

9-22-2015

The Discovery Sombrero and Other Metaphors for Litigation

William H. J. Hubbard

Follow this and additional works at: <https://scholarship.law.edu/lawreview>



Part of the [Civil Procedure Commons](#), and the [Litigation Commons](#)

Recommended Citation

William H. Hubbard, *The Discovery Sombrero and Other Metaphors for Litigation*, 64 Cath. U. L. Rev. 867 (2015).

Available at: <https://scholarship.law.edu/lawreview/vol64/iss4/6>

This Article is brought to you for free and open access by CUA Law Scholarship Repository. It has been accepted for inclusion in Catholic University Law Review by an authorized editor of CUA Law Scholarship Repository. For more information, please contact edinger@law.edu.

The Discovery Sombrero and Other Metaphors for Litigation

Cover Page Footnote

Assistant Professor of Law, University of Chicago Law School. I am grateful for comments from Douglas Baird, Emily Buss, Steve Hagenbuch, Ashish Prasad, and participants in the faculty work-in-progress workshop at the University of Chicago Law School. I thank Matthew Ladew and Hangcheng (Robert) Zhou for valuable research assistance. I thank the Paul H. Leffman Fund and the Coase-Sandor Institute for Law & Economics for research support.

THE DISCOVERY SOMBRERO AND OTHER METAPHORS FOR LITIGATION

William H. J. Hubbard⁺

I. LAW GOVERNING PRESERVATION AND DISCOVERY	875
II. THE PRESERVATION COSTS SURVEY	878
A. <i>Obstacles to Empirical Work on Preservation</i>	878
B. <i>Survey Methodology</i>	880
C. <i>Sample Characteristics</i>	883
III. THREE STYLIZED FACTS ABOUT LITIGATION	885
A. <i>The Discovery Sombrero</i>	885
B. <i>The Preservation Iceberg</i>	889
C. <i>The Long Tail of Costs</i>	894
IV. THREE IMPLICATIONS FOR LAW AND RULEMAKING	897
A. <i>The Sombrero and the Erie/Hanna Boundary</i>	897
B. <i>The Iceberg and the Law/Technology Boundary</i>	902
C. <i>The Long Tail and Transsubstantivity/Tailoring Boundary</i>	903
V. CONCLUSION	906

The process of discovery in civil litigation is doubly shrouded in fog. Fundamentally, as the term “discovery” connotes, the discovery process involves parties who lack complete knowledge about their dispute attempting to use the litigation process to obtain information.¹ Almost by definition, the parties and the court operate in a fog of uncertainty when they undertake discovery. There is nothing necessarily troubling about this uncertainty, of course; discovery exists precisely to dispel it.

However, there is a second, more troubling layer of obscurity. We know very little about the timing, volume, and cost of discovery in our civil justice system. In what fraction of cases does the gathering of documents in anticipation of discovery begin before a lawsuit is even filed? How much data is gathered in the average case? Setting aside the fees paid to outside counsel, how much does

⁺ Assistant Professor of Law, University of Chicago Law School. I am grateful for comments from Douglas Baird, Emily Buss, Steve Hagenbuch, Ashish Prasad, and participants in the faculty work-in-progress workshop at the University of Chicago Law School. I thank Matthew Ladew and Hangcheng (Robert) Zhou for valuable research assistance. I thank the Paul H. Leffman Fund and the Coase-Sandor Institute for Law & Economics for research support.

1. *But see* Frank H. Easterbrook, *Discovery As Abuse*, 69 B.U. L. REV. 635, 636–37 (1989) (noting that the threat of discovery is sometimes used as a bargaining tool to exact a favorable settlement rather than as a purely information gathering tool).

discovery cost the parties, in terms of time and money, in any given case? It is scarcely an exaggeration to say no one knows.

