From Proposition to Acquisition: Complying with Complex Second Requests

A Global Financial Platform Company Handles Slack Data and Expedited Timelines with HaystackID

By Michael Sarlo, EnCE, CBE, CCLO, RCA, CCPA

HAYSTACK®

Supporting a Complex DOJ Second Request

When proposing to acquire a highly regulated company, a global finance platform company and its internationally recognized outside counsel needed support responding to a Department of Justice (DOJ) Second Request. Facing limited resources and a looming deadline, the company turned to HaystackID for help adhering to this complex investigation. HaystackID got to work; the team collected and evaluated **18TB of data**, including significant Slack message and file stores, from onsite and remote locations spanning more than **17 types of data stores**. HaystackID completed nearly **300 collections** in **106 days** and built custom tools and processes, including:

Innovative Slack-specific Communications Heat Maps
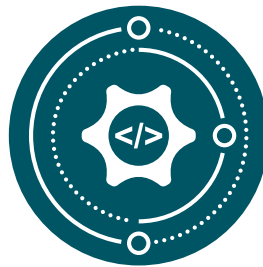
Predictive Coding Processes
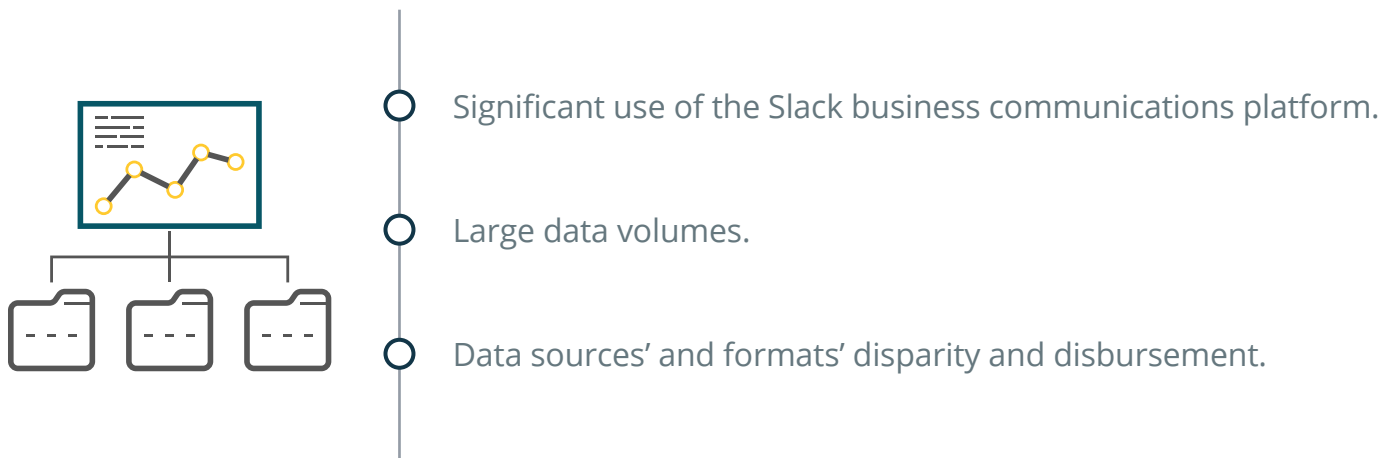
Private Message Privilege Identification

With HaystackID's tailored solutions, the global finance platform company **successfully complied with the complex investigation request**, paving the way for the completion of its proposed acquisition.

The Pitfalls of Fixed Solutions: Overcoming Second Request Challenges with a Tailored eDiscovery Approach

Antitrust investigations, with their unique and specific requests, demand customized and innovative solutions. However, many eDiscovery providers often struggle to meet these demands, resorting to rigid and inflexible solutions. HaystackID, on the other hand, takes a different approach. Our custom software and process engineering capabilities empower our team to develop precision solutions for specific data and legal discovery challenges. This approach has allowed us to maintain a successful track record in meeting antitrust mandates, leveraging the right **technologies**, **techniques**, and **talent**.

This approach and our expertise in handling large and diverse data sets proved successful in mid-2020. This COVID-constrained summer made adhering to a costly second request much more challenging. To optimize resources while meeting the request's demand, one of the world's leading financial platform companies and its outside counsel, an internationally prominent law firm with deep antitrust and competition expertise, turned to HaystackID to support an acquisition-triggered DOJ Second Request.

