DISCOVERY PROPORTIONALITY MODEL

A NEW FRAMEWORK

April 8, 2022

Rabiej Litigation Law Center

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

DISCOVERY PROPORTIONALITY MODEL

A NEW FRAMEWORK

Preface

The 2015 amendments to Rule 26(b)(1) redefined the scope of discovery and require the court and parties to consider four pertinent factors in assessing whether potentially relevant information is proportional to the needs of the case.[1] At the same time, Rule 26(g) continues to impose an obligation on lawyers to certify that every discovery request or response is "neither unreasonable nor unduly burdensome or expensive, considering the needs of the case … the amount in controversy, and the importance of the issues at stake in the action" after a reasonable inquiry.[2] Case law reveals that Rule 26(b)(1) is recognized in a growing number of cases, yet all the four pertinent Rule 26(b)(1) proportionality factors are not seriously considered, inconsistent with both Rule 26(b)(1) and Rule 26(g).

Parties do not have the luxury of time. They must make critical discovery decisions evaluating these proportionality factors, virtually from day one of a lawsuit, based on incomplete knowledge of discoverable information. The amended rule and committee note provide little guidance, even though good-faith, mistaken judgments can be harshly penalized.

With little guidance, the bench and bar grow frustrated and either neglect or ignore serious evaluation of all the Rule 26(b)(1) proportionality factors. The DISCOVERY PROPORTIONALITY MODEL: A NEW FRAMEWORK provides a standard methodology for the user to address these factors. The user's assessments are quantified and recorded on a discovery roadmap, including the user's estimates of discovery costs for every custodian with relevant information on various data sources, and assessments of burden, benefits, and costs. A bottom-line overall discovery cost is estimated in relation to the amount in controversy and importance of the issues at stake.

A team of more than 50 experts developed an excel spreadsheet with more than 100 lines of itemizations to estimate costs in discovering information from five separate data sources to assist the user. Using the calculator, the team also provided its own estimates of costs, burdens, and numbers of gigabytes, which they most often handle in typical cases based on average costs and numbers of gigabytes developed consistent with their collective experience and a literature review. Lastly, the team estimated a reasonable amount of discovery costs in relation to the amount in controversy in typical cases and adjustments for the importance of the issues at stake. The user has the options of disregarding the team's estimates, comparing them with their own estimates to evaluate reasonableness, or adopting them in whole or in part in appropriate circumstances.

Novel Approach

For the first time, costs and burdens to discover information from various data sources are separately addressed and identified, e.g., email, file shares, mobile devices, and social media. Significantly, objective metrics are proposed to address the relation between discovery costs and the amount in controversy and importance of the issues at stake, two key Rule 26(b)(1) proportionality factors that have received scant attention.

By-products of the NEW FRAMEWORK include powerful auxiliary tools, which can be used independently to estimate and compare discovery costs, assess the degree of burden in accessing data sources, and project unit costs for common data sources. Appendix F contains many of the auxiliary

tools, including a sophisticated calculator-prediction excel spreadsheet that itemizes inputs as well as a table that lists unit costs and typical numbers of gigabytes used to project average costs incurred. A preliminary draft of Appendix J contains a cost calculator based on a TAR 1.0 application.

The NEW FRAMEWORK is designed to better inform Rule 26 proportionality decisions made at all stages of litigation. It is not intended to be applied mechanically; it is not mathematically precise; it is iterative in nature; it is entirely transparent; and a judge or others can assess for themselves the inputs and reasonableness of the proportionality assessments under the modelling.

The NEW FRAMEWORK does not, and cannot, make the ultimate determination whether discovery is "proportional to the needs of the case." This is so because Rule 26(b)(1) does not assign a specific weight to any of the proportionality factors and thus any single factor may trump other factors depending on the facts in a given case. But the NEW FRAMEWORK isolates and quantifies individual assessments on a granular scale and better informs the analysis.

Drafting Process

The new analytical framework was developed by 50 practitioners, ediscovery experts, and judges. John Rabiej, former Director of Duke Law School's Center for Judicial Studies and founder of the Rabiej Litigation Law Center, launched the project in early 2020 in collaboration with the George Washington Law School's James F. Humphreys Complex Litigation Center. 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.

The NEW FRAMEWORK represents the authors' consensus views, but not necessarily their unanimous agreement with all content.

The NEW FRAMEWORK was inspired and derived from Insight Optix LLC's patented Evidence Optix® software.[3] The Center and the drafters readily acknowledge the important role Insight Optix played in the advancement of the NEW FRAMEWORK. The Center also acknowledges the studies published by the Federal Judicial Center (FJC), Tymetrix (Wolters Kluwer), and the Searle Center on Law, Regulations, and Economic Growth, Northwestern Law, which were instrumental in developing projections of discovery costs in relation to the amount in controversy as well as adjustments accounting for the importance of the issues at stake in the action.[4]

ACKNOWLEDGEMENTS

The Rabiej Litigation Law Center gratefully acknowledges the time and work of the more than 50 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
Ray, Niloy Littler Mendelson	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 Redgrave LLP

TEAM MEMBERS

Ackerman, Ethan
Morgan Lewis

Addington, John
Corning Law

Averitt, Jessica
Baker McKenzie

Baratz, Michael
Steptoe & Johnson

Bauer, Kate

Behan, Kevin
Yum

Bennitt, Jane
Global Legal Ebilling

Blaustein, Stacey
IBM

Carle, Jay
Seyfarth Shaw, LLP

Carvo, Scott
Warner Norcross & Judd

Coleman, Jennifer
Hopkins Carley ALC

Deere, Chioma
Williams, Leininger & Cosby

Diettle, Daniel
CDS

Dockterman, Mike
Steptoe & Johnson

Durocher, Skip
Dorsey & Whitney

Hamilton, Michael
Exterro

Hulliberger, Kristopher
Howard & Howard

Isaacs, Tom
Bowman & Brooke

Jackson, Matt
Sidley Austin

Judkins, Sonya
T-Mobile

Kairalla, Jason
Carlton Fields

Krezalek, Martin
Blank Rome

Lange, Michelle
VW

McGuire, Marina
Tanenbam Keale

Murao, Joy
Practice Aligned

Murray, Barb
PWC

O'Brien, Kate
Insight Optix LLC

O'Connor, Tom
Gulf Coast Legal

Pass, Robert
Carlton Fields

Patzakis, John
X1

Redgrave, Jonathan
Redgrave

Rizzolo, Dan
ePercept

Robinson, Rich
Toyota Legal One

Rollins, Graham
Capital One

Schouten, Nataliya
FIS Global

Scrutton, Alexandra Grand
3M

Shortnacy, Michael
King & Spalding

Six, Michelle
Kirkland & Ellis

Stadler, Anna Patricia
Robert Bosch GmbH

Staunton, Thomas
Miller Shakman Levin & Feldman

Sullivan, Janet
White & Case

Sweeney, Caroline
Dorsey & Whitney

Weiner, Paul
Littler Mendelson

Yates, Janice
Insight Optix LLC

DISCOVERY PROPORTIONALITY MODEL
A NEW FRAMEWORK

Introduction

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 party's 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, which underscores the fifth factor.

Two of the six factors are intended solely to be prophylactic measures. The "parties' relative access to the information" factor is intended to prevent the producing party in an asymmetric-discovery case from arguing that he was producing much more information than the requesting party, which arguably would not be proportional to what the requesting party produces.[5] As the Civil Rules Committee materials say: "If relative access to information is implicit in the present factors, what harm is there in making it explicit and reducing the opportunities for ill-founded contentiousness?"

The "parties' resources" factor is a similar prophylactic measure intended to address potential abuse by a well-financed party, not in determining specific proportionality assessments. The committee's intent was to address "the limitations on a financially weak litigant to withstand extensive opposition to a discovery program or to respond to a discovery request."[6] In other words, this factor is pertinent when a well-financed party objects to every request or requests excessive discovery, intending to exhaust the requesting party's resources. It certainly is not intended to provide one party more discovery solely because the other party has more resources. The remaining four Rule 26(b)(1) factors must be considered in every instance in a holistic manner to assess whether the discovery is proportional to the needs of the case.

Overview

The NEW FRAMEWORK follows the same steps that experienced ediscovery lawyers, consultants, and service providers routinely undertake to assess whether discovery is proportional to the needs of the case. Lawyers determine the scope of discovery, evaluate the amount in controversy, identify and prioritize individuals in accordance with the importance of information in resolving issues that they possess, ascertain the burdens/costs and benefits in retrieving and reviewing discoverable information, and conclude with an overall evaluation of whether the discovery of individuals or devices is proportional.

The NEW FRAMEWORK'S methodology starts by assessing and prioritizing the importance of the discoverable information that custodians possess or control. It also assesses the degree of burden in accessing and recovering that information from their individual sources (devices). The cost of discovery is not included in the assessment of the burden, which is considered later in the analysis.

As an aid to assist the user in assessing and prioritizing custodians, a "heat map" illustrates preliminary assessments, which can be used to identify obvious targets of discovery, facilitate party discovery negotiations, and produce leads to further refine discovery. A sample heat map is discussed in Section 02: Data-Source Burden and Effort.

The NEW FRAMEWORK builds on the preliminary assessments in the heat map and addresses the other proportionality factors. All these assessments are reported in a discovery roadmap, which records the user's estimates of the costs, burdens, and benefits of discovery of information of every custodian on their various devices. A bottom-line overall discovery cost is estimated in relation to the amount in controversy that may include an adjustment, in appropriate circumstances, for the importance of the issues at stake in compliance with Rule 26(b)(1).[7]

A blank discovery roadmap is set out below. A filled sample discovery roadmap applying the New Framework team of experts' estimates and projections to a hypothetical $20 million case is discussed in Section 03: Discovery Cost Projections.