While many practicing attorneys have rich and detailed knowledge of their own experiences, commentators have struggled to collect and organize this anecdotal information into a coherent empirical picture. To this day there is no consensus on how much litigation costs in a typical case. Reputable sources provide numbers that may seem surprisingly low—for example, \$20,000 for a single party²—or surprisingly high, to the tune of millions of dollars.³ As another example of the uncertainty, there is anecdotal evidence that many companies fear spoliation sanctions arising out of unclear preservation obligations; yet there is also evidence that the imposition of sanctions is rare.⁴ As a recent report has noted, the “actual costs of discovery have rarely been quantified in empirical studies.”⁵

This collective ignorance of judges, policymakers, and academics feeds uncertainty at both the policy and the doctrinal level. Policymaking, in the sense of rules design, is hamstrung by a lack of information about the activities that are the subject of the rules. While there is no shortage of anecdotes decrying excessive costs and burdens of discovery (usually from the defense bar) and alarm about stonewalling and evidence destruction (usually from the plaintiffs’ bar),⁶ it is hard to judge the extent of these problems or what, if anything, should be done about them.

Ignorance of how discovery tends to play out in practice leads to confusion even at a doctrinal level. The federal courts appear ambivalent about how to address perceived problems with discovery, despite discovery being the subject

2. A Federal Judicial Center study reports that the median discovery costs for defendants in civil cases in federal court are \$20,000. Emery G. Lee III & Thomas E. Willging, *National, Case-Based Civil Rules Survey*, FED. JUDICIAL CTR. 2 (2009) [hereinafter *Civil Rules Survey*].

3. A study by the Institute for the Advancement of the American Legal System (IAALS) estimates discovery costs of \$3.5 million for a “midsize” case. INST. FOR THE ADVANCEMENT OF THE AM. LEGAL SYS., *ELECTRONIC DISCOVERY: A VIEW FROM THE FRONT LINES* 4 (2008).

4. EMERY G. LEE III, *MOTIONS FOR SANCTIONS BASED UPON SPOILIATION OF EVIDENCE IN CIVIL CASES: REPORT TO THE JUDICIAL CONFERENCE ADVISORY COMMITTEE ON CIVIL RULES* 3–5 (FED. JUDICIAL CTR. 2011).

5. Nicholas M. Pace & Laura Zakaras, *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*, RAND INSTITUTE FOR CIVIL JUSTICE 4 (2012) [hereinafter *Where the Money Goes*]. See also *id.* at 3 (“A repeated lament in the academic and legal literature is that there has been little or no research into the costs imposed on the larger judicial system by the discovery process.”) (internal quotation marks omitted). The reasons for this are manifold. See *infra* Part II.A. See also *Where the Money Goes, supra*, at 4 (listing various reasons, including: “[i]nformation about pretrial expenditures is almost always in the exclusive control of litigants and their attorneys”; “[r]esearchers must collect data from multiple sources”; “[i]t may be time-consuming or costly for litigants and their attorneys to retrieve relevant data about discovery-related costs”; “[s]taff in corporate departments, such as those in legal and information technology (IT), are unlikely to track their own litigation-related time expenditures”; and “[m]ost importantly, organizations may be reluctant to share information about their legal expenditures”).

6. See, e.g., LEE III, *supra* note 4, at 7.

of an entire set of rules in the Federal Rules of Civil Procedure (Rules).⁷ For example, the Supreme Court in *Bell Atlantic v. Twombly*,⁸ the seminal case in the paradigm of plausibility pleading, famously fretted about the costs of discovery in antitrust litigation but made no use of the Rules governing discovery.⁹

Today, the most salient discovery-related issue among practitioners is “preservation”: the duty to preserve relevant documents and electronically stored information (ESI)¹⁰ when litigation is reasonably anticipated. However, the Rules do not make clear that preservation is within the scope of discovery—or for that matter, within the scope of federal procedural lawmaking power at all. The Rules assiduously avoid any mention of the preservation of documents in anticipation of litigation, presumably to avoid concerns that such rules would tread upon state-created substantive law causes of action for spoliation of evidence. Yet over the past decade the lower federal courts have treated the silence of the Rules as an invitation to create a federal common law of preservation and spoliation.¹¹ This common law of preservation and spoliation has addressed a need for judicial policing of spoliation of ESI. However, it has done so while also engendering considerable dissension among the courts themselves¹² and causing rancorous complaints from litigants about what they claim are the severe burdens of the legal obligations imposed by the case law on preservation.¹³

The need for better information about preservation and discovery has never been greater. The Federal Civil Rules Advisory Committee has recently responded to the doctrinal chaos with proposed amendments addressing, among other things, preservation and discovery of documents and ESI in federal

7. See generally FED. R. CIV. P. 26–37.

8. 550 U.S. 544 (2007).