The company knew our history in handling these requests and needed support overcoming challenges, including:

Significant use of the Slack business communications platform.

Large data volumes.

Data sources' and formats' disparity and disbursement.

Roadblocks, Recommendations, and Results

With extensive Second Request experience manifested by participation in fourteen antitrust agency investigations during 2019 and 2020, HaystackID has a deep academic and experiential understanding of FTC and DOJ Second Requests as mandated by the Hart-Scott-Rodino (HSR) Act of 1976. The HSR Act requires parties to mergers or acquisitions of specific sizes to notify the FTC or the Antitrust Division of the DOJ and provide information and documentation regarding the proposed transaction. Upon reviewing submitted information and documentation, the FTC or DOJ may make additional requests, known as Second Requests, before rendering a decision on the proposed transaction.[1]

Second Requests are discovery procedures that consist of formal requests for additional information and documentation and generally follow the framework of the Model Request for Additional Documentary Material (Second Request) as published by the FTC Premerger Notification Office.[2] While leveraging many of the technologies, techniques, and tactics used in traditional eDiscovery activities supporting audits, investigations, and litigation, Second Request discovery is different because it typically has unique characteristics that need to be considered in each case. Three key attributes of Second Requests, especially applicable to HaystackID's summer 2020 support of a global financial platform company's Second Request included:

- Disparate data and locations
- Need for advanced technologies
- A standard of substantial compliance

Notable among these three attributes is the standard of substantial compliance. Substantial compliance[3] is compliance with the significant or essential requirements of a Second Request that satisfies the request's purpose or objective even though the formal requirements may not be wholly complied with at the time of the response. This standard is unique among discovery requests in that it is time-driven and represents a qualitative best effort at compliance instead of a quantitative, time-independent approach to compliance. Based on the standard of substantial compliance, eDiscovery providers to the challenged parties must balance time, effectiveness, and efficiency to meet certification requirements for Second Requests. To meet and exceed the substantial compliance standard supporting the acquisition triggered Second Request, HaystackID was required to face and overcome three salient roadblocks.

The Roadblocks

The three roadblocks to meeting the accelerated timelines of the DOJ Second Request to the global financial platform company while meeting the standard of substantial compliance included volumes of data, data sources, and Slack discovery challenges.

Substantial Volumes of Data

As a global financial platform company seeking to acquire a multinational finance company, HaystackID's client in this Second Request effort had office locations, data repositories, and investigation-relevant individuals throughout the world, with a preponderance of data located in the United States. Given the regulatory requirements driving the disposition of data for financial companies, HaystackID faced the challenge of organizing and supporting a Second Request for high volumes of data.

To begin to determine and address this high volume of data, HaystackID worked to develop a four-phase approach to addressing eDiscovery needs in the short timelines mandated by the Second Request requirements. This four-phase approach included:

- Phase One: DOJ Notification and Planning
- Phase Two: Collection and Processing
- Phase Three: Review and Production
- Phase Four: Termination/Expiration and Completion

Structuring phases to support the expected multi-terabyte effort approaching 20TB of data and operating with communications, collections, and coordination constrained by COVID travel and social distancing restrictions, HaystackID, working with experts from both the company facing the Second Request and its outside counsel, developed detailed collection plans, identified challenges requiring custom software and process development, and designed workflows to support the entire continuum of eDiscovery tasks necessary to produce best-effort information to the substantial compliance standard.

Additionally, as part of planning to support high volumes of data, HaystackID formally established its case discovery team to assist continued administrative and operational planning and execution. This dedicated team of experts included:

- Forensics First Team: Forensics and Collections
- Early Case Insight Team: Processing and Analytics
- ReviewRight® Team: Review and Production
- Project Management Team: Case Management

With volume expectations considered and case support organization established around four phases to be delivered by four dedicated teams, HaystackID transitioned its focus to planning to address the second significant roadblock, diverse data sources.

