BENEFITS AND COST PROPORTIONALITY MODEL

A NEW FRAMEWORK

James F. Humphreys Complex Litigation Center

George Washington Law School

February 15, 2021

(WORK-IN-PROGRESS DRAFT)

Center Patrons

Western Alliance Bank

Technology Concepts & Design (TCDI)

BENEFITS AND COST PROPORTIONALITY MODEL
A NEW FRAMEWORK
GW Complex Litigation Center
February 15, 2021

Preface

The Benefits and Cost Proportionality Framework (New Framework) focuses on custodians and compares the significance of the information that they possess or control with the burden and costs in searching and retrieving it, establishing a standard and cogent approach to frame proportionality assessments.

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

The George Washington Law School's James F. Humphreys Complex Litigation Center and John Rabiej, former Director of Duke Law School's Center for Judicial Studies, launched the project in early 2020.

A steering committee of 15 lawyers, judges, and ediscovery experts developed an outline, recommended candidates to draft the New Framework, and serves as an editorial board under the auspices of the GW Complex Litigation Center.

The editorial board will review and revise the draft New Framework before publishing it for public comment. The drafting team members will review all public comments and revise the New Framework, if appropriate. The editorial board will make final edits, subject to the Center's oversight. The New Framework represents the authors' consensus views, but not necessarily unanimous approval of all content.

The New Framework as described was originally derived from Prism Litigation Technology's Evidence Optix® software (patent pending). The Center and the drafters readily acknowledge the important role Prism played in advancing the New Framework.

But the current version of the New Framework is unlike the software program in key and numerous minor respects. For example, the New Framework relies on three objective criteria, including materiality, uniqueness, and first-hand knowledge, instead of "relevance" as the criterion for prioritizing custodians. It also alters the method data sources are ranked according to burden and itemizes the individual elements of cost incurred in discovery. The New Framework discusses how these methods can be used during the discovery stages. Most significantly, the New Framework estimates standard-unit costs for the eight data sources as baseline reference points. All these variances were developed independently by the 56 experts engaged in this project.

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

ACKNOWLEDGEMENTS

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

With great gratitude, the center recognizes the contributions of the judges, lawyers, and other experts who contributed to this effort.

STEERING COMMITTEE

Arpert, Douglas (Hon.) District of New Jersey	Basile, Bobbi HBR Consulting	Baylson, Michael (Hon.) Eastern District of Pennsylvania
Gensler, Steven The University of Oklahoma	Hanks, George (Hon.) Southern District of Texas	Johnston, Iain (Hon.) Northern District of Illinois
Kaufman, Beth Schoeman Updike Kaufman & Gerber	Opsitnick, Tim TCDI	
Regard, Daniel Intelligent Discovery Solutions	Ross, Mandi Prism Litigation Technology	Yates, Christopher (Hon.) 17th Circuit Court, Michigan

TEAM LEADERS

Basile, Bobbi HBR Consulting	Belt, William CDS	Kenney, Colleen Sidley Austin
Lowe, Anthony DATALAW PLLC	Quartararo, Michael ACEDS	Regard, Daniel Intelligent Discovery Solutions
Ross, Mandi Prism Litigation Technology	Trask, Andrew Shook, Hardy & Bacon	Tully, Martin Actuate Law

TEAM MEMBERS

Addington, John
Corning Law

Baratz, Michael
Steptoe & Johnson

Bauer, Kate
Steptoe & Johnson

Behan, Kevin
Centene Corporation

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

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
Tanenbaum Keale

Murao, Joy
Practice Aligned

Murray, Barb
PWC

O'Brien, Kate
Prism Litigation Technology

O'Connor, Tom
Gulf Coast Legal

Pass, Robert
Carlton Fields

Patzakis, John
X1

Poplawski, Matthew
Winston & Strawn

Redgrave, Jonathan
Redgrave

Rizzolo, Dan
ePercept

Robinson, Rich
Toyota Legal One

Schouten, Nataliya
FIS Global

Scrutton, Alexandra Grand
3M

Shortnacy, Michael
King & Spaulding

Six, Michelle
Kirkland & Ellis

Stadler, Anna-Patricia
Robert Bosch GmbH

Staunton, Thomas
Miller Shakman Levin & Feldman

Sullivan, Janet
White & Case

Sweeney, Caroline
Dorsey & Whitney

Yates, Janice
Prism Litigation Technology

Introduction

I. Background

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

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

Although unfamiliarity with the 2015 amendments remains an issue impeding their effectiveness, the absence of a standard and cogent approach to frame proportionality assessments for the parties' discussions and a court's consideration has stymied real progress. The New Framework is designed to adequately explain and substantiate proportionality arguments, which will further the goals of the 2015 Rule 26 amendments.

II. Overview

Experienced ediscovery lawyers, consultants, and service providers have experimented with many approaches to assess proportionality. The New Framework echoes many of the best aspects of these approaches and establishes a new analytical framework, which sets out a systematic process that applies objective criteria to the rule-based elements of proportionality. It identifies significant discoverable documents and ESI for initial production while distinguishing other discoverable matter, which requires additional probing, party negotiations, and perhaps sampling to fully quantify the "benefits/cost" proportionality factor. It fully documents proportionality assessments in an organized fashion by custodian and data source, facilitating party negotiations and a judicial ruling, if necessary.

