Reducing the Cost of Electronic Data Discovery Reviews
Reducing the Cost of Electronic Data Discovery Reviews

Analysts estimate that corporations spend more than $10 billion annually to review electronic data produced in discovery. With current daily email traffic at 100 billion emails and growing, these costs stand to increase over time. Astute corporations and their counsel now leverage technology to cut discovery review costs by 50% or more.

Discovery Review Cost Drivers

The total cost of any given review comes down to a simple equation:

\[
\text{Hourly Rate} \times \text{Number Of Hours} = \text{Cost Of Attorney Review}.
\]

Hourly rates are a function of which providers the corporation has selected, which this paper treats as a constant.¹ This means that the primary opportunity for reducing costs lies in reducing the number of hours spent on the review. For the vast majority of cases, improving the efficiency of the linear review process provides the biggest opportunity for cost reduction. For the small subset of a corporation's cases that involve data sets of more than 250 gigabytes, specialized tools can be effective in reducing and organizing documents prior to the commencement of the review, generating a positive return on investment.

Improving Review Team Productivity

In every case, there is a point where lawyers and/or paralegals must review individual documents. Typically, the law firm trying the case selects a database or repository to house the documents. By standardizing on a single, central repository, corporations and their counsel can address a variety of seemingly mundane issues that, all together, sap the efficiency of their review teams. The following five areas merit special attention.

¹ Corporations have broad choices in who they select as providers. Technology can help to expand these choices by providing secure, web-based access and other tools that facilitate assigning part of the review process to lower-priced contract lawyers (on-shore or off-shore) and reserving the trial team for the review of more critical pieces.
1 Complexity: The Inhibitor

Many review platforms offer a variety of features that look great on paper but never get used because the product is too complex. This complexity can occur at multiple levels, including:

- An unintuitive interface
- Features that require a manual to locate and/or use
- Difficulty accessing the system, either because it must be accessed via the network or using plug-ins that have to be loaded on individual reviewers’ computers

In the real world, the greatest efficiencies often come from the systems that are the easiest to use. This starts with a review platform that employs a familiar and simple interface, with features that are easy to locate and intuitive to operate. Secure, web-based platforms also increase efficiencies because they don’t require plug-ins on reviewers’ machines or IT support.

2 System Lag: The Hidden Tax

Many review platforms were designed for much smaller data sets than occur in today’s typical case. Because of the sheer bulk of documents, these systems often get overloaded, with reviewers sitting idle while waiting for search results or waiting for the system to move from page-to-page or folder-to-folder. While standalone occurrences are seemingly inconsequential, system delays regularly occur thousands of times during a review, resulting in significant additional costs. For example, if a review team experiences 5 minutes of system lag each hour, review costs have just increased by over 8%.

Make sure the software you choose is designed to handle the volume of documents you need it to—and plan ahead for future growth.
3 Poor Organization: The Source of Chaos

Studies show that reviewing similar documents together can speed the review by ten times. Grouping the documents also allows potentially privileged or high-impact documents to be reviewed by senior reviewers while associates can sift through lower-level data.

-Law.com

Review efficiency suffers dramatically when the review is conducted without an overall organizational plan. Assigning documents to reviewers in a random fashion leads to documents being reviewed one-by-one, which is highly time consuming. It also can lead to rework if a document is assigned to a reviewer who is not properly qualified.

A strong organizational plan centers on sorting documents based upon logical characteristics, filing them into folders, and then assigning these folders to reviewers in a targeted fashion. This enables reviewers to develop familiarity with the common subject matter and apply judgments with greater speed and consistency.

The organizational plan often may have several stages. For example, in the first stage, documents may be sorted by custodian and assigned out to contract attorneys for a first-pass review. Then, they may be resorted based upon “issue” or “privilege” for subsequent passes.

Executing on an organizational plan requires certain features from the review platform. This starts with the ability to search large volumes of documents by key words and fields. In more complex cases with lots of documents, computer-aided grouping tools may also prove useful. Adding “batch tagging” capabilities, which enable reviewers to mark entire groupings of documents all at once, without having to review each individual document, can yield significant additional efficiencies.

4 Unnecessary Manual Steps: The Brake

Document review is all about identifying whether a document fits a certain set of characteristics: Does it relate to a certain set of issues? Is it privileged? Does it fit within the parameters of the opposing party’s request? Indicating which of these characteristics a document fits, and why, can be a highly manual process.
According to one estimate, 70% of all corporate email messages and documents (collected in discovery) are duplicates.