Diverse Data Sources

The challenge of increasing types of data continues to be rated as one of the biggest business concerns for eDiscovery specialists, with almost one in five data and legal professionals rating it as the challenge that will most impact their business in the next six months.[4] This challenge posed a formidable hurdle to the global financial platform company's Second Request as it was magnified by the disparate locations of repositories and endpoints containing data to be considered in the antitrust investigation. Data sets from multiple locations containing multiple document formats to be considered in this case included but were not limited to data types from platforms that included:

- BlueJeans Video Conferencing
- Box
- Confluence
- Custom Support Apps
- Druva
- Google Workspace
- Evernote
- LumApps
- Microsoft 365
- Mobile Device Apps
- OneDrive
- Proprietary Finance Apps
- Quickbase
- Slack
- Webex
- Websites
- Zoom

The challenge of diverse data sources was magnified by the new remote world requirements triggered by COVID-related workplace constraints in the summer of 2020. These constraints necessitated collection planning that covered both remote and onsite custodian interviews and data collections from geographically disparate repositories, endpoints, and mobile devices. With an understanding of data volume expectations and the diversity of data sources, the HaystackID team then concentrated on the third important roadblock to this complex Second Request. That being the requirement to collect, process, analyze, and review Slack communications.

Collect   Process   Analyze   Review
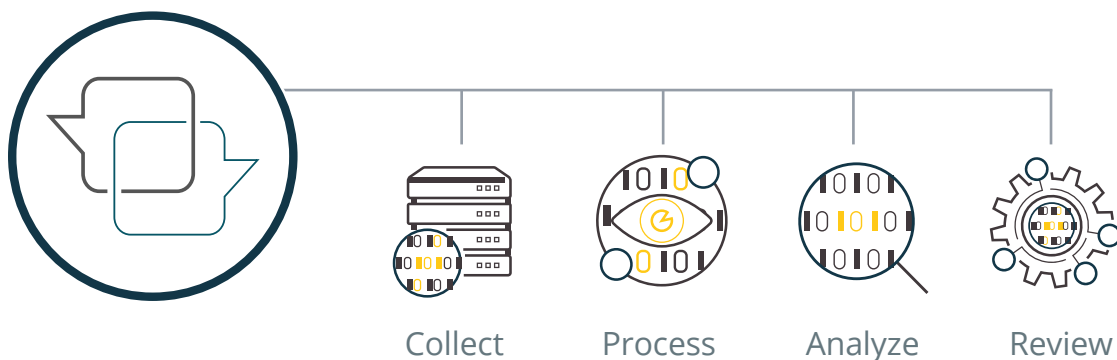
Slack Innovation and Integration

Developed as an internal communications tool for gaming company Tiny Speck, Slack launched in 2013 and has grown into one of the world's most ubiquitous business communications platforms[5] With a name that is an acronym for Searchable Log of All Conversation and Knowledge,[6] the Slack platform was not designed initially to support defensible eDiscovery-centric collections for investigations and litigation. However, over time, that capability has been integrated into recent enterprise-level implementations. Currently, Slack implementations fall into three major categories: Free (or Basic), Standard, and Plus (or Enterprise).[7]

Free Slack implementations are based on the total number of messages in a channel and have short retention. Collection for Free Slack implementations is performed by obtaining an individual API token from users, which allows for the collection of objects available for the custodian. This requirement to gain access at the individual custodian level via OAuth, an authentication protocol that enables individuals to approve one application interacting with another without a transfer of passwords, has been a time-intensive and automation-unfriendly barrier to enterprise collection of Slack communications. Thus, making the collection from Free Slack implementations generally untenable for the accelerated deadlines for Second Requests.

Both Standard and Plus (or Enterprise) Slack implementations have unlimited retention by default. Previously, Standard and Plus (or Enterprise) Slack implementations have relied on a Corporate Export capability. However, Slack's Corporate Export capability is discouraged as it continues to be dependent on individual OAuth tokens to collect private messages. Recently, Slack introduced the Slack Enterprise Grid. The Enterprise Grid is a network of two or more Slack workspace instances. Slack workspaces on the Enterprise Grid have access to the Slack Discovery API. This Discovery API lets organizations use approved third-party applications to export and act on Slack messages and files.[8] This API-enabled discovery capability enables providers such as HaystackID to accelerate the precision and speed of Slack collections, making it a tool of choice to facilitate Slack data acquisition in support of Second Requests. Nevertheless, even with Enterprise Grid's discovery support, there are still challenges requiring innovative approaches to evaluating, presenting, and preparing Slack data for processing, review, and production. One of these challenges is the use of predictive coding on Slack data.