The New Framework primarily addresses the application of the sixth proportionality factor, the benefits/cost factor, though it also implicates the factors dealing with "amount in controversy" and the "importance of discovery in resolving issues." The New Framework does not address proportionality factors addressing the parties' resources, the importance of issues at stake, and the parties' relative access to relevant information.

As its name implies, the New Framework sets out a skeletal outline that highlights the classification of custodians and their respective data sources (defined as both electronic and physical) in four broad priority and discovery-burden categories displayed in a heat map. A database table contains the custodian prioritizing, data-source burden assessments, and projected discovery costs for designated custodians and key data sources.

Because classifications and projecting costs are highly dependent on individual circumstances, the New Framework contains multiple appendices, matrices, and worksheets, which list pertinent factors and variables to provide guidance on how to classify custodians and project costs systematically. And where possible, the New Framework suggests baseline reference points for

standard-unit costs and degrees of the burden for reasonable discovery under typical circumstances, which can be adjusted for individual circumstances. The reference points are intended to facilitate the classifications and projections, especially for those less familiar with ESI and proportionality.

Proportionality assessments made under the New Framework can be adjusted periodically to account for evolving and new information learned through interactions and negotiations with opposing counsel. The New Framework is designed to be universal and to apply to all cases large and small, including complex commercial litigation (B2B), single plaintiff, class action, and investigations/trade secrets matters.

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

III. Concluding Thoughts

The New Framework's model set of data-source rankings and model set of standard-unit cost projections provide guidance. They are based on a series of assumptions and "suggest baseline-reference points and ranges" for reasonable discovery in a "typical case." They represent the collective judgment and experiences of the New Framework's lawyers, consultants, and other ediscovery experts and are grounded in numerous cost studies and surveys (see appendices).

But the results of the New Framework's model sets are subject to individual circumstances. Accordingly, the underlying worksheets and calculators used to create the model sets are explained and appended to allow adjustments to account for variances in individual cases.

Section 1: Custodians

I. Introduction

The New Framework establishes "a standard and cogent approach to frame proportionality assessments." Section 1 provides guidance on how to group custodians in four broad categories, from highest to low priority. Section 2 provides guidance on how to identify the eight most common data sources and rank them by the degree of burden in accessing information. The results of the custodian prioritizing and ranking of data sources "frame the proportionality assessments" by displaying the results 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.

Section 1 sets out objective criteria used for guidance to prioritize custodians, but in the end, all the decisions represent judgment calls *by those making them*, subject to challenge by others and subject to adjustment as the litigation progresses and new information is learned. Though not conclusive, early identification of many [most] of the high-priority custodians under the New

Framework is often apparent and without dispute. The initial results may fully satisfy the discovery objectives, or more likely, sharpen additional probing.

II. Identification of Relevant Information and Custodians

The New Framework does not create a new process to identify the scope of relevant matter or to identify custodians or data sources that are relevant. Lawyers typically take routine initial discovery steps at the outset of litigation to identify relevant information, its sources, and its custodians. The New Framework comes into play to prioritize custodians only after the relevant information, custodians, and data sources have been identified under traditional means. But a brief summary of the steps commonly taken to identify relevant information and custodians is needed to place 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. But there are common documents 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 the relevant matter and their custodians. Again, nothing new here, and many different approaches can succeed.[1] But 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 custodial or non-custodial.

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[2]). Interviews, written requests, or data sampling are techniques often used to further the investigation.[3]

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 and any key event timelines and temporal scope of preservation efforts need to be considered as well.[4]

III. Prioritizing Relevant Information

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

A. Materiality of Information

Relevant information may be more useful, significant, or important than other relevant information. The New Framework's first criteria focuses on "materiality," which is defined as significant or essential, to discern qualitative differences in relevant information.[5] 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;[6] and,
- is an essential link in a line of evidence needed to prove an assertion.[7]

B. Strength of the Information

The weight of relevant information in proving an assertion will vary. The New Framework's second criteria focuses on the "strength" of the information to distinguish the weight of relevant information based on how directly it is connected to the asserted fact.

Although considerations of materiality will overlap, the strength of the information may be indicated by:

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

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

C. Uniqueness of the Information

In a certain sense, every document and piece of information is unique.[8] The New Framework's third criteria characterizes uniqueness by distinguishing similar information from qualitatively different information. (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. And identifying unique information is a judgment call by those making the decisions,[9] 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 caselaw matures, the evolving standard will become clearer and provide more guidance.

IV. Prioritizing Custodians and Non-Custodian Data Sources

Under the New Framework, custodians and non-custodian data sources are prioritized by the level of relevant-priority information they possess or control. Along with the relevant-priority information, a custodian's position, level of knowledge, and depth of involvement in the particular issues must also be considered 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.[10] Whether the custodian has personal first-hand knowledge or whether the information is second-hand knowledge or comes from a third-party source are factors to account for ("track" or "consider").

