

The Cost of Producing Electronic Documents in Civil Lawsuits

Can It Be Sharply Reduced Without Sacrificing Quality?

RAND RESEARCH AREAS

CHILDREN AND FAMILIES
 EDUCATION AND THE ARTS
 ENERGY AND ENVIRONMENT
 HEALTH AND HEALTH CARE
 INFRASTRUCTURE AND TRANSPORTATION
 INTERNATIONAL AFFAIRS
 LAW AND BUSINESS
 NATIONAL SECURITY
 POPULATION AND AGING
 PUBLIC SAFETY
 SCIENCE AND TECHNOLOGY
 TERRORISM AND HOMELAND SECURITY

The cost of producing electronic documents in a civil lawsuit has become a policy concern in recent years. Pretrial electronic discovery, or *e-discovery*, consists of essentially three activities: (1) *collecting* potential sources of information in response to an opposing party's demand, (2) *processing* that information to reduce its volume and organize it for review, and (3) *reviewing* processed digital information to identify relevant and responsive documents to produce and privileged documents or confidential information to withhold. With the explosive growth of digital information, many believe that the cost of these activities—particularly review—accounts for an increasing proportion of total litigation costs.

A RAND study assesses e-discovery costs and describes how they can be reduced without compromising the quality of the discovery process. It also examines what litigants perceive to be the key challenges of preserving electronic information for the purposes of litigation.

The researchers worked with eight very large companies in diverse industries that were willing, with assurances of confidentiality, to provide information about their discovery expenses in 57 large-volume cases. They also interviewed company representatives about their preservation duties and concerns.

Cost Breakdown

The authors found that the review process is by far the most expensive task in producing electronic documents, as shown in the figure. At least for the large-volume cases included in the study, review accounts for 73 percent of the total cost of producing documents, while processing the information accounts for 19 percent. The work itself is largely performed by other parties: outside counsel (typically for review) and vendors (typically for processing and collection). As the figure shows, only 4 percent is spent on

Abstract

Litigants worry that producing electronic documents relevant to a civil lawsuit is consuming a growing share of litigation costs. A RAND report examines this problem and recommends options for addressing it. According to the study, the review process accounts for 73 percent of the total costs of production, and predictive coding, in which computers do the bulk of the review now performed by attorneys, is a promising option for cutting costs without compromising the quality of the process. The study also calls for rule changes to address concerns about the scope and process of preserving data in anticipation of litigation.

the litigant's own in-house counsel, information technology staff, and other employees.

Reducing the Cost of Review

Because it makes up the greatest percentage of e-discovery costs, the researchers targeted the review process in their search for ways to reduce costs. They set an ambitious goal of reducing review costs by three-quarters to bring them in line with processing costs, assuming that doing so would reduce many of the complaints now voiced about the expense of e-discovery. The authors also summarize comparisons of the quality of traditional review practices that require eyes-on review of documents by attorneys with the quality of new approaches in which computers perform the bulk of the review process.

Current Approaches

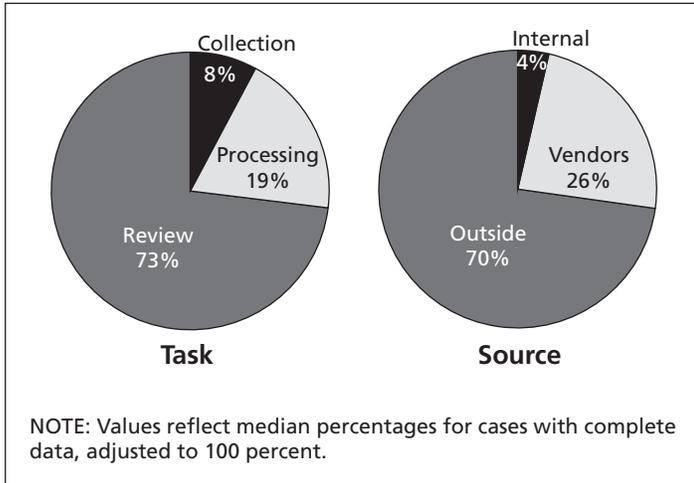
Companies are trying a variety of alternatives to the traditional use of outside law firms to reduce the cost of review, such as using lower-cost attor-

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of published, peer-reviewed documents.

Corporate Headquarters
 1776 Main Street
 P.O. Box 2138
 Santa Monica, California
 90407-2138
 TEL 310.393.0411
 FAX 310.393.4818

© RAND 2012

Relative Costs of Producing Electronic Documents



neys, increasing the rate of review, and grouping documents to streamline review. The study finds limitations in all of these approaches.

Labor Costs Can Be Lowered Only So Much

Litigants and their counsel are hiring temporary attorneys or using legal process outsourcing (LPO) companies with stables of contract attorneys. But the rates paid to such project attorneys for large-scale reviews in the United States may well have bottomed out, with further reductions of any significant size unlikely. Some companies have turned to LPOs with access to lawyers in other countries, such as India and the Philippines, but this practice may raise concerns about information security, oversight, maintaining attorney-client privilege, and logistics.

Review Speeds May Have Reached an Upper Bound

The highest estimates of review speed are about 100 documents per hour, and this number assumes that reviewers have the strongest motivations and experience and are examining documents simple enough that a decision on relevance, responsiveness, privilege, or confidentiality could be made in an average of 36 seconds. Given the trade-off between reading speed and comprehension, especially in light of the complexity of documents subject to discovery in large-scale litigation, it is unrealistic to expect much room for improvement in the rates of unassisted human review.

Grouping Documents Is Not the Answer

Litigants are also using computerized techniques to organize similar documents together:

- *Near-duplicate detection* groups together documents that contain mostly identical blocks of text or other information but differ in some minor way.

- *Clustering* identifies the keywords and concepts in each document, then groups documents by the degree to which they share keywords or concepts.
- *Email threading* groups individual emails into single conversations, sorting them chronologically and eliminating duplicate material.

In many cases, such techniques organize material rather than reducing the number of documents needing review. Unless they can be grouped to allow *bulk coding*—in which a decision on one is applied to the others in the same group—these techniques cannot make sufficiently dramatic improvements in review speed needed to minimize review costs as a source of concern.

Human Reviewers Are Highly Inconsistent

Existing research suggests that reviewers often disagree with one another when they review the same set of documents for relevance and responsiveness in large-scale reviews. In one case, for example, seven teams of attorneys, all trained in a similar manner and given the same instructions, examined 28,000 documents organized into 12,000 clusters to judge whether the clusters were responsive to the facts of the case. Results showed that the teams agreed only 23 to 54 percent of the time, depending on the pair of teams being compared.

These high levels of disagreement, confirmed in other studies, are caused by human error in applying the criteria for inclusion, rather than ambiguity in the documents or the scope of the production demand.

Predictive Coding May Be an Answer

Predictive coding is a type of computer-categorized review process that classifies documents in terms of how well they match the concepts and terms in sample documents. With predictive coding, attorneys initially examine samples of documents from the review set and make determinations about whether they are relevant, responsive, or privileged. The application then uses those decisions to assign scores to other documents in the review set. When attorneys assess samples of the computer's decisions, the application uses that input to increase the reliability of its decisions. The results of this iterative process are eventually stabilized, at which point the software's decisions should closely match those of human reviewers.

Because this is nascent technology, there are only a few studies that examine its accuracy or its costs compared with those of human review. Studies of its accuracy show that predictive coding identifies at least as many documents of interest as traditional eyes-on review, with about the same level of inconsistency as—and, in some cases, greater consistency than—human review.

Because evidence is scant and vendor pricing models are still in development, it is too early to know the exact level of

cost savings that could be achieved with predictive coding. But early studies suggest that the savings could be considerable. One such study examined a document set that had been reviewed by attorneys in the traditional way and reported that predictive coding would have saved about 80 percent in attorney review hours. A reduction of hours to this degree would likely meet the savings target of 75 percent in total review costs.