9. *Id.* at 558.

10. See A. Benjamin Spencer, *The Preservation Obligation: Regulating and Sanctioning Pre-litigation Spoliation in Federal Court*, 79 FORDHAM L. REV. 2005, 2006–07 (2011). This article will use “data,” “documents,” and “information” interchangeably to refer to both paper records and ESI.

11. *Id.* at 2005–07.

12. As a recent study noted:

Examples of conflicting holdings across and within jurisdictions include issues related to whether failure to issue a written legal-hold notice constitutes gross negligence per se, what preservation-related duties exist regarding potentially relevant evidence in the hands of third parties, whether a proportionality standard should be applied in deciding what information to retain, whether spoliation sanctions require a showing of negligence or a more stringent bad-faith standard, or whether sanctions should be imposed for the failure to properly preserve data without any need to show that the lost information was relevant or helpful to the requesting party.

Where the Money Goes, *supra* note 5, at 93 (footnotes omitted).

13. See, e.g., LAWYERS FOR CIVIL JUSTICE ET AL., PRESERVATION—MOVING THE PARADIGM 4 (2010); DISCOVERY SUBCOMM., ADVISORY COMM. ON CIVIL RULES, MINI-CONFERENCE ON PRESERVATION AND SANCTIONS 19–20 (2011).

litigation.¹⁴ This activity comes amid widespread calls for rules reform arising out of frustration with the patchwork of case law that currently governs preservation and sanctions for spoliation in federal court litigation.¹⁵ While there has been considerable debate about the merits of various proposals to amend the Rules, there is consensus on the need for further empirical research on the magnitude and nature of the costs associated with civil litigation, including the costs of discovery and preservation.

The growing awareness of the need for empirical data on the benefits and burdens of procedural rules has led to increasingly ambitious efforts to study certain aspects of the costs of civil litigation. These include the *Civil Rules Survey* by the Federal Judicial Center (FJC),¹⁶ the *Member Survey on Civil Practice* by the ABA Section of Litigation,¹⁷ and the *Litigation Cost Survey of Major Companies*.¹⁸ These studies provide essentially no discussion, however, of the cost of preservation, despite its centrality to debates about the costs of discovery and the need for Rules reform.¹⁹ The only prior, serious study of preservation costs was limited to in-depth, qualitative interviews with eight companies.²⁰

Prior to the work presented herein, *no* research had ever gathered quantitative data on preservation costs from a large sample of litigants.²¹ This article seeks to shed some light on the layers of uncertainty in and about the process of discovery. The parts that follow present new research results, propose new stylized facts about discovery, and tease out their implications for legal practice and Rules reform.

14. See Oliver H. Barber III, *Upcoming Changes to Federal Rules of Civil Procedure: Modernizing Scope of Discovery and Clarifying Consequences of Failure to Preserve*, LOUISVILLE B. ASS'N'S BRIEFS (Sept. 2014), <http://www.stites.com/learning-center/articles/upcoming-changes-to-federal-rules-of-civil-procedure-modernizing-scope>. Absent action by Congress to block the amendments, the amendments will take effect December 1, 2015.

15. See Spencer, *supra* note 10, at 2034.

16. See generally *Civil Rules Survey*, *supra* note 2.

17. See generally AM. BAR ASS'N SECTION OF LITIG., MEMBER SURVEY ON CIVIL PRACTICE: FULL REPORT § 11 (2009) [herein after ABA STUDY].

18. See generally LAWYERS FOR CIVIL JUSTICE ET AL., LITIGATION COST SURVEY OF MAJOR COMPANIES (2010) [hereinafter LITIGATION COST SURVEY].