A. Standardized Report Format

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

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

Written requests in the form of a survey, interviews, or data sampling are techniques often used to gather the information for the report. The purpose of the report is to organize the investigative results and provide a master score for each custodian based on the value of the 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, which will inform the scoring of custodians, that can be used as a template for the report.

B. Atypical Use Cases

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

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

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

c. **International Custodians:** If international custodians are involved, local data privacy law must be considered. Appendix D provides an example scenario that takes privacy laws into account.

Section 02: Defining Data Source Burden and Effort

I. Introduction

The New Framework's "standard approach to frame proportionality assessments" requires the designation of data sources by the degree of burden. Section 2 provides guidance on how to rank eight of the most common data sources in four broad categories, based on the degree of burden incurred in accessing information. Five primary variables are identified, which directly affect the degree of burden for each data source. The specific degree of burden associated with each data source is not determined, which is dependent on circumstances.

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

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

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

II. Data Sources and Types of Data

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

- Collaboration / Messaging Systems (Slack, Teams)
- Computers / laptops
- Email Systems
- File shares (departmental and personal)
- Mobile devices
- Paper documents / physical evidence
- Social Media
- Structured data (HR, finance, marketing databases)

A. Specialized Data Sources

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

- Backup media
- Computer code
- Corporate telephone data
- Ephemeral data
- Geo-location, GPS, IP addresses
- Specialized/proprietary databases, systems, or programs.
- Website data

III. Variables Affecting Burden Assessments

Five major variables affect the burden assessment, including:

A. The location of and accessibility to the data

B. The availability or state of readiness for collection

C. The methods used to preserve and collect data

D. Any specialized resources needed to effectuate collection

E. Any legal or regulatory constraints (e.g., data privacy laws) that may impact collection

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

A. Access to and Location of the Data

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

For instance, if the data is centralized in a single server room or a single cloud instance, it is more readily accessible than if it is spread across multiple data centers or cloud-storage locations that are not interconnected. The geography of US-based and international locations 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 data source.

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, 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 utilized 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 near-line 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 a data source that would typically be considered easier to collect information 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 an increased burden of collection.

E. Legal and Regulatory Restrictions and Requirements

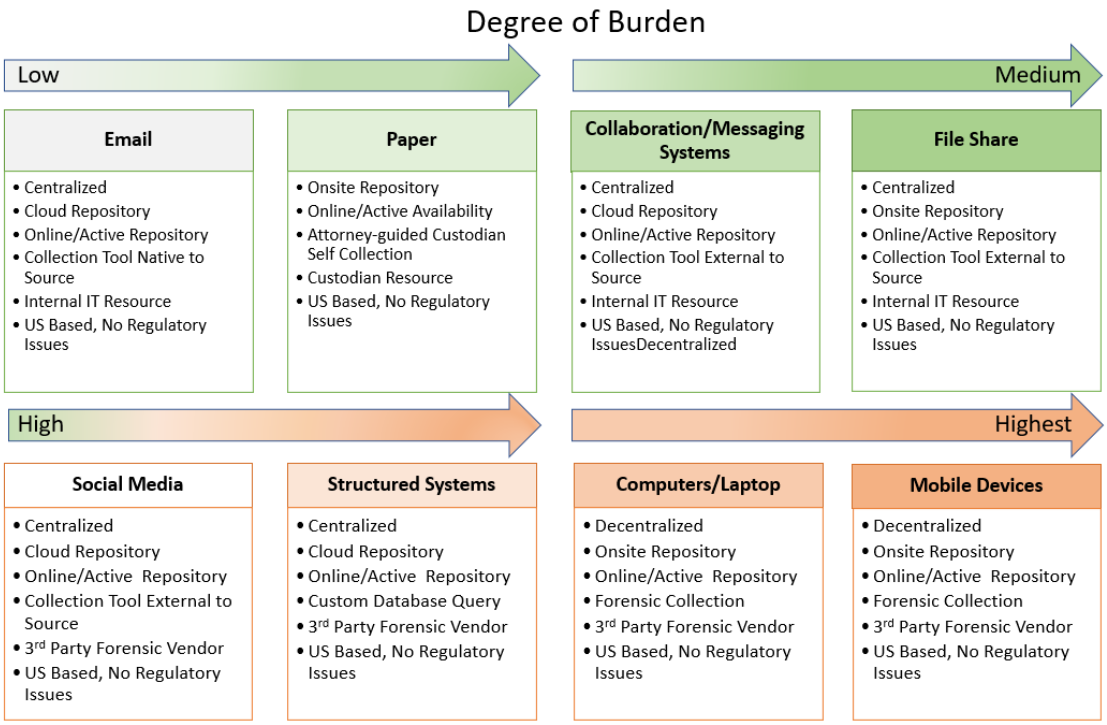
Legal requirements can also add to the burden of collection. Privacy legislation and regulations may regulate or require a party to take certain steps before the data is collected. The E.U. General Data Protection Regulation ("GDPR") applies to any processing of personal data when a European company is the controller. The GDPR has specific legal requirements for the processing of personal data (preservation, collection, processing, and production are separate data-processing activities). For instance, each data-processing activity needs a legal basis,[11] and you have obligations towards the individual.[12]

In addition, the transfer of personal data for use in litigation-discovery in the U.S. is only possible if certain conditions are met.[13] 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 results in, among others, additional

consultations and assessment steps involving local counsel or authorities, which can add to the burden of collection.