A New Framework
Discovery Roadmap

Importance of Information Held by Custodian (Least → Most)	← Incremental Burden Scale (Low to Highest) →					(Insert column per addl. data source category)	Cumulative Discovery Cost / GB
	I	II	III	IV	V		
4							
3							
2							
1							
	Email	File Shares	Social Media	Computer/Laptop	Mobile Device	Other Sources	
Average cost per custodian	$33,000	$10,500	$8,474	$47,963	$23,140		
Average per unit GB	5 GB	4.5 GB	.25 GB	30 GB	8 GB		
Average cost per GB	$6,600	$2,333	$33,896	$1,599	$2,892		

Estimate of reasonable overall discovery cost in relation to the amount in controversy $_____.
Adjustment to account for the importance of issues at stake in the action $_____.
Projected reasonable overall discovery cost $_____.

III. Conclusions

The NEW FRAMEWORK provides a standard methodology to assess the four pertinent proportionality factors and records those assessments in a matrix format that better informs the analysis not only for the benefit of the user but also for the other side and the judge, if shared by the user. It provides not only guidance on how to estimate costs and evaluate assessments, but also provides estimates of

average costs and number of gigabytes and assessments by a team of experts based on collective experience in typical cases that may be relied on by the user, the other side, and the judge.

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. Most cases do not generate serious discovery disputes. But proportionality assessments are required in every case because they are part of the definition of discoverable matter under Rule 26(b)(1).

Sections 01-03 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 04 provides guidance to assess the "amount in controversy" and "importance of issues at stake." Section 05 describes the application and strategic uses of the NEW FRAMEWORK and the generation of a record, which will inform party negotiations and judicial resolution of discovery disputes, if required.

Section 01: Custodians

I. Introduction

Section 01 provides guidance on how to group custodians in four broad categories, from highest to lowest priority. Section 02 gives instruction on how to identify the five most common data sources and rank them by degree of burden in accessing information.

Custodians are prioritized by the "importance of the (discovery) [information] they possess or control in resolving the issues," a proportionality factor under Rule 26(b)(1), which also overlaps the "benefits" of discovery, another proportionality factor.

A. 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 places the NEW FRAMEWORK in 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.

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.[8]

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 – individuals who maintain information but are not themselves fact witnesses or sources of potentially relevant information). Interviews, written requests, or data sampling are techniques often used to further the investigation.[9]

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.[10]

B. Prioritizing Discoverable Information

After the general scope of relevant matter is ascertained and the universe of custodians identified who potentially possess or control the discoverable information, the information should be classified by its level of priority, from low to high priority. The NEW FRAMEWORK proposes four criteria to prioritize discoverable information, which is important to resolving the issues, including: (1) materiality of the information; (2) strength of the information; (3) uniqueness of the information; and (4) likelihood of finding significant discoverable information on a particular data source.

1. Materiality of Information

Some discoverable information may be more useful, significant, or important than other information. The New Framework's first criterion focuses on "materiality," which is defined as significant or essential, to discern qualitative differences in relevant information.[11] 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;[12] and,
- is an essential link in a line of evidence needed to prove an assertion?[13]

2. 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.

3. 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,[14] 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.[15]

4. Likelihood of Finding Significant Discoverable Information

Although an individual custodian may possess relevant information that is important to resolving an issue, it may not be located on every data source. The New Framework's fourth criterion focuses on the likelihood that the discoverable information will be located on particular data sources. The nature of the data source often is a good indicator of the type of information that likely will be found. For example, certain discoverable information may be found more likely on one data source, such as a shared file in text as compared to social media or a mobile device. In other cases, a custodian may more likely possess discoverable information on one data source and not on another data source. The location of a particular custodian's information on different data sources is unique to that custodian. Special care should be taken to investigate which data sources of a custodian may be significant, especially with custodians whose information is important.

II. 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.[16] 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 (*see* Section 02: Data-Source Burden and Effort) 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 four criteria of materiality, strength, uniqueness, and likely to be found on a specific data source to assess the priority of the relevant information along with the custodian's connection to the relevant 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.

> **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.
>
> **Non-custodian Data:** Data is non-custodian and a traditional custodian interview is not feasible. Appendix C illustrates the adaptations to the New Framework.
>
> **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: Data-Source Burden and Effort

I. Introduction

Section 02 provides guidance on how to rank the most common data sources in four broad categories, based on the degree of burden incurred in accessing information. Five primary variables affect the degree of burden. The specific degree of burden for each data source is not predefined; rather, it is dependent on circumstances.

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:

- 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 and review

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, every data-processing activity needs a legal basis,[17] and the privacy rights of individuals must be safeguarded.[18]

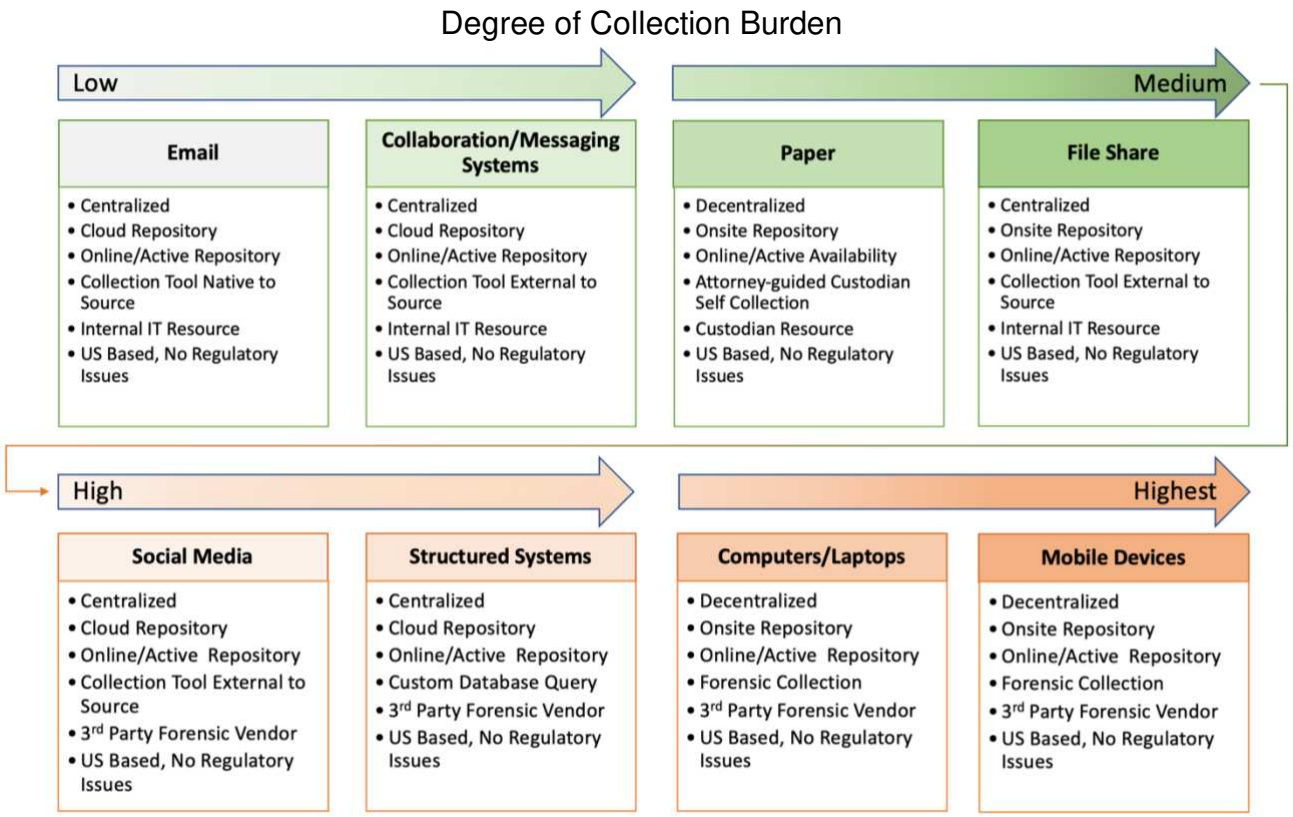
In addition, transfer of the personal data for use in litigation discovery in the U.S. is only possible if certain conditions are met.[19] Moreover, there are labor law regulations, blocking statutes (foreign and domestic), and state secrets law in other nations that should be considered. Failure to 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 user ranks the applicable data sources of discoverable information from least to most burdensome. Recognizing the possibilities that the five variables may be different for a particular data source in an individual case and require adjustments in individual cases, 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.

Unlike other estimates and projections that can vary significantly from cases to case, however, the relative burden rankings of data sources are consistent in most cases. The New Framework's team of 50 experts projected rankings of data-source burdens that can be reasonably expected in many, perhaps most, instances. The experts' rankings are based on a literature review and their collective judgment and experiences, applying the five variables as they most commonly appear in a typical case. A graphic illustrating the team of expert's rankings of eight data sources is set out below.

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. 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.

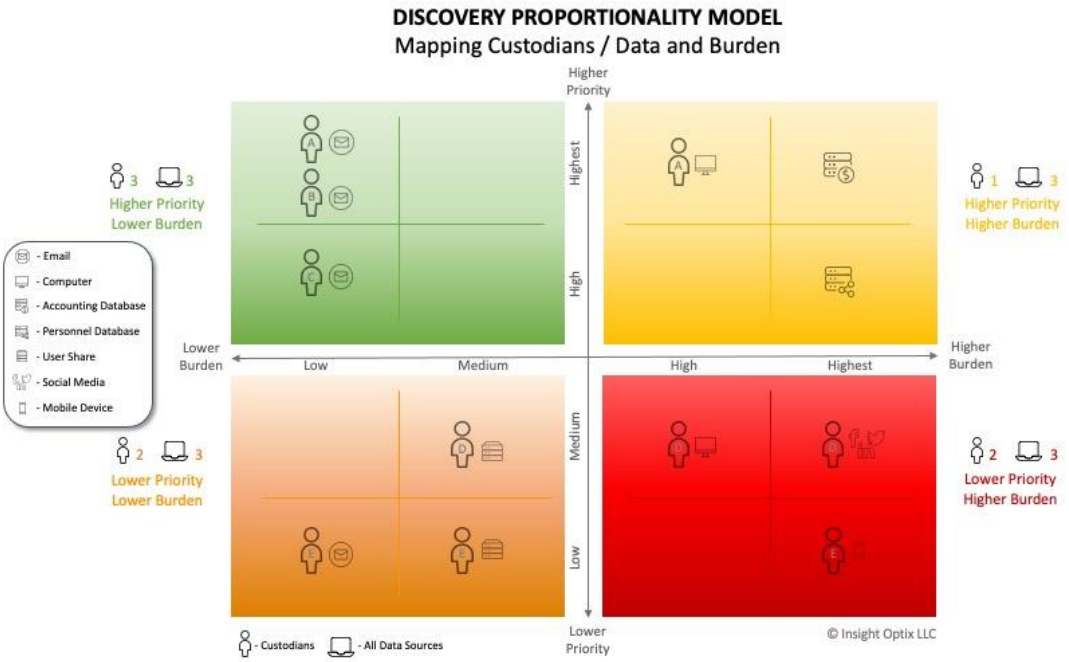


If the user adopts the team's burden rankings, the user should be aware of individual circumstances, which might require adjustments to the team's rankings. The clearest example are emails, which typically are located on a centralized server, providing relatively easy access and modest burden, compared with the less common use of decentralized email, which is located on individual computers, requiring multiple extractions and significantly increasing the burden.