19. See ABA STUDY, *supra* note 17, at 2; *Civil Rules Survey*, *supra* note 2, at 1. One reason for this is that prior studies have been surveys of outside counsel. The costs of preservation activities tend to be borne directly by the client, rather than outside counsel, and often begin before a lawsuit is filed. See William H.J. Hubbard, *Preservation Costs Survey Final Report* (Feb. 18, 2014), <http://www.regulations.gov> (search by ID number: USC-RULES-CV-2013-0002-2201) [hereinafter *Preservation Costs Survey Report*].

20. *Where the Money Goes*, *supra* note 5, at iii, 15 (“Our approach here was qualitative in nature because it was clear that gauging the magnitude of preservation expenses in individual cases would present some daunting hurdles.”).

21. *Id.* at 86 (“Despite the costs of preservation having become one of the most discussed topics in the legal press of late, we are not aware of *any* empirical research that has collected quantitative information about such costs across significant numbers of actual cases.”).

Part I briefly summarizes Rules and case law governing preservation obligations in federal civil litigation. Part II describes original, empirical research conducted on the costs of preservation and discovery. This study, referred to as the Preservation Costs Survey (Survey), is the first—and to date the only—systematic effort to measure the extent and costs of preservation activity across a cross-section of companies. Although focused on preservation costs, this Survey collected quantitative data on the volume, timing, and cost of other aspects of discovery, particularly those aspects of discovery farthest removed from court oversight (i.e., collection and processing, as opposed to review and production). It is also unique among quantitative studies in that it focuses on the costs of the *client's* own discovery-related activities rather than the costs incurred by outside counsel retained to litigate cases.

The Survey responded to a call from the Advisory Committee on Civil Rules for empirical data on the costs of preservation.²² The Survey was supported by an industry organization called the Civil Justice Reform Group, whose members include large companies concerned with the costs of preservation but, tellingly, could not quantify their own preservation costs.²³ The Survey collected information from 128 companies and gathered detailed, case-level data on preservation activity in over 3,600 separate litigation matters. Surveyed companies ranged from small companies without in-house litigation counsel to Fortune 100 companies who have entire staffs of attorneys and other professionals devoted full-time to compliance with litigation-related preservation obligations.

Part III presents key findings from the research conducted and proposes three new stylized facts about preservation and discovery, complete with three accompanying metaphors: the *discovery sombrero*, the *preservation iceberg*, and the *long tail of costs*. The usual progression of discovery activities in a given case begins with the *preservation* of information that may be relevant to ongoing or threatened litigation. Next comes the *collection* of documents for processing and review. *Processing* refers to actions such as decryption, decompression, and de-duplication of documents to render them amenable to review and to reduce redundancies and other unnecessary costs further downstream. *Review* is the work lawyers conduct to determine relevance and privilege of the documents in discovery. *Production* is the process of turning over to opposing counsel the relevant, non-privileged materials within the scope of discovery.