IV. Assigning Burden

The New Framework's model set of rankings of the eight most common data sources is based on variables that typically apply to provide baseline reference points, which are graphically displayed below. Accordingly, the low degree of burden associated with typical email is ranked at one end of the spectrum compared with messaging, which is ranked at the other end of the spectrum under the New Framework. Custodians with high-priority information at lowest discovery burden, are quickly identified.



Although this model set of rankings will apply in many, perhaps most, cases, the variables in an individual situation may be atypical and require adjustments, sometimes significant adjustments, to accurately reflect the actual burdens. Recognizing the possibilities that the five variables may be different for a particular data source in an individual case and require adjustments, a "Burden Assessment Calculator" has been developed to assist in leveraging the New Framework to account for the atypical variable and adjusting the degree of burden for the affected data source. The calculator 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.

The Burden Assessment Calculator is contained in Appendix E.

Section 03: Discovery Cost Projections

I. Introduction

In addition to the heat map of custodians and data-source burdens, a table of discovery costs projected for every custodian and data source is part of the New Framework's "standard approach to frame proportionality assessments." Section 3 guides how to project these costs, with multiple worksheets and calculators itemizing the individual costs.

Similar to the ranking of data sources by burden, estimating discovery costs must be done on a case-by-case basis relying on the itemized costs listed in the worksheets and calculators. [14] Studies have reported extremely wide ranges in per-gigabyte or per-document discovery costs, revealing tenfold to hundredfold differences in per-gigabyte or per-document costs. A model set of projected costs is set out at the end of this section, which provides baseline reference points.

II. Projecting Total Discovery Costs

The New Framework's discovery cost projections are organized around the discovery workflow found in the EDRM, a generally accepted model for ediscovery practice, including: Identification, Preservation, Collection, Processing, Review, Analysis, and Production. Similar to the approach taken by RAND in its seminal 2010 discovery-cost report, the New Framework collapses costs into three main categories: Collection, Processing & Hosting, and Review & Production. [15]

Although wide ranges of discovery costs have been reported, studies over the past decade have shown remarkable consistency in the overall comparative percentages attributed to these three stages. Ediscovery costs comprise approximately 68% for review, 19% for processing, and 13% for collection.[16] These proportions offer a useful cross-check verifying estimates.

A. Collection Costs

Section 1 describes generally the collection process of gathering the subset of potentially discoverable ESI. Preservation costs are included in this category.[17] Many different methodologies have been used to collect ESI, some more efficient, effective, and less expensive than others. Costs can be mitigated using a targeting or sampling strategy, the implementation of remote collection, or examination and elimination of redundancies before collection. Although the methodologies vary, collection costs typically account for approximately 13% 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 the hiring of external specialists.

Rates for collection vary depending upon the methodologies required. But the hourly or per-device rates for data collection have remained relatively flat over the period 2004-2020, even though data volumes, types of devices, and information stores have proliferated. During the same period, the efficiency of collection tools and the sophistication of the individuals involved

in the process (e.g., IT professionals, attorneys) have improved, which has effectively offset the proliferation of information.

B. Processing Costs

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

Collected data from different sources is typically processed by "ingesting" it, using a specialized tool designed to normalize data formats and optimize the data for the various search, analytic, and culling functions to follow.[18] The resulting data set is then prepared for and migrated to a review tool that allows either linear document review, advanced analytic options, or both. The processing stage typically accounts for approximately 19% of overall ESI project outlay.

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

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

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

C. Review and Production Costs

The review and production costs make up approximately 68% of a producing party's overall outlay. The principal driver of these costs is attorney-review time.

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

Generally, the overall cost of review turns largely on several considerations: (a) how much of the attorney time involves contract attorneys, as opposed to firm attorneys; (b) the cost of privilege review; (c) how long it takes attorneys to review the data; and (d) the volume of data or documents attorneys review or put eyes on.

The cost of review depends not only on the number of documents to be reviewed but also the amount of time it takes attorneys to review different categories of data (e.g., emails versus dense and complicated excel spreadsheets or engineering drawings). If the data has been normalized in the processing stage (i.e., text messages, emails, and Slack conversations look the same on the review platform), it can reduce review time.