V. Heat Map

A heat map is a tool used to begin preliminary assessments of discovery of custodians, which are shown in a matrix format, containing four general categories recording the prioritization of custodians and four general categories assessing burdens, ranging from low, medium, high, and highest importance and burdensomeness, respectively. The results are shown in a graphic heat map, which visually distinguishes custodians with significantly useful information from those with marginally useful information, along with the respective attendant burdens.

The user plots the custodians on the heatmap in accordance with the importance of information and burden. The closer that a custodian is plotted on the map's upper-left hand corner, the greater the importance of discoverable information in resolving the issues and the lower the burden. Conversely, the closer that a custodian is plotted on the lower right-hand corner the less important the discoverable information and the greater the burden. A sample heat map is set out below.



The prioritizing and mapping of the preliminary assessments of custodians and their data sources in the heat map addresses two of the four pertinent Rule 26(b)(1) factors, i.e., "burden of discovery," and the "importance of discovery in resolving the issues," which overlaps the "benefits" factor. As discussed in Section 04, the information collected from the heat map is added to the discovery roadmap along with costs of discovery from individual custodians and their devices, and assessments of costs in relation to the amount in controversy as adjusted by the importance of the issues at stake, the other Rule 26(b)(1) proportionality factors.

Section 03: Discovery-Cost Projections

I. Introduction

Section 03 provides guidance on how to estimate costs of discovery from custodians from various devices and data sources.

II. Estimating Discovery Costs

The New Framework's team of experts developed a cost calculator in an excel spreadsheet, which itemizes more than 100 separate elements of cost as an aid to assist the user estimate the discovery costs. The calculator includes 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." Click here to open the full calculator.

The calculator organizes costs around the discovery workflow found in the EDRM, a generally accepted model for ediscovery practice, including: Identification; Preservation; Collection; Processing; Review; Analysis; and Production. Like the approach taken by RAND in its seminal 2012

discovery cost report, *Where the Money Goes,* the cost calculator collapses costs into three main categories: A) Collection; B) Processing and Hosting; and C) Review and Production.[20] These calculations do not include *all* costs incurred in ediscovery, which can include substantial sunk equipment and software costs, incidental overhead costs, and employee compensation.

The costs involved in the EDRM steps, which are addressed in line-by-line itemizations in the cost calculator, are described below.

A. Collection Costs

Section 02 described generally the collection process of gathering the subset of potentially discoverable ESI. Preservation costs are not included in this category.[21] 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.

B. Processing and Hosting Costs

The bulk of discovery spending 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. 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). 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.[22]

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)[23] 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.[24]

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;[25] (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.[26]

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. NEW FRAMEWORK Cost Calculator

A user may rely exclusively on their own discovery-cost estimates. Alternatively, the user can rely on the New Framework's cost calculator or any other calculator tool to estimate their discovery costs.

The New Framework's cost calculator is set out as an excel spreadsheet. It contains the estimates of the New Framework's team of experts, but it permits a user to substitute their own numbers to fit particular circumstances by modifying the projected costs, rates, percentages, and other data.

Appendix F contains three of the Cost Prediction Calculator's spreadsheet tabs, which summarize the cost projections, with easily printable charts. As its name implies, the itemized cost prediction calculators list the individual costs incurred in estimating discovery for five specific data sources. They can be used to project the estimated costs in discovery. [Separate cost-calculators are being developed for TAR and CAL, which retain many of the same line-item collection and processing costs but substitute costs for analytics. A preliminary draft of Appendix J contains a cost calculator based on a TAR 1.0 application.]

The average estimates of the New Framework's team of experts in the excel spreadsheet rely on many assumptions, which are expressly disclosed. Some of the assumptions may not apply in an individual case, and the itemized costs in the excel spreadsheet may need to be adjusted. For example, the average estimates assume that a gigabyte consists of a certain number of pages. Although a gigabyte is the standard industry metric, there are no uniform metrics establishing the number of pages and documents in a gigabyte. Nonetheless, it must be converted into "pages" and "documents" to be

useful, because attorneys are compensated by the number of pages or documents they review, by far the single greatest element of discovery cost.[27]

A sampling of circumstances, which may reveal added complexities that require adjustments to the inputs to the cost calculator, can be found in Appendix G.

The NEW FRAMEWORK does not account for discovery costs associated with structured databases, e.g., personnel database. Presently, there is no way to project typical costs because the variances are too wide. Accordingly, the user must add the discovery costs of structured databases manually to the NEW FRAMEWORK. The same applies to atypical and new data sources.

IV. Average Per Gigabyte Costs of Discovery from Five Common Data Sources Projected by New Framework's Team of Experts

The New Framework's team of experts developed its own estimates of average costs and number of gigabytes from five common data sources, which they most often handle in typical cases based on their collective experience and supported by a literature review. The costs are expressed in standard units, *e.g.*, $ per gigabyte for typical volumes of data from five data sources. A user has the options of relying on their own estimates and discounting the team's estimates, using the team's estimates to compare and evaluate the reasonableness of their own cost estimates, or adopting them in appropriate circumstances in whole or in part.

The projections below were estimated using the cost calculator informed by a literature review, standard industry-wide assumptions, and the collective judgment and experiences of the New Framework's experts.[28] A graphic table containing the model set of projected costs is set out below and in Appendix H.

Discovery Proportionality Model
Judicial Detail View

STAGE	Hosted email (5GB)	File Share (4.5GB)	Social Media/ Website (0.25 GB)	Computer/ Laptop (30 GB)	Mobile Device (8 GB)	STAGE TOTAL
GB Collected	5.0	4.5	0.25	30	8	47.75
GB Produced	0.44	0.11	0.04	0.79	0.21	1.59
Collection	750.00	1,150.00	500.00	600.00	750.00	3,750.00
Pre-Processing	450.00	1,200.00	1,800.00	600.00	600.00	4,650.00
Processing	436.54	108.31	21.19	1,195.49	164.08	1,925.61
Hosting	827.33	514.14	445.21	1,166.16	612.95	3,565.79
Review	29,500.20	6,076.01	4,255.31	42,804.80	19,524.30	102,160.61
Production	233.89	171.23	157.44	301.67	191.00	1,055.23
Management/Support	1,295.53	1,295.53	1,295.53	1,295.53	1,295.53	6,477.64
DATA SOURCE TOTAL:	$ 33,493.48	$ 10,515.21	$ 8,474.67	$ 47,963.65	$ 23,137.85	$ 123,584.87

Per gigabyte costs of discovery can easily be computed from the table above, [29] and for the first time costs from five common data sources can be separately identified and compared, including:

- email--$6,698 GB
- file share--$2,336 GB
- social media/website--$33,900 GB
- computer/laptop--$1,599 GB
- mobile device--$2,892 GB

Section 04: NEW FRAMEWORK Discovery Roadmap

I. Introduction

The purpose of the New Framework's discovery roadmap is to record the user's estimates and assessments of the four pertinent Rule 26(b)(1) proportionality factors, which were identified in Sections 01-03, in order to better inform the analysis.

Section 01 addressed prioritizing custodians by the importance of the information they possessed or controlled, a good indicator of the "importance of the discovery in resolving the issues," one of the proportionality factors and one which overlaps with "benefits," a component of the expense/burden and benefit proportionality factor. Section 02 addressed the burden of discovery from individual data sources, a component of the burden/expense proportionality factor, by ranking five common data sources. Section 03 addressed the cost of discovery another proportionality factor.

Rule 26(b)(1) and Rule 26(g) require consideration of two more proportionality factors, including the amount in controversy and the importance of the issues at stake in the action. [30] These estimates and assessments are necessary to place costs of discovery from individual custodians in context by determining a reasonable overall discovery cost in relation to the amount in controversy, which can be adjusted to account for the importance of the issues at stake. [31]

To assist the user make these two remaining assessments, Section 04 provides estimates derived from respected studies addressing this matter. Again, a user has the options of discounting the New Framework's team of experts' estimates and substituting their own estimates, comparing and evaluating the team of experts' estimates to assess the reasonableness of their own assessments, or adopting them in whole or in part in appropriate circumstances.

II. Amount in Controversy

Courts have rarely interpreted the amount in controversy in depth when analyzing Rule 26(b)(1). But when they have "[m]any courts (including this one) appear to base the amount in controversy on some estimate of the defendant's worst-case scenario—that is, on the defendant's exposure if everything were to be decided against it."[32] The amount in controversy is usually the amount the plaintiff claims or could claim in good faith.[33] The New Framework's team of experts used this standard when making its assessments.[34]

A. Overall Discovery Costs in Relation to the Amount in Controversy

Whether the overall amount of discovery costs is proportional to the needs of the case can be ascertained under Rule 26(b)(1) only in the context of the amount in controversy and importance of the issues at stake. "Under the second proportionality factor, courts should 'compare[] the cost of discovery to the amount in controversy to determine [the proposed discovery's] proportionality."[35]

The NEW FRAMEWORK'S team of experts based their estimates on an analysis of several studies. In the 2010 *Litigation Costs in Civil Cases: Multivariate Analysis*, the Federal Judicial Center found that a 1 percent increase in the amount in controversy was associated with a 0.25 percent increase in total litigation costs.[36] For example, if the amount in controversy increased in a case from $100,000 to $200,000, the total litigation costs would rise by $25,000. Discovery costs can be derived from total litigation costs.[37] A 2016-17 study of attorney invoices, Tymetrix (now a subsidiary of Wolters Kluwer) found that discovery costs, excluding vendor costs, consistently represented between 29% and 33% of total litigation costs as measured by the ABA's L300 litigation-code sets.