Challenges of Preservation

Another serious challenge presented by the growth in digital information is how much of it to preserve for the purposes of future litigation. Interviews with corporate counsel from the companies in the study highlighted several concerns:

- **Companies are not tracking preservation costs.** Most interviewees reported that their companies were not tracking costs and that they were unsure how to accomplish such tracking, although collecting useful metrics would be an important first step.
- **Those costs are said to be significant.** Attorneys interviewed from all companies reported that preservation costs had become a significant portion of their total e-discovery costs, with some believing that preserving documents now cost their companies more in the aggregate than producing documents.
- **Clear legal authority is needed.** Although judicial decisions have addressed preservation scope and process, the decisions serve as precedent only in specific jurisdictions and sometimes conflict with the decisions of other courts. As a result, attorneys reported that they had no clear understanding of whether their decisions about what to preserve—and how to preserve data—were legally defensible and not at risk of serious sanctions.

Recommendations

The authors propose three recommendations to address complaints of excessive costs of review and uncertainty about legal responsibilities for preservation.

Adopt computer categorization to reduce the costs of review in large-scale electronic production. Despite its promise, predictive coding has not yet been adopted by litigants, partly because there is a dearth of judicial decisions at the trial court level that clearly address the use of the technology. The authors propose that the best way to bring the technology into the mainstream is for public-spirited litigants to use it—and proclaim its use—for large-scale e-discovery efforts. If litigants were able to win judicial approvals in several jurisdictions, then use of these techniques would spread, the technology would improve, and long-term cost savings would result for the civil justice system as a whole.

Improve tracking of the costs of production and preservation. Without such data, companies cannot develop strategies for dealing with massive data volumes, such as investing in automated legal-hold-compliance systems or advanced analytic software for early case assessment. The rule-making process has been hampered by a lack of reliable cost information related to e-discovery.

Bring certainty to legal authority concerning preservation. Steps must be taken soon to address litigant concerns about complying with preservation duties. The absence of clear, unambiguous legal authority that applies across jurisdictions appears to be thwarting thoughtful preservation efforts and may be leading to overpreservation at considerable cost. Modifying the rules of court seems to be the most practical way to bring certainty to this area. ■



This research brief describes work done for the RAND Institute for Civil Justice documented in *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*, by Nicholas M. Pace and Laura Zakaras, MG-1208-ICJ, 2012, 158 pp., \$25, ISBN: 978-08330-6876-7 (available at <http://www.rand.org/pubs/monographs/MG1208.html>). This research brief was written by Laura Zakaras. The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.

RAND Offices

Santa Monica, CA • Washington, DC • Pittsburgh, PA • New Orleans, LA/Jackson, MS • Boston, MA • Doha, QA • Abu Dhabi, AE • Cambridge, UK • Brussels, BE



Institute for Civil Justice

A RAND LAW, BUSINESS, AND REGULATION INSTITUTE

CHILDREN AND FAMILIES
EDUCATION AND THE ARTS
ENERGY AND ENVIRONMENT
HEALTH AND HEALTH CARE
INFRASTRUCTURE AND
TRANSPORTATION
INTERNATIONAL AFFAIRS
LAW AND BUSINESS
NATIONAL SECURITY
POPULATION AND AGING
PUBLIC SAFETY
SCIENCE AND TECHNOLOGY
TERRORISM AND
HOMELAND SECURITY

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from www.rand.org as a public service of the RAND Corporation.

Support RAND

[Browse Reports & Bookstore](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore the [RAND Institute for Civil Justice](#)

View [document details](#)

Research Brief

This product is part of the RAND Corporation research brief series. RAND research briefs present policy-oriented summaries of individual published, peer-reviewed documents or of a body of published work.

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND website is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).