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

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

III. Model of Projected Costs Table

The New Framework provides a "model set" of projected discovery-cost ranges expressed in standard-units, e.g., $$ per gigabyte, $$ per document, for a "typical" case based on costs that are often experienced and typically expected in most litigation. The model set of cost projections are based on a literature review, standard industry-wide assumptions, and the collective judgment and experiences of the New Framework's experts, applying the itemized costs to individual data sources as they most commonly appear in a "typical" case. [22]

The model set of cost projections are intended to provide guidance and reference points, especially to those unfamiliar with ESI, when projecting discovery costs. A graphic table containing the "model set" of projected costs is set out below.

But the use of the model set of cost projections has major limitations because it is subject to individual circumstances. To account for these circumstances, an itemized cost-projector calculator is contained in Appendix F, which permits the calculation of adjustments. Appendix G lists common variances that will require adjustments to the model set of projected costs.

(To be filled with standard-unit costs)

ESI STAGE	Hosted Email	File Shares	Computers/ Laptops	Collaboration/ Messaging Systems	Mobile Devices	Social Media	Structured Systems	Paper
Collection								
Processing								
Review & Production								

IV. Individualized Itemized Cost-Projection Calculators

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

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

V. Common Variances Requiring Adjustments

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

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

I. Introduction

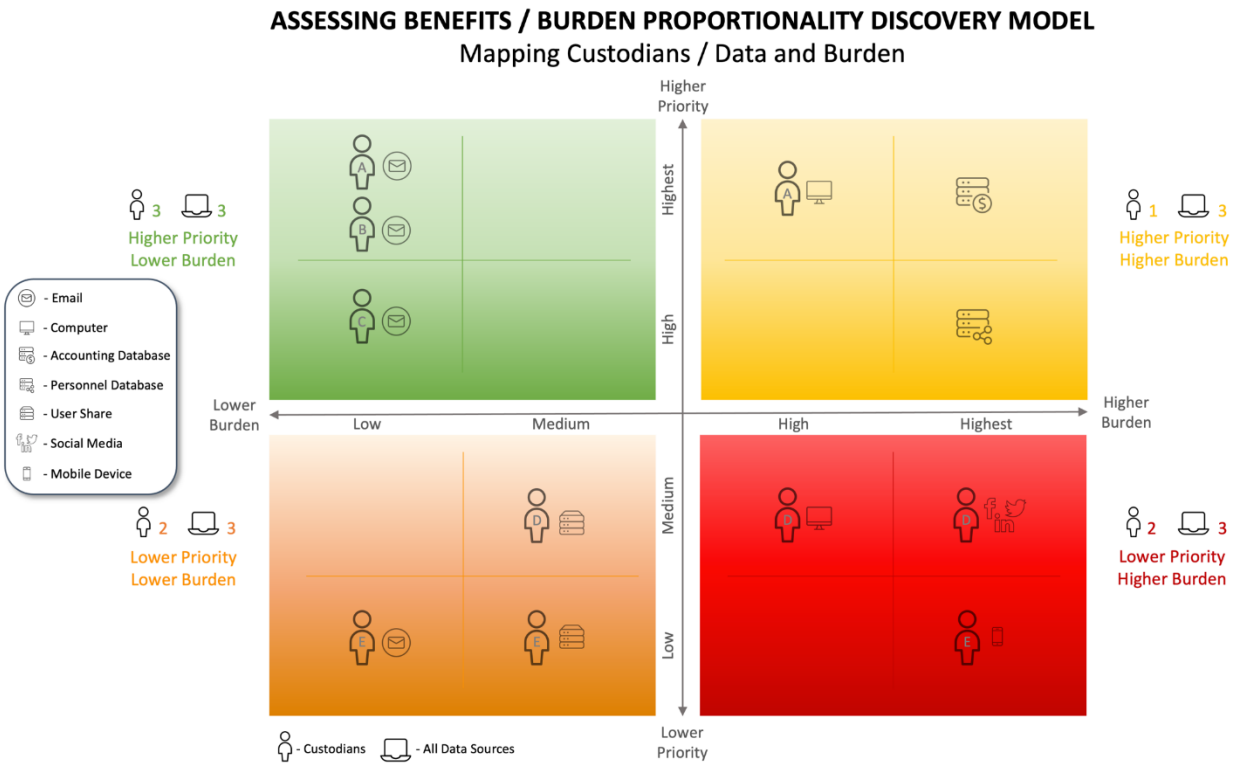
Section 4 pulls together the custodian and burden assessments and cost projections, displaying them in two formats: (a) a heat map, which organizes the custodians and their data sources on a sliding priority and burden scale and plots them in four quadrants; and (b) a database table, which sets out the assessments and adds ranges of cost projections for every custodian and data source. The heat map and database table are the outputs of the New Framework's "standard and cogent approach to frame proportionality assessments."

The weight accorded the assessments in the New Frameworks' heat map and the cost projections will be affected by the extent of the opposing party's input, if any. The more input from the opposing party, the less likely discovery disputes will occur and, if a dispute does arise, the judge will be better informed to resolve it.

II. New Framework Heatmap

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

The illustration below is an exemplar.



III. New Framework Database Table

The New Framework database table contains the data results from prioritizing custodians in Section 1, assessing data-source burden in Section 2, and projecting costs in Section 3. The table is an exemplar.