B. Overall Discovery Costs in Relation to Importance of Issues at Stake

The importance of the issues at stake must also be considered under Rule 26(b)(1). In the 2010 *Multivariate Analysis*, the FJC found that the "importance of nonmonetary stakes to the client increased defendant costs [total litigation costs] by about 25%."[38]

The NEW FRAMEWORK team of experts relied on three criteria when calibrating the impact of the importance of issues at stake, including: (1) whether the claim involves solely a monetary interest, which inures solely to the benefit of a single party; (2) whether a constitutional or other significant societal interest is at stake; and (3) whether the outcome of the case affects a single, several, or many individuals in the filed case as well as individuals in other filed or likely to be filed cases.

No Change Discovery Cost	12.5% Increase Discovery Cost	25% Increase Discovery Cost
Monetary Relief Only	Nonmonetary Relief	Nonmonetary Relief
Affecting 1-5 persons	Affecting 10-50 Persons	Affecting more than 50
None	Constitutional/Societal Interest	Constitutional/Societal Interest

III. Estimating Overall Discovery Costs in Relation to Amount in Controversy as Adjusted by Importance of Issues at Stake

The table below starts with a case involving an amount in controversy of $5 million and discovery cost of $500,000 serving as a baseline. The Tymetrix study's formula (40% of total litigation costs or 10% of amount in controversy) is used to project reasonable overall discovery costs for amounts in controversy not exceeding $5 million, adjusted by the importance of the issues at stake, in appropriate circumstances. The FJC's formula ("0.25 percent increase in total litigation costs for every 1 percent increase in the amount in controversy") is used for amounts in controversy at $40 million and above. The average difference between the two studies is used for amounts in controversy between $5 million and $40 million.[39]

Amount in Controversy	Total Litigation Cost	Overall Discovery Cost[40]	Max (25%) Adjustment for Importance of Issues
$500,000	$125,000	$50,000	$12,500
$1,000,000	$250,000	$100,000	$25,000
$2,000,000	$500,000	$200,000	$50,000
$5,000,000[41]	**$1,250,000[42]**	**$500,000[43]**	**$125,000**
$10,000,000	$1,968,750	$1,262,500	$315,625
$20,000,000	$2,460,938	$1,578,125	$394,531
$40,000,000	$3,710,937	$2,968,750	$742,188

$80,000,000	$4,638,672	$3,710,938	$927,734
$160,000,000	$5,798,339	$4,638,671	$1,159,668
$320,000,000	$7,247,924	$5,797,339	$1,449,335
$640,000,000	$9,059,904	$7,247,923	$1,811,981
$1.28 Billion	$11,324,880	$9,059,904	$2,264,976

For example, cases with amounts in controversy of $5 million, $10 million, and $40 million, have typically involved cumulative discovery costs in the range of $500,000, $1,262,500, and $2,968,750, respectively. If the importance of the issues at stake were high and the maximum 25% adjustment is applied, the cumulative discovery costs would be $625,000, $1,578,125, and $3,710,938, respectively.

IV. Populating the Discovery Roadmap (Example)

The discovery roadmap below displays the application of the New Framework team of expert's estimates of average costs and number of gigabytes to a hypothetical group of 40 custodians, who have information located on more than 100 data sources. The amount in controversy pleaded in the complaint is $20 million. The case involves partial nonmonetary relief, affects between 20-50 individuals, and implicates a societal interest.

The discovery roadmap contains the key data points to make proportionality assessments, including:

- Prioritization of custodians and their data sources by the importance of information resolving the issues;
- Assessments of burden to access information from different data sources;
- Average per gigabyte costs of discovery from five common data sources;
- Average volume of data from five common data sources;
- Average discovery costs from custodians based on average volume and average per GB cost;
- Cumulative discovery costs for selected custodians; and
- Reasonable overall level of discovery costs based on amount in controversy and importance of issues at stake

The first horizontal colored band shows discovery of three custodians who possess the most important discoverable emails and file shares of $130,500. The second colored band shows discovery of emails and file shares of the next three custodians of $87,000 for a cumulative cost of $217,500. The third colored band shows discovery of emails, file shares, social media, laptops, and mobile devices of ten custodians of $446,477 for a cumulative cost of $663,977. The final-colored band shows discovery of at least 25 other custodians of $1,967,300 for a cumulative cost of $2,631,277.

A New Framework
Illustrative Discovery Roadmap for $20 Million in Controversy

Importance of Information Held by Custodian (Least → Most)	← Incremental Burden Scale (Low to Highest) → I	II	III	IV	V	(Insert column per addl. data source category)	Cumulative Discovery Cost / GB
4	John, Joe, Mary ($99K - 15GB)	John, Joe, Mary ($31.5K - 13.5GB)					$130,500 (28.5 GB)
3	Jack, Judy ($66K - 10GB)	Jack, Judy (21K - 9GB)					$217,500 (47.5 GB)
2	Larry, Allen, Roger, Louis, Trevor, Gordy, Sue ($231K - 35GB)	Larry, Sue, Trevor ($31.5K - 13.5GB)	John, Joe ($17K - .5GB)	John, Joe, Mary ($144K - 90GB)	John ($23K - 8GB)		$663,977 (194.5 GB)
1	25 others ($825K - 125GB)	15 others ($157.5K - 67.5GB)	5 others ($42K - 1.25GB)	10 others ($479K - 300GB)	20 others ($463K - 160GB)		$2,631,277 (848.25 GB)
	Email	**File Shares**	**Social Media**	**Computer/Laptop**	**Mobile Device**	**Other Sources**	
Average cost per custodian	$33,000	$10,500	$8,474	$47,963	$23,140		
Average per unit GB	5 GB	4.5 GB	.25 GB	30 GB	8 GB		
Average cost per GB	$6,600	$2,333	$33,896	$1,599	$2,892		

- Estimate of reasonable cumulative discovery cost in relation to the amount in controversy $1,578,125.
- 12.5% adjustment to account for the importance of issues at stake in the action $197,255.
- Estimate of total discovery cost proportional to the needs of the case $1,775,390.

In this example, the cumulative discovery costs estimated for all 40 custodians exceeds the projected reasonable discovery costs under the assessments proposed by the New Framework's team of experts.

One final valuation is required. Rule 26(b)(1) is clear that no weight is assigned to individual proportionality factors, which depend on the circumstances of each case.[44] Whether less or additional discovery is "proportional to the needs of the case" depends on whether greater or less weight is assigned to any of the four pertinent proportionality factors, a matter entirely contingent on the circumstances of the case in accordance with Rule 26(b)(1). For example, a user may assign little weight to the relation of the discovery cost to the amount in controversy when a defective drug on the market is alleged to have fatal effects. Similarly, a user may assign great weight to the importance of discovery in resolving the issues when only three custodians possess virtually all the relevant information in email.

Accordingly, discovery costs may be reasonable despite exceeding the projected reasonable amount of discovery cost or unreasonable despite falling below the projected reasonable amount of discovery cost, if one or more of the proportionality factors is given an enhanced or lower weight. The New Framework's discovery roadmap sets out all the key data points, which facilitates the assigning of weights and better informs the ultimate assessments of proportionality.

Section 05: Effective Uses of Heat Map and Discovery Roadmap

I. Introduction

Thus far, the workflow is as follows:

- Section 01 identifies and assesses custodian and non-custodian data sources into four levels of priority: Highest, High, Medium, and Low.
- Section 02 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 03 offers a Cost Prediction Calculator that permits a user to estimate 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 04 estimates reasonable discovery in relation to the amount in controversy and ascertains the importance of the issues at stake.

Section 05 explains how the New Framework can be effectively applied and used.

II. Applying the Heat Map and Discovery Roadmap

Filling in the heat map and the discovery roadmap are relatively straightforward, particularly if the standard inputs are adopted. Computing the cumulative discovery costs by designated custodians and data sources is equally straightforward. The results provide the key data points necessary for meaningful assessments of proportionality on a granular scale.

III. Uses of Heat Map and Discovery Roadmap

The heat map and discovery roadmap developed under the New Framework's methodology can better inform proportionality assessments. The producing party can use them for myriad purposes throughout litigation, including:

- Setting a discovery budget
- Calculating and projecting estimated costs for the collection, processing, and review of ESI using either standard unit costs or costs determined by the user
- Evaluating cost projections and invoices submitted by outside counsel and ediscovery vendors and service providers
- Justifying discovery expenditures to the client based on standard unit costs
- Informing preservation decision making
- Prioritizing discovery of custodian and data sources
- Facilitating discovery negotiations and "meet and confers" with the opposing side
- Creating opportunities to engage court in early active judicial discovery management and establishing a common framework for consideration of early discovery limits
- Establishing a record to defend discovery proportionality assessments grounded in metrics developed by 50 experts

IV. 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 can serve several useful purposes early in litigation. Although the information available at the commencement of litigation often is inadequate to make firm proportionality assessments that are reliably certain for all custodians, sufficient information is usually available to make reasonable assessments for custodians at both priority extremes on the New Framework's heat map. Sampling may be effective to assure accuracy.

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 discovery roadmap.

The New Framework's discovery roadmap provides the key data points, which can better inform the requesting part's assessments. Much of the information in the discovery roadmap is subject to disclosure. The former rule language recognizing the discoverability of such information was eliminated, because it was so well entrenched. Nonetheless, the matter remains subject to discovery, including the "existence, description, nature, custody, condition, and location of any documents or other tangible things and the identity and location of persons who know of any discoverable matter."[45]

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

- 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 inevitably result in significant production.[46]
- 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. [47]
- 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.[48]

B. Discovery Management

As additional information becomes available during discovery, the assessments in the discovery roadmap become more definitive and the proportionality assessments firmer.[49]

The proportionality assessments made when using the methodology in the New Frameworks' discovery roadmap can be used in discovery management to:

- Better inform negotiations and decision-making at the Rule 26 meet and confer conference as well as a later Rule 16 pretrial conference;[50]
- 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;[51]
- Better inform cost estimates 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; and
- Better inform the producing party when objecting to discovery based on proportionality grounds by providing the key data points to meet its "burden of making specific objection and showing that the discovery fails Rule 26(b)('s proportionality calculation by coming forward with specific information to address the importance of the issues at stake in the action, the amount in controversy, …the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit."[52]

C. Close of Discovery

The discovery roadmap contains the information necessary to effectively close out discovery.