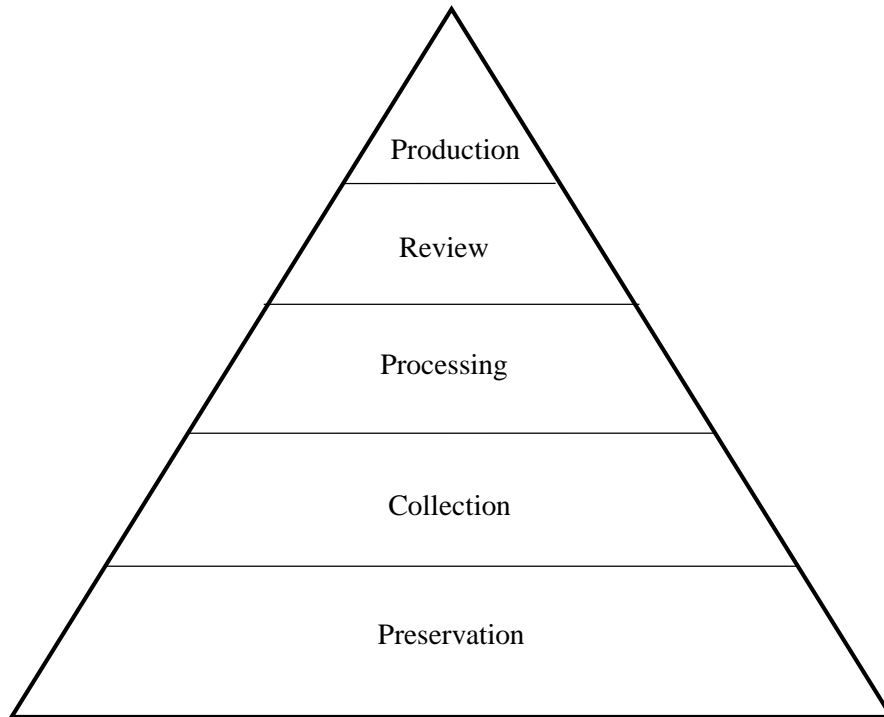
Obvious quantitative questions immediately arise: how much of what is preserved is collected? How much of what is collected is processed? One might imagine a winnowing process whereby the parties begin with a large set of

22. See *Preservation Costs Survey Report*, *supra* note 19. Earlier drafts of portions of this article were shared with the Advisory Committee.

23. As noted in the *Preservation Costs Survey Report* submitted to the Advisory Committee on Civil Rules, “The Civil Justice Reform Group describes itself as an organization formed and directed by general counsel of Fortune 100 Companies concerned about America’s justice system.” *Preservation Costs Survey Report*, *supra* note 19, at 6 n.10.

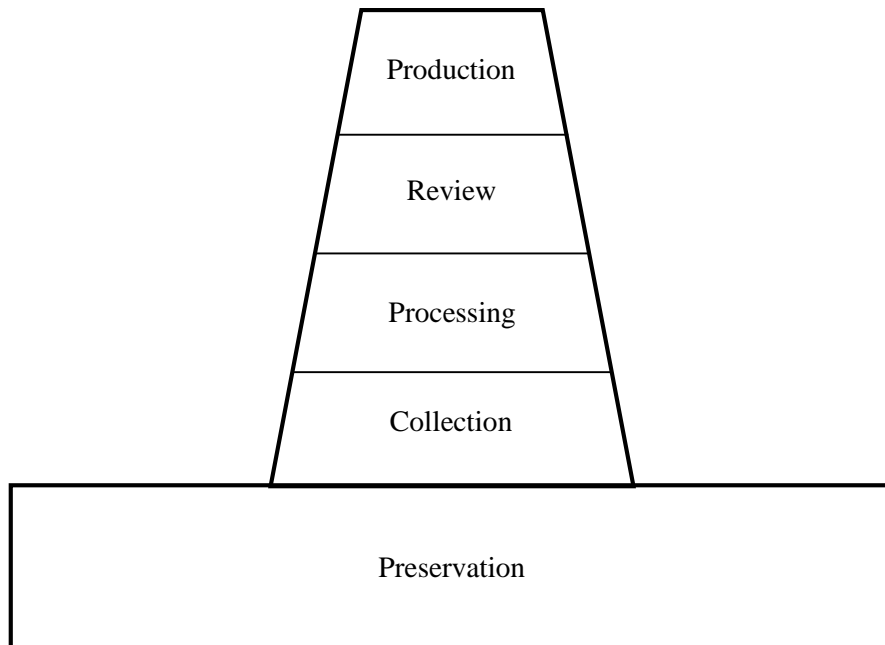
documents that are preserved, which they gradually trim down to the materials most relevant to settlement, summary judgment, or trial, as in Figure 1.

FIGURE 1: THE STAGES OF DISCOVERY



This Survey, however, indicates a different relationship between the volumes of data involved in preservation relative to the other stages. The progression is not so much a discovery pyramid as it is a discovery sombrero, introduced in Part III.A and shown below.

FIGURE 2: THE DISCOVERY SOMBRERO



The immediate implication of this fact is that preservation—a stage of discovery that to date has gone unmentioned in the Rules and is remote from judicial oversight—has the potential to be a source of substantial costs in the civil justice system.