As defined in *The Grossman-Cormack Glossary of Technology-Assisted Review,*[9] Predictive Coding is an industry-specific term generally used to describe a technology-assisted review process involving the use of a machine-learning algorithm to distinguish relevant from non-relevant documents. It is based and dependent on a subject matter expert's coding of a training set of documents to achieve maximum effectiveness. This definition of predictive coding provides a baseline description that identifies one particular function that a general set of commonly accepted machine learning algorithms may use in a technology-assisted review (TAR).[10]

Regarding Slack, predictive coding can pose unique challenges, especially for messages and files consisting mainly of numeric data, spreadsheets, image files, and short text messages. This challenge is because more text typically leads to greater accuracy in predictive coding and pulling enough text into the predictive coding process to enhance accuracy requires identifying within timeframe segments enough text to allow predictive coding analytics engines to categorize data more accurately.

HaystackID understood and was well-positioned to address Slack challenges from collection to predictive coding, given its extensive experience in developing custom tools and processes for integration into its advanced and operationalized collection, processing, and review plans.

Specific custom tools and processes requiring development and deployment by HaystackID included:

○ Slack communications heat maps to allow for the quick identification of message volume and pulse rates over time.

○ Slack-specific analytics process to address cluster concepts from names of participants.

○ Slack-specific analytics process to address not enough text in messages.

○ Slack-specific analytics process to address large control sets.

○ Bifurcating of Slack public and private channels to support requested privilege reporting agreements with the FTC.

Understanding the roadblocks posed by substantial volumes of data, diverse data sources and formats, and Slack collection and predictive coding, the HaystackID team quickly evaluated Second Request requirements and guidelines from the Department of Justice, the company, and counsel. This evaluation, based on an experiential understanding of Second Requests[11] and deep software engineering expertise in developing customized tools and processes to solve non-standard discovery challenges, enabled HaystackID to define, scope, and initiate project planning through the lens of four recommendations to the global financial platform company its outside law firm counsel.

The Recommendations

Concise in concept yet complex in components, Second Request planning led by HaystackID's data and legal discovery experts led to establishing four key recommendations to serve as guideposts for translating investigation and discovery planning into execution. These four recommendations included:

Establish a comprehensive collection plan for physical and remote acquisition of custodial and non-custodial data with iterative data pull, manipulations, and pushes to operations team.

Develop a defensible and sustainable precision workflow designed to support department of justice, company, law firm, and HaystackID requirements across entire eDiscovery continuum for high volumes of data.

Develop precision tools and techniques for detecting, identifying, collecting, and reviewing slack data to meet stringent second request timelines with a comprehensive best effort.

Develop and reach agreement from the department of justice for the use of technology-assisted review, including specific approach for slack review and privilege considerations.

Upon agreement with recommendations and approval of Slack review and privilege approaches from the DOJ, HaystackID began translating planning efforts into the execution of required tasks to deliver a comprehensive and best effort production in the accelerated time frame synonymous with Second Request investigations.

With volume expectations considered and case support organization established around four phases to be delivered by four dedicated teams, HaystackID transitioned its focus to planning to address the second significant roadblock, diverse data sources.

The Results

Based on organic and integrated eDiscovery expertise coupled with its extensive experience in supporting antitrust investigations, HaystackID was able to accomplish necessary collection, processing, review, and production tasks to support a compliant response to the DOJ. Detailed metrics and milestones of the complex case along with specific innovation and integration efforts are detailed in the following paragraphs to add context to this high-volume case with diverse data and significant Slack requirements.[12]

Collection Metrics and Milestones

From a collections perspective, HaystackID collected approximately 18TB data from approximately 300 collections over 84 days, including initial and refresh timing requirements, from both onsite and remote locations. Collection efforts included 54 custodian email acquisitions, 15 custodian interviews, and the acquisition of data from more than 17 data repositories. Additionally, HaystackID collected numerous endpoint data sets ranging from Microsoft 365, Druva, and Box to Slack, Webcasts, and Websites. Key collection highlights include:

- Total Collection Time Frame: 84 Days
- Web Storage (Box.com, Google Drive, Druva) Collection Requests: 160
- Web Storage (Box.com, Google Drive, Druva) Data Size: 14,228.4 GB
- Custodian Email Collection Requests: 54
- Custodian Email Data Size: 2,173.54 GB
- Slack Collection Requests: 44
- Slack Data Size: 2,118.51 GB
- Smart Device Collection Requests: 6
- Smart Device Data Size: 371.36 GB
- Collaboration Spaces (Quickbase, Confluence, JIRA): 16
- Collaboration Spaces (Quickbase, Confluence, JIRA) Data Size: 15.06 GB
- Additional Data: Approximately 10TB