- The decision when to release litigation holds is better informed by the New Framework's discovery roadmap, which highlights individual custodians most likely to possess information that can be released.[53] (*See* Appendix D.)
- Information in the New Framework's discovery roadmap better informs decisions to reopen discovery.[54]

The New Framework's workflow results in the discovery roadmap 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 Case Management and Resolution of Discovery Disputes

In most cases, discovery disputes do not arise or are handled by the parties at their meet and confers. In the small number of remaining cases, the parties usually focus exclusively on the importance of discovery to resolving the issue and resolve the matter without too much difficulty. But serious discovery disputes arise in a not insignificant number of cases, which can consume inordinate amounts of time of the court. And it often is impossible at the outset of litigation to determine which cases may eventually generate discovery problems.

The 2015 Committee Note to Rule 26 recognizes that there will be important occasions for "judicial management, when the parties are legitimately unable to resolve important differences and when the parties fall short of effective, cooperative management on their own."[55] A court also has an independent obligation under Rule 26 to assess whether discovery is proportional to the needs of the case.[56] The information in the New Framework's discovery roadmap provides a judge (and both parties) an ideal case-management tool, which contains essential information in an organized format that can be used to better inform evaluation of Rule 26(b)(1) proportionality disputes and evaluate

whether counsel's discovery efforts are reasonable. Sampling offers a potential practical solution to evaluate disputes about the prioritizing of custodians.

The New Framework's standard approach also offers estimates of average costs and number of gigabytes that can be used to compare the parties' estimates for reasonableness. The approach carries out the Advisory Committee on Civil Rules recommendation admonition that the "burden or expense of proposed discovery should be determined in a realistic way."[57] The discovery roadmap 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 with the parties' input can better evaluate the overall reasonableness of counsel's proportionality assessments, underlying rationales, and discovery decisions and decide whether additional discovery is appropriate focusing on individual custodians and data sources.

The New Framework's team of experts' estimates also provides a judge reference points to evaluate suspiciously high claimed costs along with an itemized cost calculator that can be used to examine the variances, which may explain the discrepancy in cost.

A judge can consider the New Framework's custodian prioritizing and data-source burden assessments and cost projections at several key litigation mileposts:

- 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.[58]
- 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.[59]
- 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 discovery roadmap 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.[60] For example, at a glance a judge could determine that the discovery of email from three additional custodians would cost $100,000 and add 90,000 pages to collect, process, and review. The judge would have the full range of data points contained in the discovery map to compare and ascertain proportionality.
- A judge can use the information to evaluate and resolve sanction motions. Under Rule 26(g), a party must certify "that the request, response, or objection was 'not unreasonable or unduly burdensome or expensive, given the needs of the case, the discovery already had in the case, the amount in controversy, and the importance of the issues at stake in the litigation.'"[61] The New Framework's discovery roadmap provides 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 can better inform the judge's decision on the reasonableness of any specific discovery action taken by counsel.[62]

Rule 26(b)(1) demands a holistic evaluation of four pertinent factors, which defies certainty and mathematical precision, to assess whether discovery is proportional to the needs of the case. The NEW FRAMEWORK clarifies the key decision points to assess proportionality, provides a team of experts' averages based on typical expectations, and lays out an analytical framework and criteria addressing each of the four Rule 26(b)(1) factors to identify the most important discoverable information.

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.

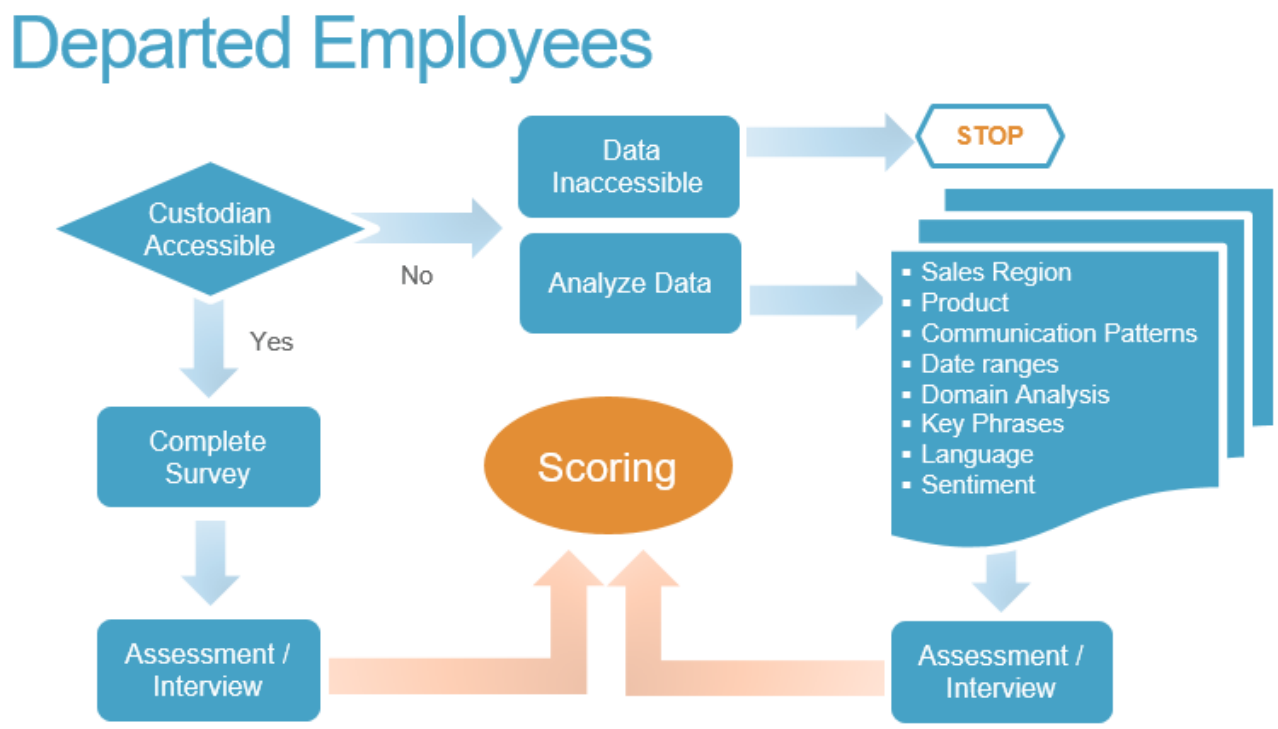
The goal is to offer a tool that provides an objective assessment of each custodian based on a series of questions, which reflect the materiality, strength, and uniqueness of their data.

This template will be made available in spreadsheet format for use in assessment creation.

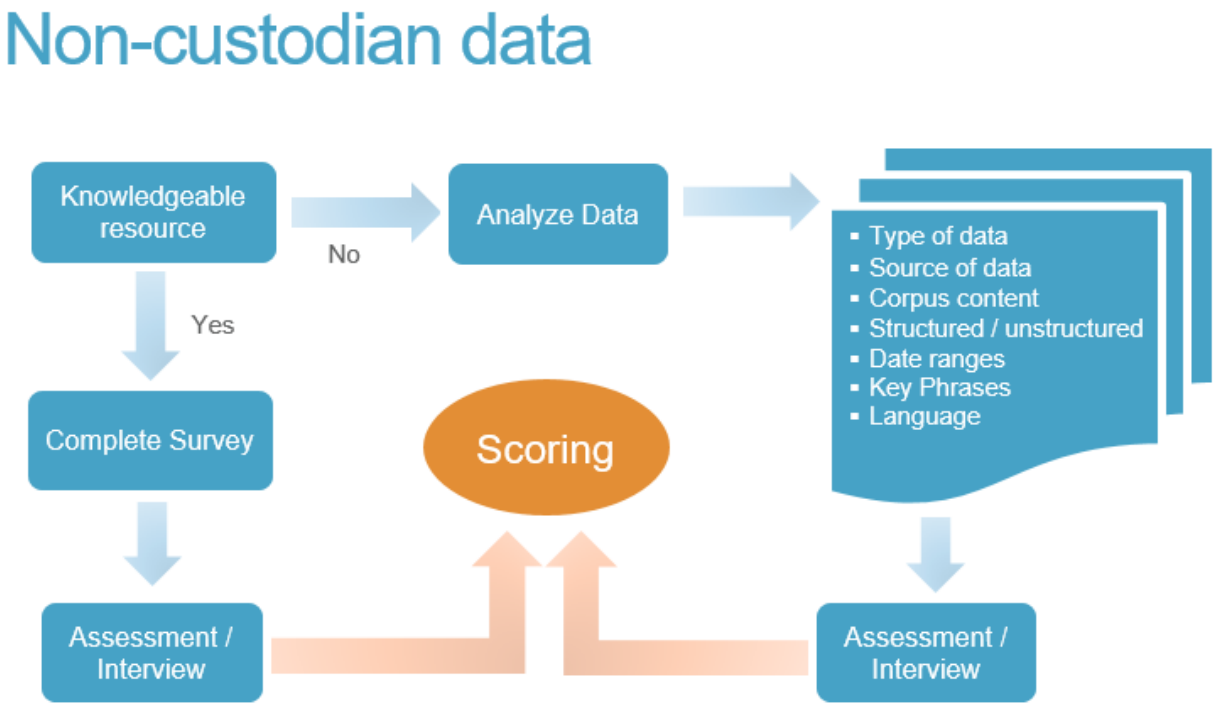
Discovery Proportionality Model Custodian Assessment	
Custodian	
Custodian Qualifying Questions	**Select Answer**
Was the custodian employed by the company within the relevant time period?	
Are the custodian position and responsibilities relevant to the matter?	
Is the department the custodian worked within relevant to the matter?	
CUSTODIAN QUALIFICATION	RELEASE
Custodian Relevancy Questions	**Select Answer**
Was the custodian involved in internal/external meetings or discussions related to the matter?	
Did the custodian create or participate in email discussions or email threads, or were they copied on email threads related to the matter?	
Did the custodian create or contribute to any documentation, presentations, design or product materials or manuals related to matter?	
Did the custodian directly or indirectly communicate with anyone outside the organization on any topic related to the matter?	
Was the custodian aware of, or participate in, any social media discussions regarding the matter?	
Data Relevancy Questions	**Select Answer**
Is this custodian's data material to the claims & defenses of the matter? ***Materiality factors:*** - addresses the heart of the case or a subsidiary issue - proves an ultimate fact or an intermediate fact - is an essential link in a line of evidence needed to prove an assertion	
Is this custodian's data strongly connected and important to the claims in the case? ***Strength/connection factors***: - direct or circumstantial evidence - complete and thorough	
Is this custodian's data unique and not in the possession or control of others? Uniqueness factors: - distinguishing similar information from that which is qualitatively different	
PRIORITY	LOW

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.