New Framework Database Table (Exemplar)

Name	Priority	Data Source		Level of Burden	Cost Ranges		
Custodian "A"	Highest		Email	Low	$	$$	$$$
			Computer	High	$	$$	$$$
Custodian "B"	Highest		Email	Low	$	$$	$$$
Custodian "C"	High		Email	Low	$	$$	$$$
Custodian "D"	Medium		User Share	Medium	$	$$	$$$
			Computer	High	$	$$	$$$
			Social Media	Highest	$	$$	$$$
Custodian "E"	Low		Email	Low	$	$$	$$$
			User Share	Medium	$	$$	$$$
			Mobile Device	Highest	$	$$	$$$
Non-custodian	Highest		Accounting Database	Highest	$	$$	$$$

Non-custodian	High		Personnel Database	Highest	$	$$	$$$

IV. Applying the New Framework at Different Stages of Litigation

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

A. Commencement of Litigation

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

The degree of confidence in these early assessments and cost projections can be affected by the extent of the opposing party's input, if any, into the development of the New Framework's heat map and database table.

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

- The database table's cost projections will provide information necessary to plan a budget covering likely discovery expenses.
- The database cost projections as well as the calculator accounting for variances provide baseline reference points to compare and evaluate RFPs, which can be particularly helpful for those less experienced in ESI.
- The information about priority custodians in the heat map will better inform the decision to make early productions in accordance with Rule 26(d)(2) and initial disclosures under Rule 26(a) by identifying obviously significant custodians which will assuredly result inevitably in discovery productions.[23]
- 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. [24]
- 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.[25]

B. Discovery Management

As additional information becomes available during discovery, including information learned from interactions with the opposing party, the assessments and cost projections can be adjusted and become firmer.[26] The extent of the opposing party's input, if any, in any adjustments to heat map and database table will affect the degree of confidence the opposing party has in the accuracy of the assessments and cost projections.

The New Frameworks' heat map and database table distill and concentrate key information in an organized format, which is necessary to make proportionality assessments. These two sources can ground the proportionality decision-making. For example:

- The granular information on data-source burdens and costs for every custodian and their respective data sources will better inform negotiations and decision-making at the Rule 26 meet-and-confer meeting as well as a later Rule 16 pretrial conference.[27]
- The identification of data sources and attendant burdens will help the parties better craft an ESI protocol, which considers all pertinent information.
- The same information can help the parties develop a phased-discovery plan, when appropriate, identifying which collections and custodians should go first and which should follow sequentially.[28]
- 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, are better informed when considered with all other pertinent information shown in the heat map and database table.

C. Close of Discovery

The New Framework's heat map and database table can be used to address proportionality-related issues, which may arise at the end of the discovery stage.

- The decision when to release litigation holds is better informed by the New Framework's heat map, which highlights individual custodians most likely to possess information that can be released.[29]
- Information in the New Framework's heat map and database table better informs decisions to reopen discovery.[30].

The New Frameworks' workflow results in the heat map and database table and provides a record documenting the decisions and actions occurring throughout the litigation lifecycle. The documentation is essential from a standpoint of establishing defensibility of process.

V. Judicial Resolution of Discovery Disputes

The information in the New Framework's heat map and database table provides a judge with essential information in an organized format to evaluate whether counsel's discovery efforts are reasonable and in good-faith and whether discovery is proportional to the needs of the case, subject to consideration of all the Rule 26(b)(1) factors. The extent to which opposing counsel provided input into the New Framework's heat map and database table will be a factor that a judge will consider when evaluating the weight to be accorded the proportionality assessments.

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

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

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

- A judge can consider the information to evaluate the scope of a requested preservation order. Custodians possessing marginally significant information in burdensome data sources are identified and can be scrutinized, which considerably narrows the preservation analysis.[31]
- 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 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.[32]
- 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 database table provides concrete information on the costs projected for each additional custodian, and most importantly, the cost is given by data source, providing the judge the capability to make more precise rulings that consider every data source.[33]
- A judge can use the information to evaluate and resolve sanction motions. Reasonableness is the overarching issue for a judge to consider when evaluating counsel's discovery actions. The New Framework's heat map and database table provide a record of all custodians, data sources, expenses, and burdens that counsel was faced with in making proportionality assessments and taking action. The record's full accounting better informs the judge's decision on the reasonableness of any specific discovery action taken by counsel.[34] The judge may also consider the extent that opposing counsel provided input, if any, to the development of the custodian prioritizing and data-source burden assessments and cost projections, in determining whether counsel's action was reasonable.

Appendix A

Survey Questions to Assess Relevant Information and Custodians

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

1. Create a few simple qualifying questions to filter immediate disqualification, i.e., employment outside date range.
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 and is ranked as follows:
    a. Highest
    b. High
    c. Medium
    d. Low

Exemplar

The following is an example of an assessment survey related to a product liability matter. 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.

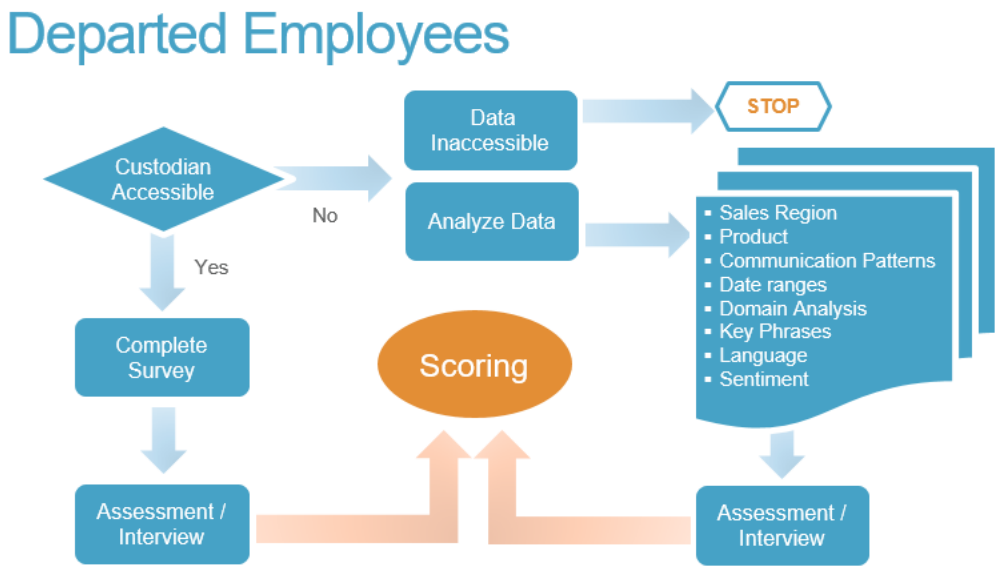
Custodian: Jim Dunbar				
Question	**Priority Level**			
	Low	**Medium**	**High**	**Highest**
Qualifying Questions:				

What period of time where you employed by Company X?	Outside relevant time period			Inside relevant time period
What position(s) do you hold and what were your responsibilities?	Irrelevant position or duties			Relevant position and/or duties
What departments did you work within?	Irrelevant department			Relevant department
Where are you physically located?		US based		International
Assessment Questions:				
Were you involved in internal or external meetings or discussions related to Product X?	Never	Sometimes	Regularly	Daily
Is the information you possess or control material?	Not material	Modestly Material	Material	Extremely Material
Is your relevant information directly connected to the claims in the case?	Not Connected	Tangentially Connected	Connected	Connected
Is your relevant information unique and not possessed or controlled by others?	No	Yes	Yes	Yes
Did you create or participate in any internal email discussions or were copied on email threads related to Product X?	Never	Sometimes	Regularly	Daily
Did you create or contribute to any documentation, presentations, design or product materials or manuals related to Product X?	Never	Sometimes	Regularly	Daily
Did you directly or indirectly communicate with anyone outside the organization on any topics related to Product X?	Never	Sometimes	Regularly	Daily
Were you aware of or participate in any social media discussions regarding Product X?	Never	Sometimes	Regularly	Daily

The responses to the interview questions can be ranked in any suitable manner from low to highest and then aggregated to create the master score.