Processing Metrics and Milestones

From a processing perspective, HaystackID processed approximately 46M documents from Slack and non-Slack collections, resulting in just under seven million post-deduplication and post data-filtering documents. This processing phase conducted over 106 days resulted in about 6.7M technology-assisted review (TAR) eligible documents and more than 1.5M TAR excluded documents requiring linear review. Key processing highlights include:

• Total Documents Processed: 45,805,850

• Slack Messages and Files Processed: 256,361

• Microsoft 365 Documents Processed: 30,150,563

• Box Documents Processed: 8,813,252

• Druva Documents Processed: 4,581,613

• Proprietary Finance Application Documents Processed: 322,260

• Google Workspace Documents Processed: 163,663

• Mobile Documents Processed: 621,819

• Additional Documents Processed: 896,319

Additionally, from the corpus of Slack messages and files, HaystackID processed data from 487 public channels and 8,104 private channels. This bifurcation of public and private channel processing supported agreements with the DOJ for privilege reporting.

Following detailed processing instructions and specifications, HaystackID applied advanced de-duplication and filtering technologies and techniques to determine the documents that would best be further evaluated via TAR (TAR Eligible Documents) or via linear review (TAR Excluded Documents). The processing and evaluation step in the eDiscovery continuum resulted in the following TAR eligible and liner review requirement file totals:

• Total Documents to be Reviewed: 8,607,391

• TAR Eligible Documents: 6,733,743

• Linear Responsiveness Review Required Documents: 521,023

This processing reduction of approximately 80% from the total complex data set of collected documents set the stage for HaystackID application of TAR expertise and technology and proprietary ReviewRight document review services to further evaluate documents for the antitrust investigation.

TAR Review Metrics and Milestones

Leveraging a TAR 1.0 workflow for non-Slack files and messages, HaystackID conducted 8 control rounds over 7 days, supported by 12 training rounds over 8 days with document sets ranging from 200 to 1,000 documents. HaystackID was able to exit the training phase for non-Slack files and messages with approximately 98% consistency, about 16% depth of recall, and approximately 75% recall, with precision percentages approaching 50% and F-Scores above 60%. This combination of control rounds and training rounds resulted in an achieved recall goal of 75%, a confidence level of 95%, a maximum margin of error rate slightly above 4%, and an estimated document set richness nearing 12%.

Additionally, given the nuances and technical expertise required for leveraging TAR 1.0 with Slack messages and files, HaystackID conducted 5 control rounds over 5 days, supported by 12 training rounds over 9 days with document sets ranging from 200 to 800 documents. HaystackID was able to exit the training phase for Slack files and messages with approximately 95% consistency, about 31% depth of recall, and approximately 75% recall, with precision percentages approaching 29% and F-Scores greater than 40%. This combination of control rounds and training rounds resulted in an achieved recall goal of 75%, a confidence level of 95%, a maximum margin of error rate under 5%, and an estimated document set richness nearing 12%.

This precision and powerful approach tested thoroughly and applied with exactness led to the TAR 1.0 evaluation and reduction of both non-Slack and Slack documents from 8,607,391 TAR eligible documents to a combined total of post-TAR non-Slack and Slack documents of 618,491 to be reviewed for privilege and key merger related documents.