Appendix B – Atypical Case – Departed Employee



Appendix C – Atypical Case – Non-Custodian Data



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.

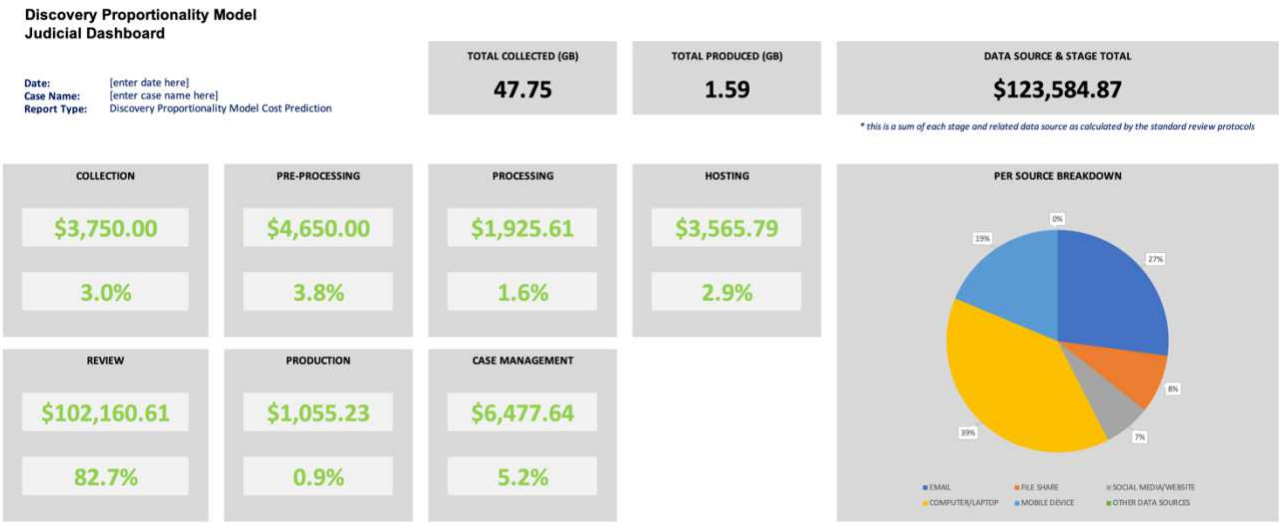
Discovery Proportionality Model BURDEN ASSESSMENT TOOL								
ESI Source	**Collaboration / Messaging**	**Computer / Laptop**	**Email**	**File Share**	**Mobile Device**	**Paper**	**Social Media**	**Structured Data**
Burden Factor	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*	*Choose from Dropdown*
Data Centralization								
Data Location								
Data Availability								
Collection Methodology								
Required Resources								
Legal & Regulatory Issues								
Burden Classification:								

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

Appendix F – Itemized Cost Prediction Calculator (Excerpts)

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

Judicial Dashboard



Judicial Detail View

Discovery Proportionality Model Judicial Detail View						
STAGE	**Hosted email** (5GB)	**File Share** (4.5GB)	**Social Media/ Website** (0.25 GB)	**Computer/ Laptop** (30 GB)	**Mobile Device** (8 GB)	**STAGE TOTAL**
GB Collected	5.0	4.5	0.25	30	8	47.75
GB Produced	0.44	0.11	0.04	0.79	0.21	1.59
Collection	750.00	1,150.00	500.00	600.00	750.00	3,750.00
Pre-Processing	450.00	1,200.00	1,800.00	600.00	600.00	4,650.00
Processing	436.54	108.31	21.19	1,195.49	164.08	1,925.61
Hosting	827.33	514.14	445.21	1,166.16	612.95	3,565.79
Review	29,500.20	6,076.01	4,255.31	42,804.80	19,524.30	102,160.61
Production	233.89	171.23	157.44	301.67	191.00	1,055.23
Management/Support	1,295.53	1,295.53	1,295.53	1,295.53	1,295.53	6,477.64
DATA SOURCE TOTAL:	$ 33,493.48	$ 10,515.21	$ 8,474.67	$ 47,963.65	$ 23,137.85	$ 123,584.87

Attorney Detail View

Discovery Proportionality Model Attorney Dashboard					
	Hosted email	File Share	Social Media/ Website	Computer/ Laptop	Mobile Device
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	6.16	843.75	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	2,916.67	29,166.67	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.

The calculator 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

Complexity Factor	Impact
Cloud based (ex. YouTube, Wiki's), Filesharing (ex. Dropbox, Google Drive) Social networking (ex. WhatsApp, Facebook) or collaborative applications (ex. Slack, Yammer, Teams)	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.
Mobile devices including phones, tablets, external drives	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.
Large volumes of hard copy data	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.
Investigations, or matters with investigative elements	Forensic images to be created and complex forensic analysis performed on various laptops, or other devices make this is a more labor intensive, time consuming analytical activity.
Unusual Source data: - Audio/video - complex data types (see Structured Systems) - non-standard files (engineering drawings, medical imaging) - legacy applications or repositories	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.
Targeted v. full collections	A full collection may increase the processing costs due to the overall size of the data set collected.
Third-party email collection	Email from private custodian accounts or non-standard email types may present a greater challenge to collection procedures.
Foreign Language	Foreign-language documents may require special processing procedures as well as translation or translation services.
High-risk discovery factors present	May include risk of spoliation of the data, an uncooperative client, or involvement of bad actors.
Unusually contentious litigation	Lack of cooperation or consistent roadblocks or objections to methodologies proposed.

Cross-border discovery	The logistics of collection and compliance with local regulations.
Privacy	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)
Structured Systems	Special expertise and skills may be needed to gather, compile, and render ESI for structured data sources. Review costs may be mitigated because fewer attorneys are reviewing data compilations and summary reports. Although higher cost, non-attorney reviewers are often needed (*e.g.*, accountants, engineers, nurses).

Appendix H – Standard-Unit Cost Table

Discovery Proportionality Model COSTS FOR TYPICAL DATA SOURCE VOLUMES AND UNIT COSTS					
ESI STAGE	**Hosted Email**	**File Shares**	**Social Media/Website**	**Computer/ Laptop**	**Mobile Device**
Assumption: Typical GB per Custodian	5	4.5	0.25	30	8
Collection	$ 750.00	$ 1,150.00	$ 500.00	$ 600.00	$ 750.00
Processing & Hosting	$ 1,713.87	$ 1,822.45	$ 2,266.40	$ 2,961.65	$ 1,377.03
Review, Production & Management/Support	$ 31,029.61	$ 7,542.77	$ 5,708.28	$ 44,402.00	$ 21,010.82
Estimated Cost applying typical GB per Custodian:	$ 33,493.48	$ 10,515.22	$ 8,474.68	$ 47,963.65	$ 23,137.85
Estimated Unit Cost per one Gigabyte:	$ 6,698.70	$ 2,336.72	$ 33,898.72	$ 1,598.79	$ 2,892.23

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

The cost table is freely available to the public.

APPENDIX I – Discovery Roadmap

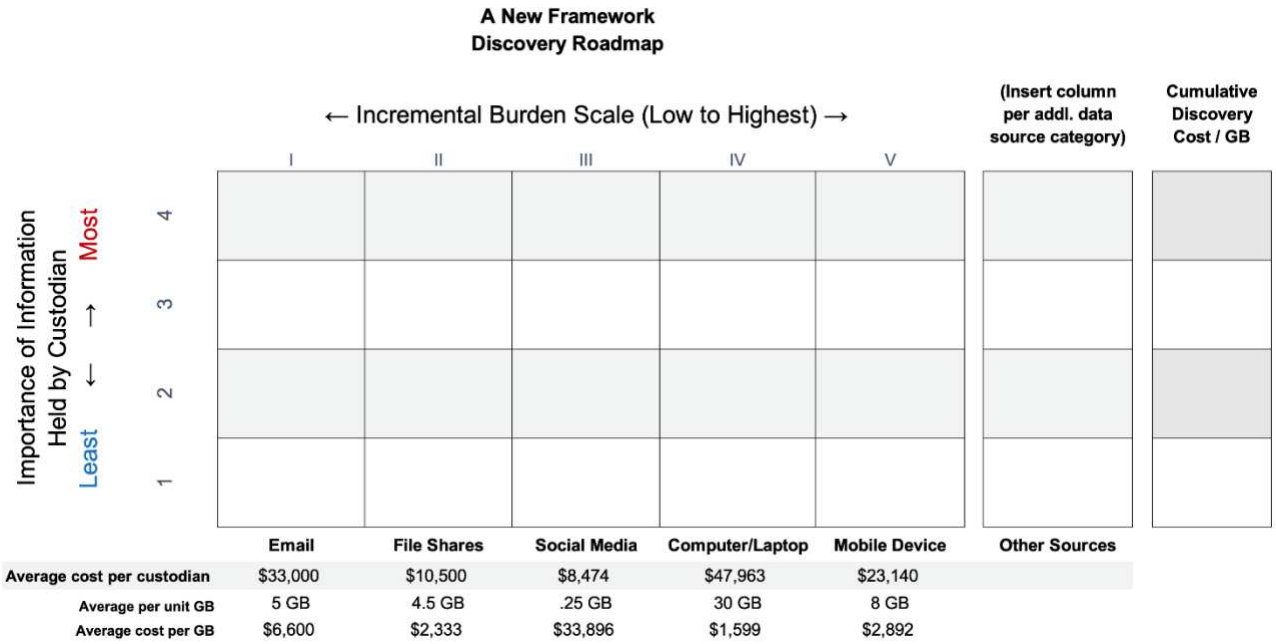


Table Estimating Overall Discovery Costs in Relation to Amount in Controversy as Adjusted by Importance of Issues at Stake