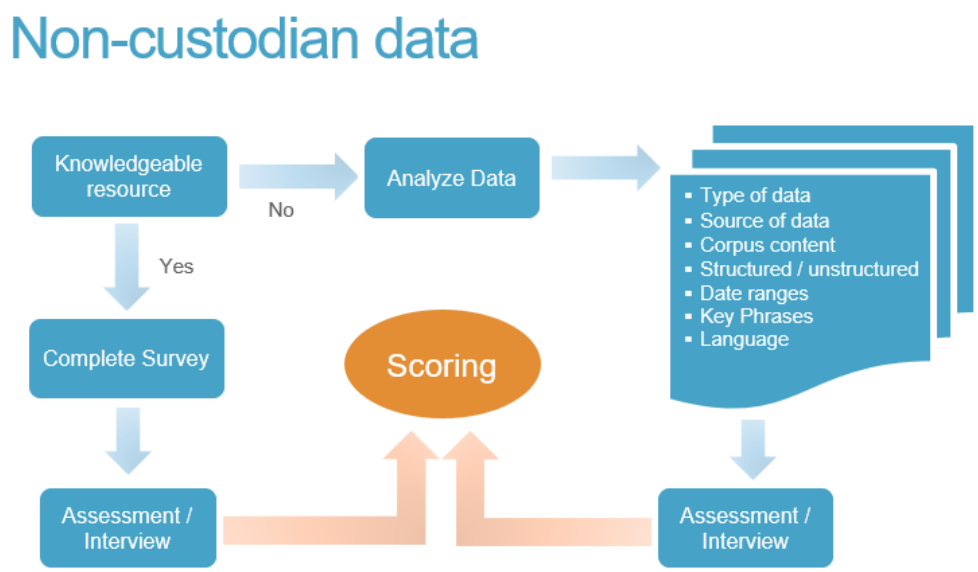
Appendix B

Atypical Case – Departed Employee



Appendix C

Atypical Case – Non-Custodian Data



Appendix D

Atypical Case – Privacy Laws

For example: If EEA custodians are involved, GDPR applies and must be considered. Art. 6 GDPR defines the lawfulness of processing (preservation / collection / review). In litigation cases, Art. 6(1)(f) of GDPR is the correct legal basis (necessary for the purposes of the legitimate interests pursued by the controller). Another legal basis could be the consent of the custodian (Art. 6(1)(a) GDPR). However, consent must be given freely, which is difficult in any employer-employee relationship. In addition, consent can be withdrawn at any time. Art. 6(1)(c) GDPR is not applicable as the term "legal obligation" means an obligation that arises out of EU law or the law of an EU Member State to which the controller is subject. If custodian or data source has no materiality, strength and uniqueness to the claims and defenses, the legal basis according to GDPR is missing. Therefore, custodian / data source cannot be preserved / collected / reviewed, and the custodian must be released from the preservation obligation. The assessment (questions, answers and scoring) should be kept as record to show the reasons for the determination of releasing the custodian.

APPENDIX E

Burden Assessment Calculator

The Burden Assessment Tool created by the team allows you to assess each of your data sources based on the criteria outlined in the narrative above. Once you select the appropriate factors for your situation, the tool will automatically assign a classification of low, medium, high, highest burden to your data source.

The tool can be downloaded here [link to excel document on GW website]

APPENDIX F

Itemized Cost-Projection Calculators

The Cost Projection Calculator developed by the Team provides itemized cost assumption for each step within each stage of ESI, from collection through Production. We provide a low, medium, high estimate based on our current, best understanding of the going rates in the market. Your assumptions or actual costs and the calculator allows you to apply your own assumptions to calculate your specific project costs.

The calculator can be downloaded here [link to excel document on GW website]

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 Data) - 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 Data Systems	Special expertise and skills may be needed to gather, compile, and render ESI for structured data sources. Review costs may be mitigated because that attorneys are reviewing data compilations and summary reports. Although higher cost, non-attorney reviewers are often needed (e.g., accountants, engineers, nurses).

ENDNOTES

---

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

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

[2] A "Data Steward" is an individual who maintains information but is not themselves a fact witness or source of potentially relevant information

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

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

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

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

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

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

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

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

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

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

[11] Art. 6 GDPR

[12] Art.13 GDPR

[13] Art. 49 GDPR

[14] The model presumes that the responding party develops the table of costs at the outset of litigation before actual costs have been incurred. Discussion of whether to ask for opposing side's input at various stages of litigation will be discussed in section 4.

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

[16] 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/

[17] NOTE: Our Framework is designed to forecast from Collection through Production. While a key step in the process, Preservation costs vary so widely based on individual circumstances, the costs associated with it are not contemplated in the Cost Projection Calculator.

[18] The objective is to prepare the data for document review through accurate and consistent culling and minimizing the quantity of files by excluding files with little or no evidentiary value (e.g., non-business files, duplicate files, files that are not germane to the claims or defenses of a case).

[19] In addition to the commonly accepted practices of the use of keywords, there are tools that permit further reduction at the review phase. One of these is email threading, which groups together different fragments of longer email conversations. This permits attorneys to review only the most complete versions of those conversations (sometimes called the "inclusive") and avoid having to review the fragments (sometimes called "non-inclusive").

Another set of tools, which some refer to generally as technology assisted review, or TAR, provide additional opportunities to reduce the number of documents that need human review. Most versions of TAR combine machine learning and human review by a subject matter expert that helps producing parties prioritize documents for review. Using statistical validation protocols and other methods, TAR also can be used to demonstrate when additional review would be unlikely to bring back additional responsive documents.

[20] The Sedona Conference has defined TAR as a "process for prioritizing or coding a collection of electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection." The Sedona Conference, *The Sedona Conference Glossary: E-Discovery and Digital Information Management*, Fourth Edition, 15 SEDONA CONF. J. 305 (2014) (definition adopted from Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola, U.S. Magistrate Judge*, 7 FED. CTS. L. REV. 1, 32 (2013)). The terms "predictive coding" and "computer assisted review" are often used interchangeably with TAR, to describe this process. TAR can also involve what is called "continuous active learning," or CAL.

[21] Moreover, the cost of privilege review and privilege logs may be reduced as parties take advantage of Rule 502 of the Federal Rules of Evidence and similar rules protecting parties against waiver.

[22] 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.

[23] 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.

[24] The parties may use the New Framework's heat map and database 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

[32] 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.

[33] The New Framework is used to help a judge understand where to draw the line in enforcing or quashing a request for production. A judge can easily balance the cost presented to the amount in controversy.

[34] Rule 37(e) notes that if electronically stored information is not properly preserved, a judge may issue sanctions. By finding those custodians and data sources that are relevant to the matter and ranking them according to priority in the information they could provide, the Framework can be a guideline for determining whether spoliation has occurred, and sanctions might be deemed necessary.