Part III.B introduces the “preservation iceberg,” which begins to unpack exactly *how* and *where* the huge data volumes being preserved for civil litigation impose costs on preserving parties. Debates about the costs of preservation and the need for Rules reform tend to be framed by anecdotes about what this article refers to as the “fixed costs” of preservation. An example of a “fixed cost” is the million dollars a large company spends on a computer system to facilitate the preservation of ESI. Given the obvious self-interest of the parties offering such anecdotes, one might wonder whether such anecdotes exaggerate the costs of preservation.

Perhaps the most surprising finding of this research is that such anecdotes *severely underestimate* the total costs of preservation activity. While a Fortune 500 company might spend \$4 million on computer systems, it is merely the “tip of the iceberg” of preservation costs, and, as with icebergs, the tip is a mere ten percent of the whole. Anecdotes about these costing \$4 million likely reflect real, but invisible, costs closer to \$40 million.

Why have the true costs of preservation evaded observation? Some costs, such as the invoice for a new computer application, are easy to observe. But this study reveals that the greatest cost of preservation activity is not the price tag of new technology, but the *human cost* in employee time diverted from business activities to litigation-related activities. The costs associated with this diversion of human effort constitute over ninety percent of total preservation costs in the largest companies and essentially 100 percent of total preservation costs in smaller companies.

Part III.C introduces the “long tail of costs,” a phenomenon that can harmonize the seemingly irreconcilable data and anecdotes that populate the rhetoric of procedural reform: on the one hand, there are documented accounts of preservation and discovery costing millions of dollars in cases that companies regularly litigate; and on the other hand, there are data showing that median costs are measured in the thousands, not millions, of dollars.

Both accounts are true: the distribution of preservation costs is such that most litigation matters involve moderate costs, but the distribution is highly skewed, with a long but thin tail of extremely expensive litigation matters. The skew is so great that even though cases with blockbuster costs are rare—maybe five percent of all litigation matters—they account for the *majority* of all costs. Interestingly, the data collected in this article on companies’ preservation costs is strikingly consistent with previous data collected on outside counsels’ litigation costs, which suggests that this “long tail of costs” reflects a deep phenomenon affecting all of litigation.

Part IV discusses the relevance of the discovery sombrero, the preservation iceberg, and the long tail of litigation costs to policymaking and legal doctrine governing discovery. While the first objective of this article is to introduce key, stylized facts on preservation and discovery, which are relevant to many questions in this field, the second objective is to explore how these stylized facts help identify the way forward in addressing three specific cleavages in the law.

First, the discovery sombrero interacts in a surprising way with *Erie Railroad Co. v. Tompkins*²⁴ and its progeny. Current federal efforts to regulate preservation through federal common law need to account for the fact that much of what is regulated occurs outside the context of federal litigation. For example, federal rules governing the conduct of preservation direct the behavior of parties who will ultimately find themselves in state, not federal, court. This raises the specter of *Erie*. Although objections have been raised against a federal rule on preservation because of Rules Enabling Act concerns, these concerns are precisely backwards. If anything, federal rulemaking solves, rather than raises, an *Erie* infirmity.

Second, the preservation iceberg interacts in an unexpected way with debates about the choice between reliance on legal reform and reliance on technological innovation to reduce costs associated with discovery. Big businesses have

24. 304 U.S. 64 (1938).

claimed that legal change is needed to control costs that have multiplied due to technological change, while their opponents have argued that technology can also lower preservation and discovery costs.²⁵ The research presented herein suggests that both of these arguments are misdirected. Preservation costs are very high—indeed, even higher than proponents of legal reform have recognized—but most of the costs are human costs, rather than technology costs.

Further, technology is not a substitute for legal reform because technological solutions are practical only for the largest companies where high-tech solutions justify their high price tag. For smaller companies—and in this study, “smaller” includes companies with under 1,000 employees—technology plays a much smaller role in the preservation process. From this point of view, legal innovation, rather than technological innovation, may be the best hope for controlling preservation costs of individuals, small businesses, and virtually everyone other than the largest and most sophisticated litigants.