- Stephanie Sabatini, The Dilemma of Duplicates, Law Technology News

Opt for a review platform that enables you to create reviews tailored to the specific characteristics required for the case at hand. This includes reviewer form profiles (1st tier reviewers get three fields, 2nd tier reviewers get ten fields, etc.), check boxes for case-specific data entry points (issue: pain & suffering, etc.) and automated tools for common tasks like redaction.

5 Poor Quality: The Killer

Nothing kills efficiency like having to redo work because of poor quality. With reviewers often working long hours, at a somewhat mundane task, under tight deadlines, and with redundant or irrelevant documents to sift through, the conditions are ripe for quality issues.

Perhaps the most common problem occurs when reviewers either fail to enter complete information or when two reviewers manually classify similar documents differently. Mandatory fields that require the reviewer to complete them before moving on to the next document decrease the risk of incomplete information. Having fields appear in the form of check boxes next to pre-defined classifications reduces the risk of inconsistent classifications. For example, a well-structured review platform will require a reviewer who designates a document as privileged to check a box specifying whether the privilege claim is one of Attorney-Client or Work Product.

Quality issues are a fact of life and savvy reviewers plan for them. This means using a review platform that includes audit trails that let you know what stage the review is in and enable you to follow up quickly on any potential quality issues.

Special Considerations in Large Cases

Discovery starts in every case with the collection of data. The data set then gets reduced - often by as much as 50%-75% - by eliminating duplicate documents, running key term searches (custodians, dates, issues, etc.), and similar techniques.

For the majority of cases which involve 10 gigabytes of data or less, simple techniques suffice to reduce data sets to manageable levels and to organize the
remaining data logically. However, a growing percentage of cases and regulatory reviews involve hundreds of gigabytes or even terabytes of data.

In these large matters, it often makes sense to incur the extra cost (which can be substantial) of an additional round of culling and organization using specialized technologies. The two most common ways for doing this are specialized searches and software that automatically groups documents by concept.

1 **Specialized Key Term Filtering**

Specialized key term filtering involves taking the reduced document data set and running more refined key term queries against it to further cull documents and to organize them in an even more refined manner. These searches may be crafted either by lawyers or by professional linguists and can run to hundreds of terms, including text and fields (e.g., date ranges, authors, etc.).

Running these searches effectively requires a platform that is designed to handle large, complex queries and massive data sets. In addition, searches can be made even more effective by using filters that take results and map them against key fields like author, recipient, date range, etc. Filters enable you to organize documents into folders with a higher degree of precision. Irrelevant document sets can be excluded without human review; many relevant document sets can be produced without human review; and the remaining documents sets are organized in a fashion that helps teams review them more efficiently.

2 **Concept Searching and Clustering**

Concept searching involves using specialized software programs to identify “concepts” within a document set. These programs look for relationships between words within the documents and can reveal links between various terms that are valid in the case in question but may not appear intuitively to someone not familiar with the producing party’s affairs.

For example, depending upon context “Java” could refer to coffee, software or an island in Indonesia. Concept searching can be used on a stand-alone basis or as a tool to assist in building stronger queries for key term filtering.
With concept clustering, the software identifies the “concepts” within a document set and files the documents into concept folders. Documents relating to relevant concepts can be assigned for further review and documents relating to irrelevant concepts can be discarded. If the size of the case justifies the cost of these tools, they can be effective.

“Electronic discovery was the number one new litigation-related burden for companies with revenues over $100 million.”

—Second Annual Litigation Trends Survey Findings at 2, Fulbright and Jaworski (2005)

About Catalyst

Catalyst builds and hosts software that allows cross-enterprise teams to review massive data sets and manage complex legal matters. Our flagship product, Catalyst CR, is an integrated, web-based review platform supporting the full range of cases for corporations and law firms. Its interface is modeled after Outlook, so review teams can get up to speed and be productive quickly.

Built on a grid of servers and running the powerful FAST search engine, CR is designed to handle terabytes of data in over 70 different languages. In certified testing, CR routinely delivered search results in seconds against 15 million documents from complex searches of up to 750 terms—fields, text and ranges. CR also offers concept search and clustering.

Catalyst CR is used by many of the world’s leading corporations and law firms. Contact sales@catalystsecure.com to learn more.

About the Author

Bill Mooz is the General Counsel and Vice President of Strategy for Catalyst Repository Systems. Bill came to Catalyst from Sun Microsystems, where he served as Senior Director in charge of Sun’s Utility Computing business and as Deputy General Counsel. Previously, Bill was a partner in Holland & Hart, LLP, an AMLAW 200 law firm. He received his JD with honors from the University of Colorado and clerked for the Honorable Stephen F. Williams on the U.S. Court of Appeals for the D.C. Circuit.