8,607,391
Documents

TAR 1.0
Evaluation

618,491
Documents
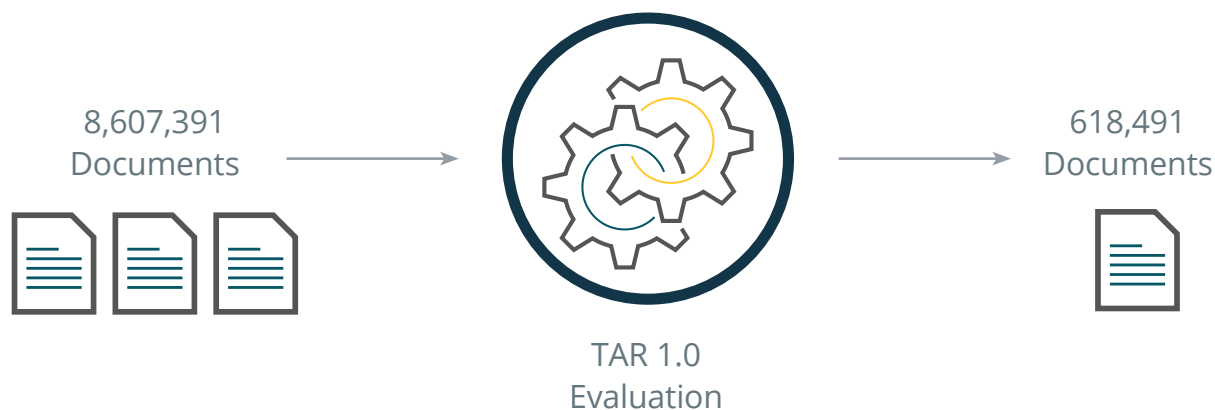
Linear Review Metrics and Milestones

In conducting a time-sensitive, content-specific document review in support of an antitrust agency-driven Second Request, HaystackID leveraged the attributes of its extensive reviewer assessment, qualification, and certification process and selected from its database of almost 20,000 potential attorney document review candidates the most qualified, most appropriate, and most immediately available reviewers to support the review of more than 1.1M documents. HaystackID's industry leading ReviewRight Match® enabled this selection process by applying a combination of proprietary technologies, innovative evaluation tools, and proven protocols that allowed for the rapid and comprehensive sourcing, testing, and qualification of reviewers.[13] This process enabled HaystackID to present and prepare 295 reviewers who were legal review experts and possessed the domain expertise congruent with the global financial platform company centric Second Request.

Given its unique position as the industry remote review leader with more than six years of experience in delivering and managing virtual review projects, HaystackID then leveraged its ReviewRight Virtual® secure remote review infrastructure to support the time-sensitive Second Request-driven review.[14] Having supported more than 1,000 successful secure remote review projects in both pre-COVID and current pandemic environments, HaystackID's combination of secure infrastructure managed by a team of not only review management experts but by virtual management review experts enabled the rapid assembly of a highly experienced virtual review team of more than 300 reviewers, review managers, and technicians.

Executed with a series of rolling reviews initiated over 96 days, the linear review portion of the Second Request-driven review resulted in the complete review of 1,139,514 documents, with more than 20,000 redactions and a privilege log of just under 80,000 documents.

Production Metrics and Milestones

Upon completion of the linear review and certification of the review results, redactions, and privilege logs, this comprehensive eDiscovery project conducted from start to finish in only 106 days and consisting of tasks ranging from remote collections to virtual review and composed of complex data types and formats, including public and private Slack channel messages and files, culminated in the production of approximately 2,000,000 documents.

Speed, Slack, and Specialization

This Second Request-driven eDiscovery project demonstrated HaystackID's integrated and organic eDiscovery expertise and capability as a specialized eDiscovery firm. Implementing innovations ranging from heat maps for presenting Slack communications patterns and density to custom technology-assisted review protocols for reviewing non-standard data types, including private and public channel Slack message and files, HaystackID enabled the successful execution of the project that allowed the leading global finance platform company and its outside counsel to respond to the DOJ Second Request compliantly. This compliant response ultimately contributed to the successful acquisition by the global finance platform company and demonstrated HaystackID's speed of execution, Slack expertise, and specialization in remote operations and Second Request investigations.
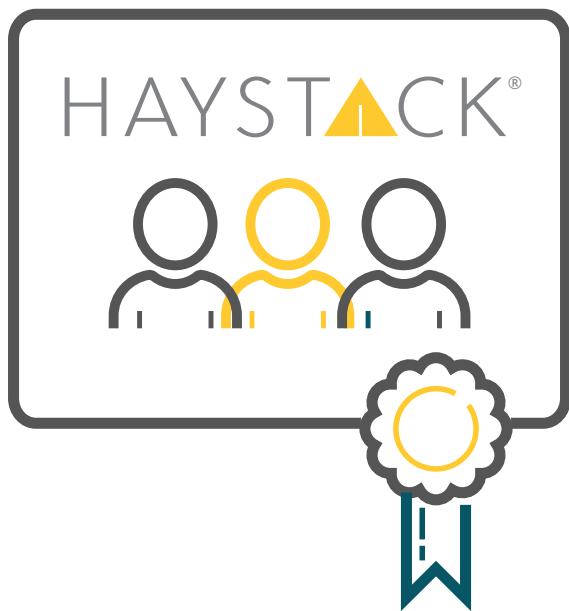
Speed of execution.

Slack expertise.

Specialization in remote operations and Second Request investigations.

References:

[1] Federal Trade Commission and the Antitrust Division of the Department of Justice (2020). Hart-Scott-Rodino Annual Report Fiscal Year 2019. [online] Federal Trade Commission. Available at: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-bureau-competition-department-justice-antitrust-division-hart-scott-rodino p110014hsrannualreportfy2019.pdf [Accessed 29 December 2020].

[2] Federal Trade Commission (2019). Model Request for Additional Information and Documentary Material (Second Request). [online] FTC Premerger Notification Office. Available at: https://www.ftc.gov/system/files/attachments/merger-review/may2019_model_second_request_final.pdf [Accessed 29 December 2020].

[3] "Substantial Compliance." The Merriam-Webster.com Legal Dictionary, Merriam-Webster Inc., https://www.merriam-webster.com/legal/substantial%20compliance. [Accessed 29 December 2020.]

[4] ComplexDiscovery https://complexdiscovery.com/holding-the-rudder-fall-2020-ediscovery-business-confidence-survey-results/

[5] Slack https://en.wikipedia.org/wiki/Slack_Technologies

[6] Slack https://en.wikipedia.org/wiki/Slack_(software)

[7] Slack Categories https://slack.com/pricing

[8] Discovery API https://slack.com/help/articles/360002079527-A-guide-to-Slacks-Discovery-APIs

[9] Grossman, M., and Cormack, G. (2013). The Grossman-Cormack Glossary of Technology-Assisted Review. [ebook] Federal Courts Law Review. Available at: https://www.fclr.org/fclr/articles/html/2010/grossman.pdf [Accessed 31 Aug. 2018].

[10] Predictive Coding https://complexdiscovery.com/casting-a-wider-net-predictive-coding-technologies-and-protocols-survey-fall-2020-results/

[11] HaystackID Second Requests - https://haystackid.com/an-integrated-approach-to-second-requests/

[12] HaystackID Project Notes (Work Product) - January 5, 2021.

[13] ReviewRight Match®. (2020, December 29). HaystackID. https://haystackid.com/review-right/review-staffing/

[14] ReviewRight Virtual®. (2020, December 29). HaystackID. https://haystackid.com/review-right/secure-remote-review-service/

Learn More. Today.

[Contact us today](#) for expert support in navigating the complexities of antitrust agency Second Requests and learn how our integrated approach can streamline your response process and ensure compliance.

About HaystackID®

HaystackID solves complex data challenges related to legal, compliance, regulatory, and cyber events. Core offerings include Global Advisory, Data Discovery Intelligence, HaystackID Core® Platform, and AI-enhanced Global Managed Review powered by its proprietary platform, ReviewRight®. Repeatedly recognized as one of the world's most trusted legal industry providers by prestigious publishers such as Chambers, Gartner, IDC, and Legaltech News, HaystackID implements innovative cyber discovery, enterprise solutions, and legal and compliance offerings to leading companies and legal practices around the world. HaystackID offers highly curated and customized offerings while prioritizing security, privacy, and integrity. For more information about how HaystackID can help solve unique legal enterprise needs, please visit HaystackID.com.

About the Author

Michael Sarlo is the Chief Innovation Officer and President of Global Investigations and Cyber Incident Response at HaystackID. In his role as Chief Innovation Officer, Sarlo spearheads the development of cutting-edge data collection, eDiscovery, and review solutions, focusing on non-traditional data types and creating award-winning applications that address critical review challenges, such as cellular device review and AI-enabled privacy protection workflows. As President of Global Investigations, he expertly handles complex regulatory, investigative, and civil litigation matters, leveraging his expertise as a certified forensic examiner to drive efficiencies in the eDiscovery process. Additionally, Sarlo leads HaystackID's Cyber Discovery and Incident Response Services division, offering pivotal support during data breaches by blending his deep knowledge in IT, data security, eDiscovery, and privacy to help organizations navigate regulatory requirements and resolve security incidents. Michael may be contacted at MSarlo@HaystackID.com.