Amount in Controversy	Total Litigation Cost	Overall Discovery Cost[63]	Max (25%) Adjustment for Importance of Issues
$500,000	$125,000	$50,000	$12,500
$1,000,000	$250,000	$100,000	$25,000
$2,000,000	$500,000	$200,000	$50,000
$5,000,00064	$1,250,000[65]	$500,000[66]	$125,000
$10,000,000	$1,968,750	$1,262,500	$315,625
$20,000,000	$2,460,938	$1,578,125	$394,531
$40,000,000	$3,710,937	$2,968,750	$742,188
$80,000,000`	$4,638,672	$3,710,938	$927,734
$160,000,000	$5,798,339	$4,638,671	$1,159,668
$320,000,000	$7,247,924	$5,797,339	$1,449,335
$640,000,000	$9,059,904	$7,247,923	$1,811,981
$1.28 Billion	$11,324,880	$9,059,904	$2,264,976

INSTRUCTIONS

1. Prioritize custodians who possess or control relevant information as well as databases in four categories by the importance of their discoverable information. Insert by name custodians in the appropriate data-source column where the information resides on in accordance with their priority—insert NONE if no custodian is plotted (columns I-IV) (*see* pages 17-18).

2. Calculate the discovery costs of every custodian by applying the "average" costs and gigabyte volume estimates at the bottom of each data-source column. For example, the cost for two custodians with 5 gigabytes each, which are plotted in the highest priority email column, would be $66,000. Alternatively, substitute the user's own average costs and number of gigabytes to fit the circumstances. Within each column and row, insert discovery cost and number of gigabytes beneath names of custodians in parenthesis.
3. Add the costs of individual custodians in each of the four rows and insert the cumulative cost in the last column for each row of custodians (cumulative discovery cost) (rows 1-4) (*see* pages 17-18).
4. Estimate the reasonable overall discovery in relation to the amount in controversy. Adopt and apply estimates calculated by team of experts in the NEW FRAMEWORK shown in table for amounts in controversy ranging from $100K to $1 billion (*see* pages 16-17). Alternatively, substitute own estimates and be prepared to support substitution with appropriate reasons. in accordance with table below the discovery roadmap. Alternatively, substitute the user's estimates.
5. Estimate and calculate an adjustment of up to 25% of discovery cost or substitute a different percentage based on own experience to account for the "importance of the issues at stake" (*see* page 16).
6. Assess whether discovery is proportional to the needs of the case taking into consideration whether any proportionality factor should be given more or less weight than the other factors (*see* page 18).

GLOSSARY

Term	Definition
Email threading	Automated process that organizes the emails in an email conversation so that they may be reviewed together in context.
Near Duping	Automated process of identifying documents that are nearly identical in content for the purposes of expedited review.
Culling	Process of reducing the overall size of the data set. May consist of targeted keyword searches, date-range filtering, deNist, and deduping.
Culling: deNist	Data-processing step of the removal of file types unlikely to have evidentiary value such as system files, executables, etc.
Culling: Dedupe	Data-processing step of identifying and removing identical duplicates from the dataset.
Custodians	Persons who have custody or control of data subject to preservation.
Data Expansion	The processing step whereby data volume is expanded due to the opening and extraction of zip files, PST files, etc.
Data Source: COLLABORATION/MESSAGING	Applications such as MS Teams, Slack, Salesforce, etc.
Data Source: COMPUTERS/LAPTOP	Company issued or BYOD hardware.
Data Source: FILE SHARE	Shared network-storage systems, cloud storage, etc.
Data Source: HOSTED EMAIL	A centralized server or a cloud-based email system.
Data Source: MOBILE DEVICE	Includes company issued or BYOD mobile devices such as Iphones and tablets.
Data Source: PAPER	Physical documents and the OCR/scanning process associated with converting the documents to a machine-readable format.
Data Source: SOCIAL MEDIA	Facebook, Twitter, Instagram, and other Social Media or web-based platforms.
Data Source: STRUCTURED SYSTEMS	Database applications such as Human Resource, Accounting, Finance systems.

ESI Stage: ANALYTIC WORKFLOW OPTIONS	May include predictive-coding methodologies such as TAR/CAL, data visualization, or topic classification.
ESI Stage: COLLECTION	Process of identifying and physical collecting data from targeted data sources.
ESI Stage: HOSTING	Storage of collected data in readily accessible platform for document review, includes user fees or licenses.
ESI Stage: PRE-PROCESSING CULLING	Targeted collection strategies such as keyword, date filtering, or data-source selection.
ESI Stage: PROCESSING	Multi-step function whereby data is extracted, normalized, and prepared for document review and production.
ESI Stage: PRODUCTION	Process of turning the results of the attorney review into something that can be produced to the requesting party.
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 production.

Endnotes

[1] Committee Note, Fed. R. Civ. P. 26(b)(1) (2015).
[2] "The present amendment restores the proportionality factors to their original place in defining the scope of discovery. This change reinforces the Rule 26(g) obligation of the parties to consider these factors in making discovery requests, responses, or objections." Committee Note, Fed. R. Civ. P. 26(b)(1) (2015).
[3] The New Framework is freely available to the public and posted on the Center's website. The New Framework does not grant any rights to the use of Evidence Optix® software or intellectual property.

[4] *See* endnotes 27, 28, and 29 *infra*.
[5] "The direction to consider the parties' relative access to relevant information adds new text to provide explicit focus on considerations already implicit in present Rule 26(b)(2)(C)(iii). Some cases involve what often is called 'information symmetry.' One party – often an individual plaintiff – may have very little discoverable information. The other party may have vast amounts of information, including information that can be readily retrieved and information that is more difficult to retrieve. In practice these circumstances often mean that the burden of responding to discovery lies heavier on the party who has more information, and properly so." Committee Note, Fed. R. Civ. P. 26(b)(1) (2015).
[6] "So too, consideration of the parties' resources does not foreclose discovery requests addressed to an impecunious party, nor justify unlimited discovery requests addressed to a wealthy party. The 1983 Committee Note cautioned that '[t]he court must apply the standards in an even-handed manner that will prevent use of discovery to wage a war of attrition or as a device to coerce a party, whether financially weak or affluent.'"
[7] "This first Rule 26 factor calls for the court to 'examine[] the significance of the substantive issues [at stake in the litigation], as measured in philosophic, social, or institutional terms. BlueAlly, 2017 U.D. Dist. LEXIS 30558, 2017 WL 876266, at *4 (quoting Fed. R. Civ. P. 26 advisory committee's note). For example, courts should carefully scrutinize discovery requests in 'cases in public policy spheres, such as employment practices, free speech, and other matters,' which often 'seek[] to vindicate vitally important personal and public values' and 'may have importance far beyond the monetary amount involved." Oxbow Carbon & Minerals LLC v. Union Pac. R.R. Co., 322 F.R.D. 1 (D.D.C. 2017).
[8] 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.
[9] 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.

[10]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.

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.

[11] "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…."

[12] An 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.

[13] For instance, determining materiality in a theft of 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. On the issue of whether the information allegedly taken was actually secret, materiality may involve information tending to show the steps taken to protect the trade secrets and information tending to show that trade secrets were not generally known.

[14] 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.

[15] *See* extensive list of citations supporting the proposition at Endnote 20, GUIDELINES AND BEST PRACTICES FOR IMPLEMENTING THE 2015 DISCOVERY AMENDMENTS CONCERNING PROPORTIONALITY, Bolch Judicial Institute, Duke Law School, 3d Ed. (April 2021).

[16] 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.

[17] Art. 6 GDPR.

[18] Art.13 GDPR.

[19] Art. 49 GDPR.

[20] https://edrm.net/edrm-model/.

[21] The NEW FRAMEWORK is designed to forecast costs from collection through production. A key part of the process -- preservation costs -- varies so widely based on individual circumstances that the costs associated with it are not included in the Cost-Projection Calculator.

[22] 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.

[23] 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. *See also Technology Assisted Review (TAR) Guidelines,* Bolch Judicial Institute, Duke Law School (January 2019).

[24] Robinson, R. (2020). A 2020 Look at eDiscovery Collection: Task, Spend, and Cost Data Points, https://complexdiscovery.com/a-2020-look-at-ediscovery-collection-task-spend-and-cost-data-points/.

[25] 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 can be significant.

[26] 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.

[27] Depending on the source and format of the data, the number of pages and documents contained in a single gigabyte will vary significantly. For example, there may be 30,000 pages and 4,838 documents of email (1.5 pages per document) in a single gigabyte and 14,300 pages and 1,059 documents of Word text (13.5 pages per document) in a single gigabyte. These numbers will also vary by the type of discoverable information.

[28] List of references here, *e.g.*, EDRM, Sedona Conference, Federal Judicial Center references for judges https://www.fjc.gov/subject/electronic-discovery, 7th Circuit Council on eDiscovery and Digital Information, Northern District of California E-Discovery (ESI) Guidelines.

[29] The costs of discovery from individual custodians can be calculated using the average volume of data from designated sources along with the per gigabyte costs typically expected, e.g., cost of "average" email from a custodian is $33,493 (5GB typical volume x $6,697 average per GB email cost).

[30]Despite Rule 26 requirements, parties usually neglect or ignore them. In fact, most conflate Rule 26(b)(2)(C)(i) with the proportionality rule. They focus solely on whether the discovery provides marginal utility and whether "discovery sought is unreasonably cumulative or duplicative or can be obtained from some other source that is more convenient, less burdensome, or less expensive." Although this argument can resolve many, perhaps most, discovery disputes, Rule 26(b)(1) has a broader scope.

