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.