Third, the stark differences among cases involving different substantive fields—compare the typical scope of discovery in an antitrust case versus an employment discrimination case, or compare the information asymmetries in those cases with those in, say, contract cases—put constant pressure on the transsubstantive design of the Rules. The long tail of costs, however, points the way to a Rules-based approach to controlling discovery that does not require the Rules to abandon a commitment to transsubstantive standards. Given that most preservation and discovery costs are concentrated in a small share of cases, one can structure the Rules to set presumptive limits on discovery that leave most cases unaffected but facilitate party bargaining and judicial oversight in the fraction of cases where the issue of cost control may deserve careful attention.

I. LAW GOVERNING PRESERVATION AND DISCOVERY

The Rules do not explicitly address preservation. The Rules do, however, provide the framework for addressing discovery generally. For example, Rule 1 dictates that the Rules “should be construed and administered to secure the just, speedy, and inexpensive determination of every action and proceeding.”²⁶ Rule 26(b)(1) outlines the scope of discovery: “[p]arties may obtain discovery regarding any nonprivileged matter that is relevant to any party’s claim or defense.”²⁷ Rule 26(b)(2) sets out the limits of permissible discovery.²⁸ In particular, Rule 26(b)(2)(C) outlines the bases for limiting discovery and imposes a mandatory requirement on courts to limit discovery, even *sua sponte*, if “the burden or expense of the proposed discovery outweighs its likely

25. See ABA STUDY, *supra* note 17, at 110 (finding that defense counsels were less optimistic about technological advances improving cost efficiency than were plaintiffs’ lawyers).

26. FED. R. CIV. P. 1.

27. FED. R. CIV. P. 26(b)(1).

28. FED. R. CIV. P. 26(b)(2).

benefit.”²⁹ In this way, Rule “26(b)(2)(C) cautions that all permissible discovery must be measured against the yardstick of proportionality.”³⁰

While it appears these Rules present a set of guidelines for discovery that might arguably apply to preservation, these Rules have been by all accounts ineffective at providing meaningful guidance to courts and litigants on questions of preservation.³¹ Federal case law on preservation has largely ignored the Rules and has done little to settle the question of what needs to be preserved and by what standards a failure to preserve will be judged. In fact, courts do not even agree on “whether a proportionality standard should be applied in deciding what information to retain.”³² In *Pippins v. KPMG LLP*,³³ the court found the proportionality standard too “amorphous” to be useful and instead concluded that “[u]ntil a more precise definition is created by rule, prudence favors retaining all relevant materials.”³⁴

Of course, a few principles governing preservation are fairly well-settled. For example, the duty to preserve relevant data attaches when a party reasonably anticipates litigation.³⁵ Failure to take appropriate steps to preserve data can subject a party to sanctions, which a federal court may impose under its inherent power.³⁶

One key step—perhaps *the* key step—in complying with the duty to preserve is the issuance of a litigation hold. A “litigation hold” is a set of actions taken by a company to comply with preservation obligations in a litigation matter.³⁷ A litigation hold will define the scope of documents and data that must be preserved.³⁸ A “litigation-hold notice” is an instruction from legal counsel to an employee that the employee must retain all documents and data in her custody that are within the scope of the litigation hold; for example, in a products liability case, the scope might be all documents relating to the safety of a particular

29. FED. R. CIV. P. 26(b)(2)(C).

30. *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 269 F.R.D. 497, 523 (D. Md. 2010).

31. See DISCOVERY SUBCOMM., *supra* note 13, at 2–3; see also LAWYERS FOR CIVIL JUSTICE ET AL., *supra* note 13, at 11 (noting that “only two courts have considered the application of proportionality to the scope of preservation pursuant to FRCP Rule 26(b)(2)(C) although neither court specifically analyzed its application”).

32. *Where the Money Goes*, *supra* note 5, at 93 (citations omitted).

