic discovery, digital forensics, cybersecurity, and data analytics to info@digitalmountain.com.

UPCOMING INDUSTRY EVENTS

THE SEDONA CONFERENCE WORKING GROUP 1 TOWN HALL ON SEARCH AND RETRIEVAL METHODS 2022
Virtual, January 20, 2022

NETDILIGENCE CYBER RISK SUMMIT CONFERENCE 2022
Fort Lauderdale, FL: January 31, 2022 - February 1, 2022

ABA TECHSHOW 2022
Chicago, IL: March 2-5, 2022

11TH ANNUAL ASU-ARKFELD EDISCOVERY, LAW AND TECHNOLOGY CONFERENCE
Phoenix, AZ: March 8-9, 2022

LEGALWEEK 2022
New York City, NY: March 8-11, 2022

Click here to see more upcoming events and links.

Digital Mountain, Inc. Founder and CEO, Julie Lewis, will be presenting at various upcoming industry events. Please send requests for speaker or panel participation for her to marketing@digitalmountain.com.

DIGITAL MOUNTAIN, INC.
4633 Old Ironsides Drive, Suite 401
Santa Clara, CA 95054
866.DIG.DOCS

www.digitalmountain.com

Contact us today!

FOLLOW US AT: