DISCOVERY PROPORTIONALITY MODEL
A NEW FRAMEWORK
James F. Humphreys Complex Litigation Center
George Washington Law School
July 16, 2021
(DRAFT)

Center Senior Patrons
Western Alliance Bank
Technology Concepts & Design (TCDI)
Preface

The Discovery Proportionality Model: A New Framework establishes a standard approach that identifies and prioritizes custodians by the significance and burden of discoverable matter they possess or control as well as the projected cost, so that the Rule 26(b)(1) proportionality factors can be applied systematically.

The new analytical framework was developed by 56 practitioners, ediscovery experts, and judges to better inform Rule 26 proportionality decisions made at all stages of litigation. The New Framework will strengthen counsels’ strategic decisions in anticipation of litigation and in later Rule 26 proportionality assessments, help further counsels’ meet and confer discussions concerning proportionality issues in discovery, and better inform the judicial resolution of outstanding discovery disputes.

The Washington Law School’s James F. Humphreys Complex Litigation Center and John Rabiej, former Director of Duke Law School’s Center for Judicial Studies, launched the project in early 2020. A two-day bench-bar online conference with more than 250 registrants, including 50 federal judges, reviewed a draft in March 2021. The draft was revised in light of comments at the conference. It will be published for public comment for an eight-week period. A bench-bar conference will be held in 2022 to begin considering developing best practices applying each Rule 26(b)(1) proportionality factor using the New Framework as the foundation.

The drafting team will review all public comments and revise the New Framework, if appropriate. A steering committee of 15 lawyers, judges, and ediscovery experts developed an outline, recommended candidates to draft the New Framework, and serves as an editorial board, under the auspices of the GW Complex Litigation Center. The editorial board will make final edits, subject to the Center’s oversight. The New Framework represents the authors’ consensus views, but not necessarily unanimous agreement with all content.

The New Framework as described was originally derived from Insight Optix LLC’s patented Evidence Optix® software. The Center and the drafters readily acknowledge the important role Insight Optix played in the advancement of the New Framework.

The current version of the New Framework, however, is unlike the software program in key and numerous minor respects. For example, the New Framework relies on three objective criteria, including materiality, uniqueness, and firsthand knowledge, instead of “relevance” as the criterion for prioritizing custodians. It also specifies eight data sources and ranks them according to the degree of burden in accessing information from them. Individual elements of cost are itemized. Most significantly, the New Framework estimates standard-unit costs for the five data sources as baseline reference points. All these variances were developed independently by the 56 experts engaged in this project.

Although the New Framework will be freely available to the public and posted on the Center’s website, this New Framework does not grant any rights to the use of Evidence Optix® software or intellectual property.
ACKNOWLEDGEMENTS

The James F. Humphreys Complex Litigation Center gratefully acknowledges the time and work of the 56 lawyers and experts who drafted the New Framework. Nine of the team members assumed greater drafting responsibilities and served as team leaders. In addition, the input and feedback of six judges has been invaluable and materially improved the final product.

It is with great gratitude that the center recognizes the contributions of the judges, lawyers, and other experts who contributed to this effort.

STEERING COMMITTEE

Arpert, Douglas (Hon.)
District of New Jersey

Basile, Bobbi
HBR Consulting

Baylson, Michael (Hon.)
Eastern District of Pennsylvania

Gensler, Steven
The University of Oklahoma

Hanks, George (Hon.)
Southern District of Texas

Johnston, Iain (Hon.)
Northern District of Illinois

Kaufman, Beth
Schoeman Updike Kaufman & Gerber

Opsitnick, Tim
TCDI

Regard, Daniel
Intelligent Discovery Solutions

Ross, Mandi
Insight Optix LLC

Yates, Christopher (Hon.)
17th Circuit Court, Michigan

TEAM LEADERS

Basile, Bobbi
HBR Consulting

Belt, William
CDS

Kenney, Colleen
Sidley Austin

Lowe, Anthony
DataLaw

Quartararo, Michael
ACEDS

Regard, Dan
Intelligent Discovery Solutions

Ross, Mandi
Insight Optix LLC

Trask, Andrew
Shook, Hardy & Bacon

Tully, Martin
Actuate Law
| TEAM MEMBERS | | | |
|---|---|---|
| Addington, John | Baratz, Michael | Bauer, Kate |
| Corning Law | Steptoe & Johnson | Steptoe & Johnson |
| Behan, Kevin | Bennitt, Jane | Blaustein, Stacey |
| Centene Corporation | Global Legal Ebilling | IBM |
| Carle, Jay | Carvo, Scott | Coleman, Jennifer |
| Seyfarth Shaw, LLP | Warner Norcross & Judd | Hopkins Carley ALC |
| Deere, Chioma | Dockterman, Mike | Durocher, Skip |
| Williams, Leininger & Cosby | Steptoe & Johnson | Dorsey & Whitney |
| Hamilton, Michael | Hulliberger, Kristopher | Isaacs, Tom |
| Exterro | Howard & Howard | Bowman & Brooke |
| Jackson, Matt | Judkins, Sonya | Kairalla, Jason |
| Sidley Austin | T-Mobile | Carlton Fields |
| Krezalek, Martin | Lange, Michelle | McGuire, Marina |
| Blank Rome | VW | Tanenbaum Keale |
| Murao, Joy | Murray, Barb | O’Brien, Kate |
| Practice Aligned | PWC | Insight Optix LLC |
| O’Connor, Tom | Pass, Robert | Patzakis, John |
| Gulf Coast Legal | Carlton Fields | X1 |
| Redgrave, Jonathan | Rizzolo, Dan | Robinson, Rich |
| Redgrave | ePerceopt | Toyota Legal One |
| Schouten, Nataliya | Scrutton, Alexandra Grand | Shortnacy, Michael |
| FIS Global | 3M | King & Spalding |
| Six, Michelle | Stadler, Anna Patricia | Staunton, Thomas |
| Kirkland & Ellis | Robert Bosch GmbH | Miller Shakman Levin & Feldman |
| Sullivan, Janet | Sweeney, Caroline | Yates, Janice |
| White & Case | Dorsey & Whitney | Insight Optix LLC |
DISCOVERY PROPORTIONALITY MODEL
A NEW FRAMEWORK

Introduction

I. Background

The scope of discovery under Federal Rule of Civil Procedure 26(b)(1) was revised in 2015 to require not only that the matter be relevant to any claim and defense, but that it also be proportional to the needs of the case.

Six factors must be considered in assessing whether a matter is proportional to the needs of the case, including: “(1) the importance of the issues at stake in the action; (2) the amount in controversy; (3) the parties’ relative access to relevant information; (4) the parties’ resources; (5) the importance of the discovery in resolving the issues; and (6) whether the burden or expense of the proposed discovery outweighs its likely benefit.” In addition, Rule 26(b)(2)(C)(i) limits discovery if it is unreasonably cumulative or duplicative.

Although unfamiliarity with the 2015 amendments remains an issue impeding their effectiveness, the absence of a standard and cogent approach to frame proportionality assessments for the parties’ discussions and a court’s consideration has stymied real progress.

II. Overview

Experienced ediscovery lawyers, consultants, and service providers have developed many approaches to assess proportionality under Rule 26(b)(1). The New Framework adopts many of these well-developed practices. It sets out a standard framework to apply the proportionality factors in a systematic process in assessing whether discovery is proportional to the needs of the case.

As its name implies, the New Framework provides a skeletal outline that classifies custodians and their respective data sources (defined as both electronic and physical) in four priority and discovery-burden quadrants displayed in a heat map. Custodians with significant and less burdensome discoverable matter are identified in one quadrant, while other custodians who may possess or control less significant and more burdensome discoverable matter are identified in other quadrants. A heat map table contains the custodian prioritizing, data-source burden assessments, and projected discovery costs for designated custodians and key data sources.

The New Framework sets out this key data in an organized format, so that the respective Rule 26(b)(1) proportionality factors can be assessed in a focused and concrete context, which facilitates holistic considerations instead of examining matters indiscriminately in isolation. For example, the differences in assessing the “parties’ resources,” one of the six Rule 26(b)(1) proportionality factors, in relation to a custodian ranked in the heat map’s fourth quadrant as having little significant discovery at high burden, would be highlighted in a mass-tort MDL when both parties are well-financed, when compared to an employment discrimination case handled by a solo practitioner against a corporation.

The New Framework provides the foundational data for these Rule 26(b)(1) proportionality assessments, which must account for all six proportionality factors. Best practices will be developed that build on the New Framework and provide guidance to apply the six factors in relation to the significance, burden, and projected cost of the discovery.
Unlike cost and burden where objective criteria are used, rankings of custodians according to the significance of their discoverable matter involves more subjective judgments, upon which reasonable minds can differ. Although the New Framework provides three criteria to assess the significance of discoverable matter, best practices will also be developed that provide more guidance, including a sampling process to substantiate individual custodian rankings, if necessary.

Because classifications and projecting costs are highly dependent on individual circumstances, the New Framework contains multiple appendices, matrices, and worksheets, which list pertinent factors and variables to provide guidance on how to classify custodians and project costs in a systematic fashion. And where possible, the New Framework suggests baseline reference points for standard-unit costs and degrees of burden for reasonable discovery under typical circumstances, which can be adjusted for individual circumstances. The reference points are intended to facilitate the classifications and projections, especially for those less familiar with ESI and proportionality.

The New Framework’s heat map and cost tables can be adjusted to account for evolving and new information learned through interactions and negotiations with opposing counsel. The New Framework is designed to be universal and to apply to all cases large and small, including complex commercial litigation (B2B), single plaintiff, class action, and investigations/trade secrets matters.

Sections 1-3 describe the New Framework’s main functions, including the priority grouping of custodians and non-custodian data sources, data-source burden and effort, and elements used in estimating costs. Section 4 describes the plotting of custodians on priority and burden-sliding scales, the strategic uses of the New Framework, party cooperation considerations, and the generation of a record, which will inform party negotiations and judicial resolution of discovery disputes, if required.

III. Concluding Thoughts

The New Framework’s model sets of data source rankings and standard-unit cost estimates provide guidance. They are projections of reasonable ranges of costs at the outset of litigation, based on wide experience with diverse data sources and volume. They represent the collective judgment and experiences of the New Framework’s lawyers, consultants, and other ediscovery experts and are grounded in numerous cost studies and surveys (see appendices).

The projections and estimates of the New Framework’s model sets, however, are subject to individual circumstances. Accordingly, the underlying worksheets and calculators used to create the model sets are explained and appended to allow adjustments to account for variances in individual cases.
Section 01: Custodians

I. Introduction

The New Framework establishes a standard and cogent approach to apply the Rule 26(b)(1) proportionality factors in assessing whether discovery is proportional to the needs of the case. Section 1 provides guidance on how to group custodians in four broad categories, from highest to lowest priority. Section 2 gives instruction on how to identify the eight most common data sources and rank them by degree of burden in accessing information. The results of the custodian prioritizing and ranking of data sources provide a foundation for displaying the results in a graphical heat map, which visually distinguishes custodians with significantly useful information from those with marginally useful information, along with the respective attendant burdens.

Section 1 sets out objective criteria used for guidance to prioritize custodians; but in the end, all the decisions represent judgment calls by those making them, subject to challenge by others and subject to adjustment as the litigation progresses and new information is learned. Though not conclusive, early identification of many [most] of the high-priority custodians under the New Framework is often apparent and without dispute. The initial results may fully satisfy the discovery objectives, or more likely, sharpen additional probing. If challenged or otherwise appropriate, the classifications of custodians can be sampled for quality control purposes.

II. Identification of Relevant Information and Custodians

The New Framework does not create a new process to identify the scope of relevant matter or to identify custodians or data sources that are relevant. Lawyers typically take routine initial discovery steps at the outset of litigation to identify relevant information, its sources, and its custodians. The New Framework comes into play to prioritize custodians only after the relevant information, custodians, and data sources have been identified under traditional means. A brief summary of the steps commonly taken to identify relevant information and custodians is needed to place the New Framework into context.

The first step in discovery is to identify the scope of relevant matter. Every case is unique, and there is no standard procedure. There are common documents, however, that are typically examined to define the scope of relevant matter in an individual case, including:

- Complaint(s) / charges / other pleadings
- Demand correspondence / other correspondence with the opposition
- Key documents, interviews, witness statements, and investigative materials
- Initial disclosures / Rule 34 early document requests and productions or pre-suit discovery

After the broad scope of relevant matter is defined, the second step is to identify data sources containing relevant information and all corresponding custodians. Again, this is not a new idea, and many different approaches can succeed. Every approach should consider all reasonably available resources to identify likely persons with knowledge of the relevant facts, as well as potential sources of pertinent data and information, whether custodian or non-custodian.
The following is a list of items that are typically considered to start the process by identifying obviously key custodians with relevant information:

- Organizational charts
- Discovery in similar litigation
- Industry / Market / Business assessment

Further fact investigation of the obvious main custodians, their communication patterns, and their role in the underlying issues will help identify other potential custodians with relevant information (e.g., subordinates, managers, assistants, predecessors, successors, colleagues, or data stewards). Interviews, written requests, or data sampling are techniques often used to further the investigation.

Following the preliminary investigations, a list of custodians, their data sources, and categories of likely relevant information, including title, position, dates of employment, and relationship to the issues is typically compiled to begin the discovery search. Third-party custodians who may be in possession or control of potentially relevant information, key event timelines, and temporal scope of preservation efforts need to be considered as well.

III. Prioritizing Relevant Information

After the general scope of relevant matter is ascertained and the universe of custodians identified who potentially possess or control the relevant information, the analysis classifies the information by its level of priority, from low to high priority. The New Framework applies three criteria to prioritize relevant information, including: (1) materiality of the information; (2) strength of the information; and (3) uniqueness of the information.

A. Materiality of Information

Some relevant information may be more useful, significant, or important than other relevant information. The New Framework’s first criterion focuses on “materiality,” which is defined as significant or essential, to discern qualitative differences in relevant information. All relevant information falls on a spectrum of significance, and where materiality starts on that spectrum is a matter of judgment. The key is whether the information is material because it is of such a nature that its knowledge would affect a person’s decision-making process. The stronger the materiality of information, the higher the priority it is assigned.

There are several indicators of materiality, including whether the information:

- goes to the heart of the case or addresses a subsidiary issue;
- proves an ultimate fact or an intermediate fact; and,
- is an essential link in a line of evidence needed to prove an assertion.

B. Strength of the Information

The weight of relevant information in proving an assertion will vary. The New Framework’s second criterion focuses on the “strength” of the information to distinguish the weight of relevant information based on how directly it is connected to the asserted fact.
Although considerations of materiality will overlap, the strength of the information may be indicated by:

- whether it provides direct or circumstantial evidence; and
- whether the information is complete and thorough or limited and partial.

The stronger the information, the higher the priority it is assigned.

C. Uniqueness of the Information

In a certain sense, every document and piece of information is unique. The New Framework’s third criterion characterizes uniqueness by distinguishing similar information from information that is qualitatively different. (Exact duplicates are not pertinent, because they are eliminated as part of routine deduplication processes.) The stronger the uniqueness of the information, the higher the priority it is assigned.

There are no bright-line tests to distinguish unique information, which will depend on the circumstances. Identifying unique information is a judgment call by those making the decisions, like many other decisions prioritizing custodians and information. But a growing number of courts have posited “unique, relevant information” as a standard in their proportionality analyses to distinguish discovery that is not proportionate to the needs of the case. As case law matures, the evolving standard will become clearer and provide more guidance.

IV. Prioritizing Custodians and Non-Custodian Data Sources

Under the New Framework, custodians and non-custodian data sources are prioritized by the level of relevant-priority information they possess or control. Along with this information, a custodian’s position, level of knowledge, and depth of involvement in the particular issues must also be taken into account when prioritizing them.

The role of the custodian within an organization, the nature of the custodian’s involvement, and the pertinent time period of the custodian can add critical gloss to the priority of information that they possess or control. Whether the custodian has personal firsthand knowledge, the information is secondhand knowledge, or comes from a third-party source are also factors to consider.

A. Standardized Report Format

In a case with few custodians, prioritizing them can be readily apparent after minimal investigation. And every custodian can be promptly plotted on the New Framework’s heat map in one of four quadrants: (1) highest priority, (2) high priority, (3) medium priority, and (4) low priority. But in cases involving a multitude of custodians, standard procedures regarding gathering and recording the results of investigations are needed to provide more uniform results.

A standardized report format can facilitate the prioritizing of information and custodians by recording the investigation results of applying the three criteria of materiality, strength, and uniqueness to assess the priority of the relevant information along with the custodian’s connection to the information.

Written requests in the form of a survey, interview, or data sampling are techniques often used to gather the information for the report. The purpose of the report is to organize the results of the investigations and provide a master score for each custodian based on the value of information they possess or control as determined by the three criteria and the custodian’s connection to the information. The master scores of the custodians are used to plot the custodians on the New Framework’s heat map.
Although no single format can effectively handle all cases, Appendix A suggests an initial survey containing a series of questions to inform the scoring of custodians. This can then be used as a template for the report.

B. Atypical Use Cases

Special situations or atypical cases will require different handling. Atypical cases may involve a departed employee, non-custodian data, or information in other countries subject to foreign privacy laws. The New Framework can be adapted to fit those circumstances.

a. Departed Employees: A legal hold is in place, but one of the subject employees is no longer with the company. Appendix B illustrates the adaptations to the New Framework.

b. Non-custodian Data: Data is non-custodian and a traditional custodian interview is not feasible. Appendix C illustrates the adaptations to the New Framework.

c. International Custodians: If international custodians are involved, local data privacy laws must be considered. Appendix D provides an example scenario that takes privacy laws into account.
Section 02: Defining Data-Source Burden and Effort

I. Introduction

The New Framework’s standard approach requires the designation of data sources by degree of burden. Section 2 provides guidance on how to rank five of the most common data sources in four broad categories, based on the degree of burden incurred in accessing information. The five primary variables identified directly affect the degree of burden for each data source. The specific degree of burden associated with each data source is not predefined; rather, it is dependent on circumstances.

Although the ranking of data sources by burden using the five variables can be done on a case-by-case basis, the New Framework projects rankings of data-source burdens that can be reasonably expected in many, perhaps most, instances. The model rankings are based on a literature review and the collective judgment and experiences of the New Framework’s experts, applying the five variables as they most commonly appear in a typical case. They are intended to provide guidance and reference points, especially to those unfamiliar with ESI, when ranking the data sources. A graphic illustrating the model rankings of eight data sources is set out at the end of this section.

Individual circumstances will require adjustments to the model rankings. The clearest example are emails, which typically are located on a centralized server, providing relatively easy access and modest burden, compared with the less common use of decentralized email, which is located on individual computers, requiring multiple extractions and significantly increasing the burden.

The following discussion identifies the eight most common data sources and explains the variables that affect the burden assessments.

II. Data Sources and Types of Data

Listed below are the eight most common sources of ESI, which are considered under the New Framework:

- Collaboration / Messaging Systems (Slack, Teams)
- Computers / Laptops
- Email Systems
- File Shares (departmental and personal)
- Mobile Devices
- Paper / Physical Evidence
- Social Media
- Structured Systems (HR, finance, marketing databases)

Specialized Data Sources

Listed below are specialized data sources not considered under the New Framework, but can arise in an individual case:

- Backup media
- Computer code
- Corporate telephone data
- Ephemeral data
- Geolocation, GPS, IP addresses
- Specialized/proprietary databases, systems, or programs
- Website data
III. Variables Affecting Burden Assessments

Five major variables affect the burden assessment, including:

A. The location of and accessibility to the data
B. The availability or state of readiness for collection
C. The methods used to preserve and collect data
D. Any specialized resources needed to effectuate collection
E. Any legal or regulatory constraints (e.g., data privacy laws) that may impact collection

These variables are central in determining the burden associated with collecting ESI.

A. Access to and Location of the Data

The accessibility and physical location of data sources can significantly impact the effort and burden of collection. The more centralized the data is, the easier it is to access the information, thereby lowering the burden.

For instance, if the data is centralized in a single server room or in a single cloud instance, it is more readily accessible than if it is spread out across multiple data centers or cloud storage locations that are not interconnected. The geography of US-based and international locations, as well as legal and regulatory requirements described in Section E below, may also place a greater burden on coordination and collection efforts.

It is important to determine whether sources like desktops, laptops, and mobile devices can be accessed remotely from a central location, or whether the collection can only be accomplished by physically traveling to the actual place where the data source resides.

Many organizations now allow employees to use their own personal devices to connect to their organizational networks and access work-related systems and potentially sensitive or confidential data. BYOD situations present a new challenge to logistical collection efforts, and also raise significant privacy concerns.

B. Availability of Data

The state of the data and its ready availability impact the degree of burden. Accessing data from a departed employee increases the burden.

Thus, if the data belongs to a current employee, it is more readily available than data belonging to a former employee, which may have been archived or dispersed. A current employee’s data is typically in an active state and readily available for collection by the custodian, IT, or vendor. If the employee is no longer employed and the data source is no longer in use or has been deactivated, availability to the data becomes more challenging.

C. Preservation and Collection Methods

The preservation method may impact the burden as well. If special processes are required to preserve information before it can be accessed, the burden is higher.

In some instances, the only way to defensibly preserve data without altering it is through collection. In other circumstances, like those in which an advanced document management or filing system is in place, it may be possible to preserve data in place with little or no effort.
The method of collection can also impact the burden. A decentralized organization might require collection by the custodian, which would most likely entail custodian training to ensure against spoliation. On the other hand, some organizations may have a native collection tool that allows for in-place or nearline collection.

Finally, specialized data sources may require forensic or customized queries and system knowledge to collect data efficiently and safely.

**D. Specialized Resources to Collect**

There are instances in which the information from a data source that would typically be considered easier to collect may impose added burdens because of age, corruption, or malintent. In such circumstances, it may be necessary to engage specialized tools or resources that significantly increase the degree of burden of collection.

Additionally, there may be corporate resources such as structured databases that require expert knowledge and understanding of the system in order to correctly and safely extract the pertinent data, leading to increased burden of collection.

**E. Legal and Regulatory Restrictions and Requirements**

Legal requirements can also add to the burden of collection. Privacy legislation and regulations may regulate or require a party to take certain steps before the data is collected. The E.U. General Data Protection Regulation (“GDPR”) applies to any processing of personal data when a European company is the controller. The GDPR has specific legal requirements for the processing of personal data (preservation, collection, processing, and production are separate data processing activities). For instance, each data processing activity needs a legal basis, and you have obligations towards the individual.

In addition, transfer of the personal data for use in litigation discovery in the U.S. is only possible if certain conditions are met. Moreover, there are labor law regulations, blocking statutes (foreign and domestic), and state secrets law in other nations that should be considered. Failure to observe or properly comply with these legal and regulatory restrictions and requirements may lead to fines or even criminal consequences. Therefore, compliance with these legal and regulatory restrictions and requirements result in, among others, additional consultations and assessment steps involving local counsel or authorities, which can add to the burden of collection.

**IV. Assigning Burden**

The New Framework’s model set of rankings of the eight most common data sources is based on variables that typically can be expected to apply to provide baseline reference points, which are graphically displayed below. Accordingly, the low degree of burden associated with typical email is ranked at one end of the spectrum, compared with mobile devices, which are ranked at the other end of the spectrum under the New Framework. Custodians with high-priority information at lowest discovery burden are quickly identified. In its current form, the chart below represents burden factors relating only to collection of the data sources.
Although this model set of rankings will apply in many, perhaps most cases, the variables in an individual situation may be atypical, and sometimes require significant adjustments to accurately reflect the actual burden. Recognizing the possibilities that the five variables may be different for a particular data source in an individual case and require adjustments, a Burden Assessment Tool has been developed to assist in leveraging the New Framework to account for atypical variables and adjusting the degree of burden for the identified data source. The tool can be used to select specific factors that are then evaluated to assign an approximate burden level of Highest, High, Medium, or Low to each data source type.

The Burden Assessment Tool is contained in Appendix E.
Section 03: Discovery Cost Projections

I. Introduction

In addition to the heat map of custodians and data-source burdens, a table of discovery costs projected for every custodian and data source is part of the New Framework’s key data, which provides the base for applying the Rule 26(b)(1) proportionality factors. Similar to the ranking of data sources by burden, estimating discovery costs must be done on a case-by-case basis, relying on the itemized costs listed in the worksheets and calculators.14

Studies have reported varied ranges in per-gigabyte or per-document discovery costs. A model set of projected costs for five common data sources is set out at the end of this section, which provides baseline reference points. Section 3 explains how ranges of discovery costs are projected at the outset of litigation, based on wide experience with diverse data sources and volumes.

Experienced ediscovery lawyers and service providers may be routinely using customized search techniques and methods, which result in cost and efficiencies different from those that are reasonably expected in general. Guidance and multiple worksheets and calculators are provided to make adjustments in the projected discovery costs in individual cases to accommodate these situations.

II. Projecting Total Discovery Costs

The New Framework’s discovery cost projections are organized around the discovery workflow found in the EDRM, a generally accepted model for ediscovery practice, including: Identification, Preservation, Collection, Processing, Review, Analysis, and Production. Similar to the approach taken by RAND in its seminal 2012 discovery cost report, Where the Money Goes, the New Framework collapses costs into three main categories: A) Collection; B) Processing and Hosting; and C) Review and Production.15 The New Framework does not profess to include all costs incurred in ediscovery, which can include substantial sunk equipment and software costs, incidental overhead costs, and staff compensation.

Although various ranges of discovery costs have been reported, studies over the past decade have shown consistency in the overall comparative percentages attributed to these three stages. Ediscovery costs are comprised of approximately 65%-75% for review, 10%-20% for processing, and 5%-15% for collection.16 These proportions offer a useful crosscheck in verifying estimates.

A. Collection Costs

Section 2 describes generally the collection process of gathering the subset of potentially discoverable ESI. Preservation costs are not included in this category.17 Many different methodologies have been used to collect ESI; some more efficient, effective, and less expensive than others. Costs can be mitigated using a targeting or sampling strategy, the implementation of remote collection, or examination and elimination of redundancies prior to collection. Although the methodologies vary, collection costs typically account for approximately 5%-15% of the overall cost, with the caveat that the proportion may vary dramatically depending upon the circumstances of an individual matter.

Once identified, the data must be collected in a format compatible with standard processing methodologies, offered by ediscovery vendors, or in-house tools to allow processing to normalize data formats. Collection costs include fees for in-house technicians, software, and hiring of external specialists.
Rates for collection vary depending upon the methodologies required. But the hourly or per-device rates for data collection have remained relatively flat over the period 2004-2020, even though data volumes, types of devices, and information stores have proliferated. During the same period, the efficiency of collection tools and the sophistication of the individuals involved in the process (e.g., IT professionals, attorneys) have improved, which has been effective in reducing the volume of information being collected.

B. Processing and Hosting Costs

The bulk of discovery spend for processing consists of costs for culling the volume of data to eliminate nonrelevant data, which may include purchasing or leasing software and the attendant staff expense, in order to minimize the data that must be reviewed by attorneys for relevancy and privilege, the single most costly outlay.

Collected data from different sources is typically processed by ingesting it, using a specialized tool designed to normalize data formats and optimize the data for the various search, analytic, and culling functions to follow.18 The resulting data set is then prepared for and migrated to a review tool that allows either linear document review, advanced analytic options, or both. The processing stage typically accounts for approximately 10%-20% of overall ESI project outlay.

The fees for routine processing are generally standard and are typically charged based on volume or gigabyte. The process usually includes ingestion of data, deNisting and deduping, metadata extraction, creation of search and analytics indices, culling based on selected criteria such as date or domains, and the preparation of various exception reports for chain of custody purposes. Additionally, recent tools that remove near-duplicate files and provide email threading can reduce the volume of data in a logical and systematic manner.19

In addition to standard fees, however, other fees may be incurred depending on the chosen workflow or additional actions that may be applied to the data, such as advanced analytics (sometimes referred to as artificial intelligence), Technology Assisted Review (TAR)20 or Continuous Active Learning (CAL).

Regardless of the workflow chosen, data is typically hosted externally and incurs a monthly hosting (storage and maintenance) cost.

C. Review and Production Costs

The review and production costs make up approximately 65%–75% of a producing party’s overall outlay. The principal driver of these costs is attorney review time.

When the volume of the reviewable data is not large, accounting for the cost of a handful of lawyers reviewing can be straightforward. But in intensive discovery litigation, review is frequently assigned to many lawyers at various per-hour rates. If a law firm is involved, work is frequently assigned to at least two different types of attorneys: independent contract attorneys, not employed by firms, often as part of an agency (sometimes referred to as managed review); and firm attorneys, which can involve both associate and more senior attorneys, particularly for privilege review. The difference in cost between and among these different types of attorneys is significant.

Generally, the overall cost of review turns largely on several considerations: (a) how much of the attorney time involves contract attorneys, as opposed to firm attorneys; (b) the cost of privilege review; (c) how long it takes attorneys to review the data; and (d) the volume of data or documents attorneys review or put eyes on.
The cost of review depends not only on the number of documents to be reviewed, but also the amount of time it takes attorneys to review different categories of data (e.g., emails versus dense and complicated excel spreadsheets or engineering drawings). If the data has been normalized in the processing stage (i.e., text messages, emails, and Slack conversations look the same on the review platform), it can reduce review time.

In addition to the overall number and nature of documents to be reviewed, the balance of time and utilization between contract attorneys and firm attorneys may depend on several factors, including the potential sensitivity of production decisions (e.g., stakes involved, overall sensitivity of the corpus) as well as workflow decisions regarding the number of tasks required of the reviewers created by the handling of email threads, families, and redactions. 21

Production is the process of turning the results of the attorney review into something that can be produced to the requesting party. The costs associated with production are lower and more predictable than the cost of review. Production is typically handled by the vendor hosting the data being reviewed.

III. Model of Projected Costs Table

The New Framework provides a model set of projected discovery cost ranges at the outset of litigation, which are expressed in standard units, e.g., $ per gigabyte, $ per document, for typical volumes of data for five data sources. (See Appendix H.) The projections are based on a literature review, standard industry-wide assumptions, and the collective judgment and experiences of the New Framework’s experts. 22 They are intended to provide guidance and reference points, especially to those unfamiliar with ESI, when projecting discovery costs. A graphic table containing the model set of projected costs is set out below.

The use of the model set of cost projections has limitations, however, because it is subject to individual circumstances. To account for these circumstances, and to permit a user to modify the projected costs, rates, percentages, and other data to fit particular circumstances, an excel spreadsheet is provided, which includes a comprehensive Cost Prediction Calculator with eight separate tabs. The first seven tabs explain and summarize the more than 100 line-by-line itemized data and projections set out in the eighth and last tab, "Calculations.” The formatting of the line-by-line itemizations under the Calculations tab can be easily modified. Click here to open the full calculator.

Appendix F contains three of the Cost Prediction Calculator’s spreadsheet tabs, which summarize the cost projections, with easily printable charts. Appendix G lists common variances that will require adjustments to the model set of projected costs.

IV. Individualized Itemized Cost Prediction Calculators

As its name implies, the itemized cost prediction calculators list the individual costs incurred in estimating discovery for eight specific data sources. They can be used to project the estimated costs in discovery.

The calculators were used to produce the aggregate discovery costs in the model set of projected costs for each of the eight most common data sources with data based on the collective judgment and experience of the experts developing the New Framework and the results of an extensive literature review. The calculators can be modified to make adjustments to the model set of projected costs to account for individual circumstances.
V. **Common Variances Requiring Adjustments**

Circumstances can arise that may add complexity to the process or procedures that could increase costs. A sampling of such factors can be found in Appendix G, which includes a multiplier that can be applied to adjust costs estimated by the model set of projected costs.

**Section 04: Heat Map, Heat Map Table, and Application of New Framework**

I. **Introduction**

Thus far, the workflow is as follows:

- **Section 1** identifies and assesses custodian and non-custodian data sources into four levels of priority: Highest, High, Medium, and Low.
- **Section 2** examines the burden and effort involved in the collection of various types of data sources, which are then categorized into four levels of burden: Highest, High, Medium, and Low.
- **Section 3** offers a Cost Prediction Calculator that estimates the cost of discovery for moving various data source categories downstream through collection, processing, review, and production. Factors that may add a layer of complexity to an ESI project are also identified.

Section 4 pulls together the custodian prioritizing, burden assessments, and cost projections, displaying them in two formats: (a) a heat map, which organizes the custodians and their data sources on a sliding priority and burden scale and plots them in four quadrants; and (b) a heat map table, which sets out the assessments and adds ranges of cost projections for every custodian and data source.

The weight accorded the custodian prioritizing and burden assessments in the New Framework’s heat map and cost projections will be affected by the extent of the opposing party’s input, if any. The more input from the opposing party, the less likely discovery disputes will occur and, if a dispute does arise, the judge will be better informed to resolve it. Sampling may offer a practical method in assessing particular custodian prioritizing to ensure accuracy or if challenged.

II. **New Framework Heat Map**

The custodian prioritization and data-source burden assignments under Section 1 and Section 2 are placed into a heat map for a simple and clear graphical representation. The custodians and non-custodian sources are plotted on a quadrant grid on the Y-axis from Highest to Low Priority, and the data-source burdens are plotted on the X-axis from Highest to Low Burden. Each quadrant is further broken down into an additional four sectors, showing a total of 16 groupings of Priority / Burden scale.

The following illustration is an exemplar.
III. New Framework Heat Map Table

The New Framework heat map table lists the prioritized custodians in Section 1, the data-source burdens in Section 2, and the projected costs in Section 3. The table is an exemplar.

### New Framework Heat Map Table (Exemplar)

<table>
<thead>
<tr>
<th>Name</th>
<th>Priority</th>
<th>Data Source</th>
<th>Level of Burden</th>
<th>Cost Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custodian “A”</td>
<td>Highest</td>
<td>Email</td>
<td>Low</td>
<td>$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer</td>
<td>High</td>
<td>$</td>
</tr>
<tr>
<td>Custodian “B”</td>
<td>Highest</td>
<td>Email</td>
<td>Low</td>
<td>$</td>
</tr>
<tr>
<td>Custodian “C”</td>
<td>High</td>
<td>Email</td>
<td>Low</td>
<td>$</td>
</tr>
<tr>
<td>Custodian “D”</td>
<td>Medium</td>
<td>User Share</td>
<td>Medium</td>
<td>$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Computer</td>
<td>High</td>
<td>$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social Media</td>
<td>Highest</td>
<td>$</td>
</tr>
<tr>
<td>Custodian “E”</td>
<td>Low</td>
<td>Email</td>
<td>Low</td>
<td>$</td>
</tr>
<tr>
<td>Non-custodian</td>
<td>Highest</td>
<td>Accounting</td>
<td>Highest</td>
<td>$</td>
</tr>
<tr>
<td>Non-custodian</td>
<td>High</td>
<td>Personnel Database</td>
<td>Highest</td>
<td>$</td>
</tr>
</tbody>
</table>
IV.  Applying the New Framework at Different Stages of Litigation

The New Framework can be applied at every stage of the litigation lifecycle, including commencement of litigation, discovery management, and close of discovery. By establishing a standard approach to frame proportionality assessments, the New Framework focuses attention on key issues, creates a common analysis with standard vocabulary, and sharpens proportionality assessments.

A.  Commencement of Litigation

The New Framework’s heat map and heat map table can serve several useful purposes early in litigation. Although the information available at the commencement of litigation often is preliminary and inadequate to make firm proportionality assessments that are reliably certain about all custodians, sufficient information is usually available to make early assessments of custodians identified at both priority extremes on the New Framework’s heat map.

The degree of confidence in these early assessments and cost projections can be affected by the extent of the opposing party’s input, if any, into the development of the New Framework’s heat map and heat map table. Sampling may be effective to assure accuracy.

The early assessments as well as cost projections can have immediate dividends, including:

- The heat map table’s cost projections will provide information necessary to plan a budget covering likely discovery expenses.
- The database cost projections as well as the calculator accounting for variances provide baseline reference points to compare and evaluate RFPs, which can be particularly helpful for those less experienced in ESI.
- The information about priority custodians in the heat map will better inform the decision to make early productions in accordance with Rule 26(d)(2) and initial disclosures under Rule 26(a) by identifying obviously significant custodians which will assuredly result inevitably in discovery productions.25
- Custodians not tagged at either priority extreme in the heat map are identified who need further investigation, which may involve additional written questions, interviews, or sampling.26
- The preliminary assessments in the heat map will better inform preservation decisions by highlighting information that is obviously important and unquestionably preservable compared with information that is less significant.27

B.  Discovery Management

As additional information becomes available during discovery, including information learned from interactions with the opposing party, the custodian prioritizing, burden assessments, and cost projections can be adjusted so that the application of the proportionality factors can become firmer.28

The New Frameworks’ heat map and heat map table distill and concentrate key information in an organized format, which is necessary to apply the six Rule 26(b)(1) proportionality factors. These two sources are critical to ground the proportionality decision-making. The granular information on data-source burdens and costs for every custodian and their respective data sources enhances the analysis of the six proportionality factors, which will, for example:
• Better inform negotiations and decision-making at the Rule 26 meet and confer conference as well as a later Rule 16 pretrial conference;\textsuperscript{29}
• Help the parties better craft an ESI protocol, which considers all pertinent information;
• Help the parties develop a phased-discovery plan, when appropriate, identifying which collections and custodians should go first and which should follow sequentially;\textsuperscript{30} and,
• Better inform cost implications when negotiating production format decisions, such as whether to produce data in native, image, or mixed native/image format, whether to produce images with color, how to handle redactions, and how to log privileged documents and confidentiality designations, when considered with all other pertinent information shown in the heat map and heat map table.

C. Close of Discovery

The application of the Rule 26(b)(1) proportionality factors based on the New Framework’s heat map and heat map table can be used to address proportionality-related issues, which may arise at the end of the discovery stage.

• The decision when to release litigation holds is better informed by the New Framework’s heat map, which highlights individual custodians most likely to possess information that can be released.\textsuperscript{31} (See Appendix D.)
• Information in the New Framework’s heat map and heat map table better informs decisions to reopen discovery.\textsuperscript{32}

The New Framework’s workflow results in the heat map and heat map table and provides a record documenting the decisions and actions that occurred throughout the litigation lifecycle. The documentation is essential from a standpoint of establishing defensibility of process.

V. Judicial Resolution of Discovery Disputes

The information in the New Framework’s heat map and heat map table provides a judge with essential information in an organized format to apply the Rule 26(b)(1) proportionality factors and evaluate whether counsel’s discovery efforts are reasonable and in good faith and whether discovery is proportional to the needs of the case. The extent to which opposing counsel provided input into the New Framework’s heat map and heat map table will be a factor that a judge will consider when evaluating the weight to be accorded the proportionality assessments. Sampling offers a potential practical solution to evaluate disputes about the prioritizing of custodians.

Most significantly, the New Framework’s standard approach framing proportionality assessments presents a fuller picture of all the potential data sources, custodians, and attendant costs that counsel must consider and evaluate under Rule 26. Under this approach, a judge can better evaluate the overall reasonableness of counsel’s proportionality assessments, underlying rationales, and discovery decisions.

The New Framework’s model set of projected costs for each data source provides a judge a reference point to evaluate suspiciously high claimed costs along with an itemized cost calculator that can be used to verify the variances, which may explain the discrepancy in cost.

A judge can rely on the New Framework’s custodian prioritizing and data-source burden assessments and cost projections at several key litigation mileposts, including:
• A judge can consider the information to evaluate the scope of a requested preservation order. Custodians possessing marginally significant information in burdensome data sources are identified and can be scrutinized, which considerably narrows the preservation analysis.33

• A judge can consider the same information at the Rule 16(b) conference, in evaluating and deciding the scope and sequencing of discovery. The New Framework’s custodian prioritization and cost projections provide a base to apply the proportionality factors and provide a ready-made roadmap for the sequencing of discovery, starting with custodians with high-value information at low burden and moving to custodians with less significant information at high burden.34

• A judge can use the information to evaluate and resolve discovery disputes, involving a motion to compel or a motion for a protective order. Judges are routinely requested to rule on motions asking discovery for “x” number of additional custodians or to limit the number of custodians to “x.” The heat map table provides concrete information on the costs projected for each additional custodian, and most importantly, the cost is given by data source, providing the judge the capability to make more precise rulings that consider every data source.35

• A judge can use the information to evaluate and resolve sanction motions. Reasonableness is the overarching issue for a judge to consider when evaluating counsel’s discovery actions. The New Framework’s heat map and heat map table provide a record of all custodians, data sources, expenses, and burdens that counsel was faced with in making proportionality assessments and taking action. The record’s full accounting better informs the judge’s decision on the reasonableness of any specific discovery action taken by counsel.36 The judge may also consider the extent that opposing counsel provided input, if any, to the development of the custodian prioritizing and data-source burden assessments and cost projections, in determining whether counsel’s action was reasonable.
Appendix A – Survey Questions to Assess Relevant Information and Custodians

Begin by creating a series of questions designed to assess the outlined criteria in relation to the custodian. Some of these questions may be generic and used in all matters, while some will drill down on the specific knowledge level or understandings pertinent to the specific claims and defenses. The survey questions may also include disqualifying questions such as employment period, position in organization, and duties and responsibilities. The questions may also include an overriding qualifying question that would immediately place a custodian in the Highest category regardless of any other assessment scores.

1. Create a few simple qualifying questions to filter immediate disqualification; e.g., employment outside date range. This will generate a custodian qualification of RELEASE or RETAIN.
2. Limit the number of substantive assessment questions to no more than five to six to reduce ambiguity and maintain objectivity.
3. The mechanics of ranking the responses may vary and include:
   a. Weighting certain responses over others
   b. If / then logic applied to certain responses
   c. AND, OR, NOT logic used with certain responses
4. The rank of each question is aggregated to create an overall master score for each custodian.
5. This master score identifies the custodian’s overall potential importance (PRIORITY DETERMINATION) and is ranked as follows:
   a. Highest
   b. High
   c. Medium
   d. Low

Exemplar

The following is an example of an assessment survey. Note that the assessment questions and relative answers should be customized to the matter at hand.

The goal is to offer a tool that provides an objective assessment of each custodian based on a series of questions which reflect the materiality, strength, and uniqueness of their data.

This template will be made available in spreadsheet format for use in assessment creation.
The responses to the assessment questions can be ranked in any suitable manner from low to highest and then aggregated to create the master score.

Developed under the George Washington Law School James F. Humphreys Complex Center and is freely available to the public. The Center acknowledges Insight Optix for its original thinking that inspired the New Framework and its critical logistical support.
Appendix B – Atypical Case – Departed Employee

Departed Employees

- Custodian Accessible?
  - Yes: Complete Survey
  - No: Analyze Data
- Data Inaccessible
- Scoring
  - Sales Region
  - Product
  - Communication Patterns
  - Date ranges
  - Domain Analysis
  - Key Phrases
  - Language
  - Sentiment
- Assessment / Interview
- STOP
Appendix C – Atypical Case – Non-Custodian Data

Non-custodian data

Knowledgeable resource

Complete Survey

Yes

No

Analyze Data

Scoring

Assessment / Interview

Assessment / Interview

- Type of data
- Source of data
- Corpus content
- Structured / unstructured
- Date ranges
- Key Phrases
- Language
Appendix D – Atypical Case – Privacy Laws

If discoverable matter from EEA custodians (data subjects) is involved, GDPR applies and must be considered. Art. 6 GDPR defines the general scope of lawful processing (preservation / collection / review). Art. 6(1)(f) of GDPR requires that processing in litigation must be necessary for the purposes of a legitimate interest pursued by the controller and not overridden by privacy rights of a data subject.

Other GDPR provisions may be relevant in litigation. Art. 6(1)(a) GDPR permits processing with the custodian’s consent. But consent must be given freely, which is difficult to assess in any employer/employee relationship. In addition, consent can be withdrawn at any time. Although Art. 6(1)(c) GDPR permits processing if necessary to comply with a legal obligation, the provision applies only to an obligation that arises out of EU law or the law of an EU Member State to which the controller is subject.

If the information from a custodian or a data source has no materiality, strength, and uniqueness to the claims and defenses, there likely is no legitimate interest in processing under the GDPR. Therefore, such information from custodians and data sources cannot be processed, i.e., preserved, collected, or reviewed.

The responses to Appendix A’s survey questions to assess relevant information and custodians should be kept for documentation purposes to show the reasons for the determination of releasing and processing the custodian’s information.
Appendix E – Burden Assessment Tool

The Burden Assessment Tool aids in assessing each data source based on the criteria outlined in the narrative above. Once the appropriate factors are selected for a specific situation, the tool will automatically assign a classification of low, medium, high, highest burden to each data source.

This template will be made available in spreadsheet format for use in determining burden classification.

Developed under the George Washington Law School James F. Humphreys Complex Center and is freely available to the public. The Center acknowledges Insight Optix for its original thinking that inspired the New Framework and its critical logistical support.
Appendix F – Itemized Cost Prediction Calculator

The calculator includes seven tabs – three summary sheets are shown below. Click here to open the full calculator.

Judicial Dashboard

GW Discovery Proportionality Model
Judicial Dashboard

<table>
<thead>
<tr>
<th>STAGE</th>
<th>Hosted email (5GB)</th>
<th>File Share (4.5GB)</th>
<th>Social Media/Website (0.25 GB)</th>
<th>Computer/Laptop (30 GB)</th>
<th>Mobile Device (8 GB)</th>
<th>STAGE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB Collected</td>
<td>5.0</td>
<td>4.5</td>
<td>0.25</td>
<td>30</td>
<td>30</td>
<td>69.75</td>
</tr>
<tr>
<td>GB Produced</td>
<td>0.44</td>
<td>0.11</td>
<td>0.04</td>
<td>0.79</td>
<td>0.21</td>
<td>1.59</td>
</tr>
<tr>
<td>Collection</td>
<td>750.00</td>
<td>1,150.00</td>
<td>500.00</td>
<td>600.00</td>
<td>750.00</td>
<td>3,750.00</td>
</tr>
<tr>
<td>Pre-Processing</td>
<td>450.00</td>
<td>1,200.00</td>
<td>1,800.00</td>
<td>600.00</td>
<td>600.00</td>
<td>4,650.00</td>
</tr>
<tr>
<td>Processing</td>
<td>436.54</td>
<td>108.31</td>
<td>21.19</td>
<td>1,195.49</td>
<td>164.08</td>
<td>1,925.61</td>
</tr>
<tr>
<td>Hosting</td>
<td>827.33</td>
<td>514.14</td>
<td>445.21</td>
<td>1,166.16</td>
<td>612.95</td>
<td>3,565.79</td>
</tr>
<tr>
<td>Review</td>
<td>29,500.20</td>
<td>6,076.01</td>
<td>4,255.31</td>
<td>42,804.80</td>
<td>19,524.30</td>
<td>102,160.61</td>
</tr>
<tr>
<td>Production</td>
<td>233.89</td>
<td>171.23</td>
<td>157.44</td>
<td>301.67</td>
<td>191.00</td>
<td>1,055.23</td>
</tr>
<tr>
<td>Management/Support</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>6,477.64</td>
</tr>
</tbody>
</table>

DATA SOURCE TOTAL: $33,493.48 $10,515.21 $8,474.67 $47,963.65 $23,137.85 $123,584.87

Developed under the George Washington Law School James F. Humphreys Complex Center and is freely available to the public. The Center acknowledges Insight Optix for its original thinking that inspired the New Framework and its critical logistical support.
### GW Discovery Proportionality Model

#### Judicial Detail View

<table>
<thead>
<tr>
<th>STAGE</th>
<th>Hosted email (5GB)</th>
<th>File Share (4.5GB)</th>
<th>Social Media/ Website (0.25 GB)</th>
<th>Computer/ Laptop (30 GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB Collected</td>
<td>5.0</td>
<td>4.5</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>GB Produced</td>
<td>0.44</td>
<td>0.11</td>
<td>0.04</td>
<td>0.79</td>
</tr>
<tr>
<td>Collection</td>
<td>750.00</td>
<td>1,150.00</td>
<td>500.00</td>
<td>600.00</td>
</tr>
<tr>
<td>Pre-Processing</td>
<td>450.00</td>
<td>1,200.00</td>
<td>1,800.00</td>
<td>600.00</td>
</tr>
<tr>
<td>Processing</td>
<td>436.54</td>
<td>108.31</td>
<td>21.19</td>
<td>1,195.49</td>
</tr>
<tr>
<td>Hosting</td>
<td>827.33</td>
<td>514.14</td>
<td>445.21</td>
<td>1,166.16</td>
</tr>
<tr>
<td>Review</td>
<td>29,500.20</td>
<td>6,076.01</td>
<td>4,255.31</td>
<td>42,804.80</td>
</tr>
<tr>
<td>Production</td>
<td>233.89</td>
<td>171.23</td>
<td>157.44</td>
<td>301.67</td>
</tr>
<tr>
<td>Management/Support</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>1,295.53</td>
<td>1,295.53</td>
</tr>
<tr>
<td><strong>DATA SOURCE TOTAL:</strong></td>
<td><strong>$ 33,493.48</strong></td>
<td><strong>$ 10,515.21</strong></td>
<td><strong>$ 8,474.67</strong></td>
<td><strong>$ 47,963.65</strong></td>
</tr>
</tbody>
</table>

*Developed under the George Washington Law School James F. Humphreys Complex Center and The Center acknowledges Insight Optix for its original thinking that inspired the New Framework*
<table>
<thead>
<tr>
<th>Mobile Device (8 GB)</th>
<th>STAGE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>69.75</td>
</tr>
<tr>
<td>0.21</td>
<td>1.59</td>
</tr>
<tr>
<td>750.00</td>
<td>3,750.00</td>
</tr>
<tr>
<td>600.00</td>
<td>4,650.00</td>
</tr>
<tr>
<td>164.08</td>
<td>1,925.61</td>
</tr>
<tr>
<td>612.95</td>
<td>3,565.79</td>
</tr>
<tr>
<td>19,524.30</td>
<td>102,160.61</td>
</tr>
<tr>
<td>191.00</td>
<td>1,055.23</td>
</tr>
<tr>
<td>1,295.53</td>
<td>6,477.64</td>
</tr>
<tr>
<td><strong>$ 23,137.85</strong></td>
<td><strong>$ 123,584.87</strong></td>
</tr>
</tbody>
</table>

is freely available to the public.
and its critical logistical support.
Attorney Detail View

<table>
<thead>
<tr>
<th>GW Discovery Proportionality Model</th>
<th>Attorney Dashboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosted email</td>
<td>File Share</td>
</tr>
<tr>
<td>Social Media/Website</td>
<td>Computer/Laptop</td>
</tr>
<tr>
<td>Mobile Device</td>
<td></td>
</tr>
</tbody>
</table>

| Subtotal: Collection per custodian | -                   |
| Subtotal: Collection per data source/device | 750.00 | 1,150.00 | 500.00 | 600.00 | 750.00 |
| Subtotal: Collection hourly rate  | -                   |
| Subtotal: Forensic hourly rate    | -                   |
| SUBTOTAL COLLECTION               | 750.00 | 1,150.00 | 500.00 | 600.00 | 750.00 |
| SUBTOTAL ESTIMATED GB             | 5.00   | 4.50     | 0.25  | 30.00  | 8.00   |
| SUBTOTAL PRE-PROCESSING           | 450.00 | 1,200.00 | 1,800.00 | 600.00 | 600.00 |
| Subtotal: Ingestion total fee     | 200.00 | 59.06    | 6.15  | 843.76 | 69.00  |
| Subtotal: Analytics fee           | 168.00 | 31.89    | 8.94  | 227.81 | 61.58  |
| Subtotal: Processing fee          | 68.54  | 17.35    | 6.08  | 123.93 | 33.50  |
| SUBTOTAL PROCESSING               | 436.54 | 108.31   | 21.19 | 1,195.49 | 164.08 |
| Subtotal: Hosting GB fee          | 419.33 | 106.14   | 37.21 | 758.16 | 204.95 |
| Subtotal: User licensing          | -       | -       | -     | -     | -      |
| SUBTOTAL HOSTING                  | 827.33 | 514.14   | 445.21 | 1,166.16 | 612.95 |
| Subtotal: 1st Pass review fee     | 7,257.60 | 1,530.90 | 1,073.32 | 10,935.00 | 4,729.54 |
| Subtotal: 2nd Pass review fee     | 20,000.00 | 4,166.67 | 2,916.87 | 29,166.87 | 13,333.33 |
| Subtotal: Privilege review fee    | 2,177.28 | 367.42   | 257.60 | 2,624.40 | 1,418.86 |
| Subtotal: Attorney Redaction fee  | 65.32  | 11.02    | 7.73  | 78.73  | 42.57  |
| SUBTOTAL REVIEW                   | 29,500.20 | 6,076.01 | 4,255.31 | 42,804.80 | 19,524.30 |
| Subtotal: Production per/GB fee   | 83.89  | 21.23    | 7.44  | 151.67 | 41.00  |
| Subtotal: Production hourly rate  | 150.00 | 150.00   | 150.00 | 150.00 | 150.00 |
| SUBTOTAL PRODUCTION COST          | 233.89 | 171.23   | 157.44 | 301.67 | 191.00 |
| Subtotal: Tech support            | -      | -       | -     | -     | -      |
| Subtotal: Project Management support | -     | -       | -     | -     | -      |
| Subtotal: Outside Counsel         | -      | -       | -     | -     | -      |
| Subtotal: Supervisor/Manager fee  | -      | -       | -     | -     | -      |
| SUBTOTAL MANAGEMENT/ADMIN          | 1,295.53 | 1,295.53 | 1,295.53 | 1,295.53 | 1,295.53 |
| TOTAL:                            | $ 33,493.48 | $ 10,515.21 | $ 8,474.67 | $ 47,963.65 | $ 23,137.85 |

The Cost Prediction Calculator is designed to help lawyers estimate the costs associated with document productions, and at the same time provide a template that lawyers and judges can use to facilitate discussion and resolution of disputes over the burdens associated with discovery of ESI. The numbers and assumptions in the calculator are based on current publicly available pricing and the collective knowledge and experience of the drafting team. The appropriate numbers and assumptions in any individual case may vary greatly and may need to be adjusted. The numbers and assumptions are also time sensitive and will need to be modified as technology and costs change.

The calculator provides flexibility to compute costs for an entire project or a single stage, and for one data source or multiple data sources.

The calculator contains eight sections:

1. Instructions: provides a step-by-step outline of using the calculator.
2. Judicial Dashboard: an abbreviated summary of the calculator results depicted by EDRM stage.
3. Judicial Detail View: an abbreviated view of the calculator results by data source type and EDRM stage.
4. Attorney Detail View: a more detailed dashboard view that summarizes the total cost for each EDRM stage.
5. Data Source Estimates: a spreadsheet that allows you to estimate data volumes for each data source type.

6. Costs: presents the costs estimates based on current publicly available pricing. You may enter your own pricing as appropriate.

7. Assumptions: data value assumptions pertaining to volume reductions, review rates, etc. You may enter your own assumptions as appropriate.

8. Calculator: using input from the 3 steps above, this section allows you to calculate the overall costs for each stage of EDRM for each different type of data source.

Each individual case may be different, and thus may warrant different pricing and assumptions than the ones included in the calculator. These numbers are intended only as a benchmark to gain insight into document production costs and their potential impact on the overall burden of discovery in civil cases.

Developed under the George Washington Law School James F. Humphreys Complex Center and is freely available to the public. The Center acknowledges Insight Optix for its original thinking that inspired the New Framework and its critical logistical support.
### Appendix G – Variances that Require Adjustment to Model Set of Projected Costs

<table>
<thead>
<tr>
<th>Complexity Factor</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud based (ex. YouTube, Wiki's), Filesharing (ex. Dropbox, Google Drive) Social networking (ex. WhatsApp, Facebook) or collaborative applications (ex. Slack, Yammer, Teams)</td>
<td>To the extent the data locations are external, cloud based, or not otherwise under full control of party, there can be numerous complications. Specialized attorneys, consultants, and technical experts may be retained to complete the work.</td>
</tr>
<tr>
<td>Mobile devices including phones, tablets, external drives</td>
<td>More involved custodian interviews may be necessary to pinpoint precise applications on specific devices. Specialized software and technicians necessary to retrieve data from these locations</td>
</tr>
<tr>
<td>Large volumes of hard copy data</td>
<td>Determining locations of relevant data at external storage locations and then unitizing, scanning, OCR is labor intensive, especially if there are hundreds of boxes of hard copy materials</td>
</tr>
<tr>
<td>Investigations, or matters with investigative elements</td>
<td>Forensic images to be created and complex forensic analysis performed on various laptops, or other devices make this a more labor intensive, time consuming analytical activity.</td>
</tr>
<tr>
<td>Unusual Source data:</td>
<td>Non-standard, non-business data types such as those listed here may increase the processing costs, the attorney review rate or require a specialized resource to interpret the information.</td>
</tr>
<tr>
<td>- Audio/video</td>
<td></td>
</tr>
<tr>
<td>- complex data types</td>
<td></td>
</tr>
<tr>
<td>(see Structured Systems)</td>
<td></td>
</tr>
<tr>
<td>- non-standard files (engineering drawings, medical imaging)</td>
<td></td>
</tr>
<tr>
<td>- legacy applications or repositories</td>
<td></td>
</tr>
<tr>
<td>Targeted v. full collections</td>
<td>A full collection may increase the processing costs due to the overall size of the data set collected.</td>
</tr>
<tr>
<td>Third-party email collection</td>
<td>Email from private custodian accounts or non-standard email types may present a greater challenge to collection procedures.</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>Foreign language documents may require special processing procedures as well as translation or translation services.</td>
</tr>
<tr>
<td>High risk discovery factors present</td>
<td>May include risk of spoliation of the data, an uncooperative client or involvement of bad actors.</td>
</tr>
<tr>
<td>Unusually contentious litigation</td>
<td>Lack of cooperation or consistent roadblocks or objections to methodologies proposed.</td>
</tr>
<tr>
<td>Cross border discovery</td>
<td>The logistics of collection and compliance with local regulations.</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Privacy</td>
<td>Cost of recovering when business and personal data is co-mingled on mobile devices; when discovery is impacted by data privacy law (e.g., HIPPA, EU GDPR, CCPA)</td>
</tr>
<tr>
<td>Structured Systems</td>
<td>Special expertise and skills may be needed to gather, compile, and render ESI for structured data sources. Review costs may be mitigated because that attorneys are reviewing data compilations and summary reports. Although higher cost, non-attorney reviewers are often needed (e.g., accountants, engineers, nurses).</td>
</tr>
</tbody>
</table>
APPENDIX H – Standard-Unit Cost Table

GW Discovery Proportionality Model
COSTS FOR TYPICAL DATA SOURCE VOLUMES AND UNIT COSTS

<table>
<thead>
<tr>
<th>ESI STAGE</th>
<th>Hosted Email</th>
<th>File Shares</th>
<th>Social Media/Website</th>
<th>Computer/ Laptop</th>
<th>Mobile Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption: Typical GB per Custodian</td>
<td>5</td>
<td>4.5</td>
<td>0.25</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Collection</td>
<td>$ 750.00</td>
<td>$ 1,150.00</td>
<td>$ 500.00</td>
<td>$ 600.00</td>
<td>$ 750.00</td>
</tr>
<tr>
<td>Processing &amp; Hosting</td>
<td>$ 1,713.87</td>
<td>$ 1,822.45</td>
<td>$ 2,266.40</td>
<td>$ 2,961.65</td>
<td>$ 1,377.03</td>
</tr>
<tr>
<td>Review, Production &amp; Management/Support</td>
<td>$ 31,029.61</td>
<td>$ 7,542.77</td>
<td>$ 5,706.28</td>
<td>$ 44,402.00</td>
<td>$ 21,010.82</td>
</tr>
<tr>
<td>Estimated Cost applying typical GB per Custodian:</td>
<td>$ 33,493.48</td>
<td>$ 10,515.22</td>
<td>$ 8,474.68</td>
<td>$ 47,963.65</td>
<td>$ 23,137.86</td>
</tr>
<tr>
<td>Estimated Unit Cost per one Gigabyte:</td>
<td>$ 6,698.70</td>
<td>$ 2,336.72</td>
<td>$ 33,898.72</td>
<td>$ 1,598.79</td>
<td>$ 2,892.23</td>
</tr>
</tbody>
</table>

This chart reflects the costs from collection through production for each type of data source within a representative ESI matter. The totals are based on an assumption of one (1) custodian and typical amount of data for each of the itemized data sources. We understand each project is different and would reflect individual circumstances, assumptions and costs. For example, not every data source will typically be present in every case. The table is offered solely as a baseline to aid in estimating the per gigabyte costs for your specific matter.

Assumptions
- Collection calculation for each data source = 1 device
- Processing, hosting, review & production calculations = 1 custodian
- 24 month duration hosting
- 1 user license
- 10 hours project management
- 5 hours tech support

Developed under the George Washington Law School James F. Humphreys Complex Center and is freely available to the public. The Center acknowledges Insight Optix for its original thinking that inspired the New Framework and its critical logistical support.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email threading</td>
<td>Automated process that organizes the emails in a email conversation so that they may be reviewed together in context.</td>
</tr>
<tr>
<td>Near Duping</td>
<td>Automated process of identifying documents that are nearly identical in content for the purposes of expedited review</td>
</tr>
<tr>
<td>Culling</td>
<td>Process of reducing the overall size of the data set. May consist of targeted keyword searches, date range filtering, deNist &amp; deduping.</td>
</tr>
<tr>
<td>Culling: deNist</td>
<td>Data processing step of the removal of file types unlikely to have evidentiary value such as system files, executables, etc.</td>
</tr>
<tr>
<td>Culling: Dedupe</td>
<td>Data processing step of identifying and removing identical duplicates from the dataset</td>
</tr>
<tr>
<td>Custodians</td>
<td>Persons who have custody or control of data subject to preservation.</td>
</tr>
<tr>
<td>Data Expansion</td>
<td>The processing step whereby data volume is expanded due to the opening and extraction of zip files, PST files, etc.</td>
</tr>
<tr>
<td>Data Source: COLLABORATION/MESSAGING</td>
<td>Applications such as MS Teams, Slack, Salesforce, etc.</td>
</tr>
<tr>
<td>Data Source: COMPUTERS/LAPTOP</td>
<td>Company issued or BYOD hardware.</td>
</tr>
<tr>
<td>Data Source: FILE SHARE</td>
<td>Shared network storage systems, cloud storage, etc.</td>
</tr>
<tr>
<td>Data Source: HOSTED EMAIL</td>
<td>A centralized server or a cloud-based email system.</td>
</tr>
<tr>
<td>Data Source: MOBILE DEVICE</td>
<td>Includes company issued or BYOD mobile devices such as phones and tablets.</td>
</tr>
<tr>
<td>Data Source: PAPER</td>
<td>Physical documents and the OCR/scanning process associated with converting the documents to a machine-readable format.</td>
</tr>
<tr>
<td>Data Source: SOCIAL MEDIA</td>
<td>Facebook, Twitter, Instagram and other Social Media or web-based platforms.</td>
</tr>
<tr>
<td>Data Source: STRUCTURED SYSTEMS</td>
<td>Database applications such as Human Resource, Accounting, Finance systems.</td>
</tr>
<tr>
<td>ESI Stage: ANALYTIC WORKFLOW OPTIONS</td>
<td>May include predictive coding methodologies such as TAR/CAL, data visualization or topic classification.</td>
</tr>
<tr>
<td>ESI Stage: COLLECTION</td>
<td>Process of identifying and physical collecting data from targeted data sources.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ESI Stage: HOSTING</td>
<td>Storage of collected data in readily accessible platform for document review, includes user fees or licenses.</td>
</tr>
<tr>
<td>ESI Stage: PRE-PROCESSING CULLING</td>
<td>Targeted collection strategies such as keyword, date filtering or data source selection.</td>
</tr>
<tr>
<td>ESI Stage: PROCESSING</td>
<td>Multi-step function whereby data is extracted, normalized and prepared for document review and production.</td>
</tr>
<tr>
<td>ESI Stage: PRODUCTION</td>
<td>Process of turning the results of the attorney review into something that can be produced to the requesting party</td>
</tr>
</tbody>
</table>

Forensics

Utilized to collect all data on a device, including system and deleted files without altering metadata. Computer forensics is not routinely employed for civil litigation matters unless warranted in special circumstances.

GB: document ratio

Estimated number of documents contained in a GB of data.

Ingestion

Initial phase of loading data for review. Many vendors charge a reduced rate for ingestion and allow the application of culling and keyword processes to the dataset prior to full processing.

Non-custodian data sources

Data sources not associated with a specific custodian such as central repositories, databases or shared cloud storage.

OCR

"Optical Character Recognition" OCR technology recognizes text within a digital image and converts it to a searchable format.

Redaction

Process of removing sensitive or confidential information from documents being produced.

Scanning

Process to convert physical paper and photos to a digital format.

TAR/CAL

Predictive coding techniques used to automate the identification of potentially relevant documents and prioritize the documents for review and/or production.
Organizational Structure - Not only is it critical to have a thorough understanding of the technical infrastructure, but also required is an understanding of the departments, teams, and individuals who may possess knowledge or understanding of the subject matter and possible repositories in question. This understanding can be gained through surveys or interviews within the various groups, including the custodians and non-custodian data stewards, compliance or privacy personnel, department heads, HR, and internal ediscovery or GC office personnel.

1 Data Mapping - An identification plan to capture the data sources and types of information required should start with a discussion with the IT team to assess the overall technical infrastructure and network architecture utilized within the organization. This would include an outline of the internal networks, cloud storage, archives and enterprise software systems. IT should also provide a data map to define who has access to what systems, what timeframes each system has been in use, and what data stores are likely to be found on them. They should provide a comprehensive list of custodians and their personal data sources, whether company owned or BYOD (Bring Your Own Device), including mobile devices, computers or laptops.

2 A “Data Steward” is an individual who maintains information but is not themselves a fact witness or source of potentially relevant information.

3 Included in the collection workflow is the process of identifying which witnesses may possess information relevant to the litigation (aka “custodians”), where that ESI is stored, how it may be collected, and who to contact regarding the preservation and collection of the ESI. This identification of appropriate data sources may occur concurrently with collection, as the technical resources (e.g., those individuals most familiar with a party’s use of ESI) are often suited to both identify and collect said data. These individuals often include IT, computer forensics, ediscovery professionals, or members of the legal team with a strong understanding of information technology, working closely with custodians and under the supervision of counsel.

4 Details about likely relevant data should be extracted from the initial fact gathering steps such as custodian interviews, which will facilitate processing the information at a later stage. Understanding who may have created and used what data types, in what time frame, storing them in what location are crucial details that can inform decisions on which filtering options will be used during the collection stage.

5 For example, if the facts crucial to an early phase of ediscovery are more likely to be found in a series of reports typically created and used by several custodians using specific document formats in a six-month period, filtering should be applied initially to locate that narrow, defined set.

If at any point more data is needed from additional sources or custodians, the methods mentioned before can be easily adjusted to include, for example, a broader date range. It can also be staggered in an iterative fashion as the case develops. This approach avoids premature activity and also helps to moderate and balance use of resources in a proportional manner.

5 “Having some logical connection with the consequential facts…Of such a nature that knowledge of the item would affect a person’s decision-making process…significant; essential….”

6 The ultimate fact may be one that, if accepted, causes the claim or defense to prevail, whereas the intermediate fact is combined with other facts to make that determination.

7 For instance, determining materiality in a trade secret case may involve information tending to show the custodian's access to trade secrets, information tending to show the devices the custodians used to access the trade secrets and whether the trade secrets were taken, information tending to show the steps taken to protect the trade secrets, and information tending to show that trade secrets were generally known.

8 Of course, ESI frequently creates exact duplicates, but for our purposes those duplicates are eliminated by deduplication and other methods and are not further considered.

9 In some cases, the decision-maker may be the lawyer representing the producing party. In other cases, the lawyer may get input from the opposing lawyer.

10 For example, a manager at the time of the fact giving rise to an action, who later becomes Vice President and has further involvement with additional facts of the matter, may be more material than other fact witnesses. With regard to
data sources, if the manager utilized a different laptop as a manager than as a Vice President, one laptop could be more material than the other.

11 Art. 6 GDPR
12 Art. 13 GDPR
13 Art. 49 GDPR
14 The model presumes that the responding party develops the table of costs at the outset of litigation before actual costs have been incurred. Discussion of whether to ask for opposing side’s input at various stages of litigation will be discussed in section 4.
15 https://edrm.net/edrm-model/.
17 NOTE: Our Framework is designed to forecast from Collection through Production. While a key step in the process, Preservation costs vary so widely based on individual circumstances, the costs associated with it are not contemplated in the Cost Projection Calculator.
18 The objective is to prepare the data for document review through accurate and consistent culling and minimizing the quantity of files by excluding files with little or no evidentiary value (e.g., non-business files, duplicate files, files that are not germane to the claims or defenses of a case).
19 In addition to the commonly accepted practices of the use of keywords, there are tools that permit further reduction at the review phase. One of these is email threading, which groups together different fragments of longer email conversations. This permits attorneys to review only the most complete versions of those conversations (sometimes called the “inclusive”) and avoid having to review the fragments (sometimes called “non-inclusive”).

Another set of tools, which some refer to generally as technology assisted review, or TAR, provide additional opportunities to reduce the number of documents that need human review. Most versions of TAR combine machine learning and human review by a subject matter expert that helps producing parties prioritize documents for review. Using statistical validation protocols and other methods, TAR also can be used to demonstrate when additional review would be unlikely to bring back additional responsive documents.
20 The Sedona Conference has defined TAR as a “process for prioritizing or coding a collection of electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection.” The Sedona Conference, The Sedona Conference Glossary: E-Discovery and Digital Information Management, Fourth Edition, 15 SEDONA CONF. J. 305 (2014) (definition adopted from Maura R. Grossman & Gordon V. Cormack, The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola, U.S. Magistrate Judge, 7 FED. CTS. L. REV. 1, 32 (2013)). The terms “predictive coding” and “computer assisted review” are often used interchangeably with TAR, to describe this process. TAR can also involve what is called “continuous active learning,” or CAL.
21 Moreover, the cost of privilege review and privilege logs may be reduced as parties take advantage of Rule 502 of the Federal Rules of Evidence and similar rules protecting parties against waiver.
23 ©Insight Optix LLC.
24 ©Insight Optix LLC.
25 The Higher Priority Custodians / Lower Burden classification in the upper left quadrant pinpoints the priority group to review and assess. Isolating this group early and focusing on disclosure and production of the custodians and data sources in this quadrant provides a win-win for both parties because the approach results in early and expeditious
calibration of the highest priority custodians and their data. Further data analytics can also be performed within each quadrant to inform future negotiations.

26 The parties may use the New Framework’s heat map and heat map table to plan processing, analysis, and review workflows. They may decide that the case and the data is suited to keyword searches. The assessment may also impact processing decisions, including whether to use analytics tools like email threading, TAR, or continuous active learning, all of which require upfront costs with the expectation of cost savings during review.

27 It informs decisions by the producing parties on whether and how to preserve and collect data from a particular source. The party may decide to preserve the data through collection or using preserve-in-place technology. This may turn on the type of data source. For example, whether to collect a logical or physical image of a computer, or whether to collect mobile devices or home computers.

28 During discovery and with the exchange of interrogatories, document requests, deposition notices and third-party discovery, a clearer picture forms as to the theories and themes that the parties develop to support their claims and defenses. The parties may update and fine-tune the proportional assessments, adding new custodians and sources and re-evaluating the importance of each custodian and source’s data. They may re-evaluate whether preservation and collection of each data source identified in the process is proportionate.

29 The upper right quadrant of the New Framework’s heat map includes data sources that are from Higher Priority Custodians / Higher Burden class. Certain data sources in this quadrant may be moved forward, but the evidentiary value should be weighed, and negotiation of potential sampling or cost shifting to address the highly burdensome nature of the discovery.

The lower left quadrant represents the Lower Priority Custodian / Lower Burden group. Data sources should be assessed for relative evidentiary value due to less likelihood of finding relevant information. This evaluation can be reinforced through data sampling to ensure that relevant content does not exist in the data associated with these custodians.

The final quadrant of Lower Priority Custodian / Higher Burden should also be assessed for relative value. Parties may determine that moving this data forward is disproportionate to the needs of the case and will not lead to the discovery of valuable information. Negotiation may lead to possible exclusion due to the burdensome nature.

30 Priority custodians with low burden data sources can be placed into a phased discovery plan, starting with high-yield sources before proceeding to those that are more burdensome and less likely to yield relevant information. With costs generated for each source, the 26(b) consideration of “whether the burden or expense of the proposed discovery outweighs its likely benefit” is much easier to quantify.

31 Maintaining data has cost, even if it is just the cost of storage. In cases where there is a large amount of electronically stored information, that cost can be more significant than initially expected. In these cases, a producing party can identify the storage costs of information related to released claims and evaluate (and document) whether to request permission to release holds for certain custodians or certain information or release other preservation obligations.

32 A Westlaw search of the term for the year 2020 yields 474 cases in which a party requested reopening discovery over the course of the year.) See, e.g., Stephen C. v. Bureau of Indian Educ., 2020 WL 4464398, *1 (D. Ariz. Aug. 4, 2020) (noting scope dispute). See, e.g., Campos-Elbeck v. C.R. Bard, Inc., 2020 835305,*3 (S.D. Cal. Feb. 20, 2020) (denying request as not proportional). In these cases, if the proportional assessment is being applied cooperatively, the requesting party can identify the needed information and its location on the heat map to argue that limited reopening of discovery is proportional to the needs of the case. In addition, the requesting party can use the Framework to identify and limit the scope of any reopening. If it is being applied in an adversarial situation, then the producing party can use the Framework to show that the additional information is not proportional or point out appropriate limits.

33 With costs generated for each source, the Rule 26(b) consideration of “whether the burden or expense of the proposed discovery outweighs its likely benefit” is much easier to quantify.

34 Priority custodians with low burden data sources can be placed into a phased discovery plan, starting with high-yield sources before proceeding to those that are more burdensome and less likely to yield relevant information.

35 The New Framework is used to help a judge understand where to draw the line in enforcing or quashing a request for production. A judge can easily balance the cost presented to the amount in controversy.
Rule 37(e) notes that if electronically stored information is not properly preserved, a judge may issue sanctions. By finding those custodians and data sources that are relevant to the matter and ranking them according to priority in the information they could provide, the Framework can be a guideline for determining whether spoliation has occurred, and sanctions might be deemed necessary.