[31] 253 F.R.D. 354, 364 (D. Md. 2008). The usefulness of comparing discovery costs and the amount in controversy was recognized by Judge Paul Grimm in the seminal *Mancia v. Mayflower Textile Servs. Co*., when he noted that "While admittedly a rough estimate, this range [worst and best likely outcomes to estimate the amount in controversy] is useful for determining what the 'amount in controversy' is in the case, and what is 'at stake' for purposes of Rule 26(b)(2)(C)'s [26(b)(1) later amended in 2015] proportionality analysis. The goal is to attempt to quantify a workable 'discovery budget' that is proportional to what is at issue in the case."

[32] U.S. EEOC v. George Washington University, 2020 WL 3489478, at *6 (D.D.C. June 26, 2020).

[33] Guideline 2(B), REVISED GUIDELINES AND PRACTICES FOR IMPLEMENTING THE 2015 DISCOVERY AMENDMENTS TO ACHIEVE PROPORTIONALITY, Duke Law Center for Judicial Studies (Jan. 20, 2017*); see also Preventive Energy Sol. v. nCap Ventures 5*, 2021 WL 2930653 at *4 (D. Utah, July 11, 2021) ("Knowing the amount of claimed damages affects how a court resolves discovery disputes (i.e., Rule 26(b)(1) requires the court to consider the 'amount in controversy') and how the parties approach settlement and the resources necessary to devote to discovery and trial. To be sure, in the beginning of a case, this type of precise calculation may not be possible because necessary information may be in the hands of the opposing party. But the parties are expected to disclose all the details they can with the information that they have and cannot refuse to disclose because they not yet received discovery."

[34] If no damages are pleaded in the complaint, estimating the amount of controversy is a subjective judgment, which is fraught with difficulty and imprecision. But the definition of discoverable matter under Rule 26(b)(1) requires its consideration and, as a practical matter, most producing parties take the amount in controversy into account when determining how much discovery is reasonable. Similar to other projections, the estimated amount in controversy may change during litigation when more information becomes available, and conclusions are firmer. Perhaps at no stage is this clearer than after a ruling on a motion to dismiss.

Notwithstanding a presumption of good faith, a party's estimate of the amount in controversy may be vulnerable to gamesmanship and a fair degree of skepticism. This estimate as well as many other similar judgments made throughout the proportionality assessment analysis is subject to self-policing. Unsupportable projections of the amount in controversy would work against self-interests. Lowballing projected amounts in controversy might temporarily limit discovery, but it would undermine the legitimacy of all proportionality assessments, damage credibility, require expensive makeovers, incur additional costs, and invite a court's rebuke, or worse, sanctions.

See Preliminary Report to the Judicial Conference Advisory Committee on Civil Rules, National, Case-Based Civil Rules Survey, Federal Judicial Center (October 2009). The Federal Judicial Center analyzed the results of a survey of more than 2,500 attorneys in closed civil cases. At the 95th percentile they found that $5 million was an appropriate amount of controversy projected by the defendants. The median stake was $160,000.

[35] Oxbow Carbon & Minerals LLC v. Union Pac. R.R. Co., 322 F.R.D. 1 (D.D.C. 2017).
[36] Litigation Costs in Civil Cases: Multivariate Analysis, Report to the Judicial Conference Advisory Committee on Civil Rules, Emery Lee & Thomas Willging, Federal Judicial Center (March 2010).
[37] Most studies estimate discovery costs as a percentage of the "total litigation costs" and not as a percentage of the "amount in controversy." Separate studies estimate the "total litigation cost" as a percentage of the "amount in controversy." Analysis of both types of studies is necessary to calculate what percentage of the amount in controversy discovery costs represent. Discovery costs typically represent 30 - 45 percent of total litigation costs, which themselves represent 25 percent of the amount in controversy. The FJC study found that at the 95th percentile total litigation costs were $400,000-$600,000, depending upon how many types of discovery were used. The median was $40,000. Preliminary Report to the Judicial Conference Advisory Committee on Civil Rules, at 37, supra n. 27. Discovery costs as a percentage of total litigation costs ranged from 80 percent at the 95th percentile to a median of 30 percent. Id. at 38-39.

The FJC found that defendants projected the ratio of discovery costs to the amount in controversy to be 30 percent at the 95th percentile and as low as 3.3 percent as the median. Id. at 43.

The Tymetrix study analyzed attorney invoices submitted by AmLaw 100, 101-200, and unranked law firms in patent, product liability, and personal injury/wrongful death litigation. Copy of study on request to Center.

See also, LITIGATION COST SURVEY OF MAJOR COMPANIES, Lawyers for Civil Justice; Civil Justice Reform Group; and U.S. Chamber Institute for Legal Reform, administered and data compiled by Searle Center on Law, Regulation, and Economic Growth, Northwestern Law (May 10-11, 2010). The average discovery cost per case (exceeding $250,000) in 2008 among 20 major corporations participating in the study was $621,880 compared with an average $2,019,248 total litigation cost or 30%.

[38] LITIGATION COSTS IN CIVIL CASES: MULTIVARIATE ANALYSIS, at 8 supra. n. 25
[39] The findings of the Tymetrix study are based on hard empirical data. But calculating discovery costs in high-end cases involving more than $20 million using the 40% of total litigation cost formula resulted in unreasonably low discovery costs. On the other hand, the FJC's formula raising discovery costs by 25% for every 100% increase in the total litigation costs, resulted in unreasonably high discovery costs in cases at or below $10 million. Accordingly, the table relies on the Tymetrix study for setting the discovery cost for the baseline $5 million case and cases below $5 million in the amount in controversy, while using an average of the discovery costs computed under both the Tymetrix and FJC formulae for the $10 million and $20 million cases. The FJC formula is applied to all cases exceeding $40 million.
[40] Based in part on 2009 FJC study, which projects 80% of total litigation costs incurred in discovery in cases at the 95th pecentile (Tables 6 and 7). Also, based in part on Wolters Kluwer (Tymetrix division) study of attorney-invoices study, which shows that discovery costs represent 28%-33% of total litigation costs. The study covered: (i) personal injury cases – 800-1200 hours billable discovery; (ii) patent litigation -- +1500 hours billable discovery; and (iii) product liability -- 1200-2000 hours billable discovery. Percentage was increased to 40% to account for discovery-vendor costs, which were excluded from their study.
[41] Baseline $5 million amount in controversy based on FJC 2009 study median at 95th percentile (Table 9), which is consistent with discussions with lawyers and experts handling civil cases in federal courts with intensive discovery.
[42] Based on 25% of amount in controversy, which is consistent with discussions with lawyers and experts handling civil cases in federal courts with intensive discovery regarding a $5 million case.
[43]Based on FJC 2009 study, which shows that discovery costs represented 30.5% of the median $5 million case at 95th percentile (Table 10).
[44]But "'no single factor is designed to outweigh the other factors in determining whether the discovery sought is proportional,' and all proportionality determinations must be made on a case-by-case basis." Oxbow Carbon & Minerals LLC v. Union Pac. R.R. Co., 322 F.R.D. 1 (D.D.C. 2017), quoting in part Williams v. BASF Catalysts, LLC 2017 U.S. Dist. LEXIS 122053, 2017 WL 3317295, at *4 (D.N.J Aug. 3, 2017).

[45] Id. "Framing intelligent requests for electronically stored information, for example, may require detailed information about another party's information systems and other information resources."
[46] 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.

[47] 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.

[48] 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.

[49] 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 reevaluate whether preservation and collection of each data source identified in the process is proportionate.

[50] 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.

[51] 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.

[52] *Banks v. St. James Parish Sch. Bd*., 2021 WL 2070132 (E.D. Lou. May 15, 2021).
[53] 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.

[54] 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 NEW 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.

[55] Committee Note, Fed. R. Civ. P. 26 (2015).

[56] *See Durbin v. C & L Tiling Inc.,* 2019 U.S. Dist. LEXIS 161967, at *20 (W.D. Ky. Sept. 23, 2019) ("the lack of valid objection by [parties] does not obviate the Court's independent obligation" to limit discovery. *See* extensive list of citations supporting the proposition at Endnotes 46 and 48, GUIDELINES AND BEST PRACTICES FOR IMPLEMENTING THE 2015 DISCOVERY AMENDMENTS CONCERNING PROPORTIONALITY, Bolch Judicial Institute, Duke Law School, 3d Ed. (April 2021).

[57] Id.

[58] 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.

[59] 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.

[60] The NEW FRAMEWORK can be used to help a judge understand where to draw the line in enforcing or quashing a request for production. A judge can readily compare the cost presented with the amount in controversy.

[61] Committee Note, Fed. R. Civ. P. 26(b)(1) (2015) citing provisions of Fed. R. Civ. P. 26(g). "This change [restoring proportionality factors to their original place] reinforces the Rule 26(g) obligation of the parties to consider these factors in making discovery requests, responses, or objections." Committee Note to Rule 26(b)(1) (2015).

[62] 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 NEW FRAMEWORK can be a guideline for determining whether spoliation has occurred, and sanctions might be deemed necessary.

[62] Based in part on 2009 FJC study, which projects 80% of total litigation costs incurred in discovery in cases at the 95th pecentile (Tables 6 and 7). Also, based in part on Wolters Kluwer (Tymetrix division) study of attorney-invoices study, which shows that discovery costs represent 28%-33% of total litigation costs. The study covered: (i) personal injury cases – 800-1200 hours billable discovery; (ii) patent litigation -- +1500 hours billable discovery; and (iii) product liability -- 1200-2000 hours billable discovery. Percentage increased to 40% to account for discovery-vendor costs, which are excluded from their study.

[64] Baseline $5 million amount in controversy based on FJC 2009 study median at 95th percentile (Table 9), which is consistent with discussions with lawyers and experts handling civil cases in federal courts with intensive discovery.

[65] Based on 25% of amount in controversy, which is consistent with discussions with lawyers and experts handling civil cases in federal courts with intensive discovery regarding a $5 million case.

[66]Based on FJC 2009 study, which shows that discovery costs represented 30.5% of the median $5 million case at 95th percentile (Table 10).