33. No. 11 Civ. 0377(CM)(JLC), 2011 WL 4701849, at *6 (S.D.N.Y. Oct. 7, 2011).

34. *Id.* at 6 (citations omitted) (internal quotation marks omitted).

35. See, e.g., *Pension Comm. of the Univ. of Montreal Pension Plan v. Banc of America Sec., LLC*, 685 F.Supp.2d 456, 466 (S.D.N.Y. 2010); *Surowiec v. Capital Title Agency, Inc.*, 790 F.Supp.2d 997, 1005 (D. Ariz. 2011) (“It is well established that the duty to preserve arises when a party knows or should know that certain evidence is relevant to pending or future litigation.”) (internal quotation marks omitted).

36. *Surowiec*, 790 F.Supp.2d at 1008.

37. Stephanie F. Stacy, *Litigation Holds: Ten Tips in Ten Minutes*, <http://www.ned.uscourts.gov/internetDocs/cle/2010-07/LitigationHoldTopTen.pdf> (last visited Sept. 21, 2015).

38. *Id.*

product that the company produces.³⁹ The usual practice is to send a litigation-hold notice to the set of “key players” who are likely to have data relevant to the dispute in question.⁴⁰ As described below in Part II, one primary measure of preservation activity is the number of litigation-hold notices issued.

However, many questions regarding preservation remain unsettled. Courts have diverged on questions such as “whether failure to issue a written legal-hold notice constitutes gross negligence per se”⁴¹ and “what preservation-related duties exist regarding potentially relevant evidence in the hands of third parties.”⁴² Most notably, courts have not even converged on a standard for the two essential prerequisites for imposing spoliation sanctions: the alleged spoliator’s state of mind and prejudice to the other party.

As to the former, most courts require bad faith—in the form of intentional destruction of data to prevent its use in litigation—before imposing serious sanctions such as entering judgment against the offending party or giving an adverse inference instruction to the jury.⁴³ But some courts explicitly disclaim any requirement of bad faith.⁴⁴ Further, some courts are willing to infer negligence from the mere fact that *any* data whatsoever was lost.⁴⁵ This split is complicated by the idiosyncratic terminology applied by some courts that distinguish between “willfulness” and bad faith, such that a merely volitional act—such as good faith deletion of data without awareness of its potential relevance to litigation—is “willful” spoliation.⁴⁶ As for the standard for finding prejudice, some courts will presume relevance and prejudice from gross

39. *Id.*

40. *See id.*

41. *Where the Money Goes*, *supra* note 5, at 93 (citations omitted).

42. *Id.* (citation omitted).

43. *See, e.g.*, *Rimkus Consulting Grp., Inc. v. Cammarata*, 688 F.Supp.2d 598, 607, 653 (S.D. Tex. 2010) (imposing sanctions including an adverse inference instruction against defendants that intentionally deleted emails after a duty to preserve had arisen).

44. *See, e.g.*, *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 220 (S.D.N.Y. 2003) (holding that negligent spoliation is sufficient to incur sanctions).

45. *See id.* (“Once the duty to preserve attaches, any destruction of documents is, at a minimum, negligent.”). For example, in *Pension Committee v. Banc of America Securities, LLC*, a party had failed to produce a number of emails that were later discovered in another production. 685 F. Supp. 2d 456, 462, 489 (S.D.N.Y. 2010). The court stated that, “[t]his, alone, demonstrates that the [party’s] effort to find and produce all relevant documents was insufficient.” *Id.* at 489. *See* Michael W. Deyo, *Deconstructing Pension Committee: The Evolving Rules of Evidence Spoliation and Sanctions in the Electronic Discovery Era*, 75 ALB. L. REV. 305 (2011/2012) (discussing *Pension Committee* in detail). *See* *Surowiec v. Capital Title Agency, Inc.*, 790 F.Supp.2d 997, 1007 (D. Ariz. 2011) (rejecting this approach as “too inflexible”).

46. *See, e.g.*, *Nucor Corp. v. Bell*, 251 F.R.D. 191, 194 (D.S.C. 2008) (holding that if the spoliation is “willful” adverse inference instructions are appropriate “even in the absence of bad faith,” and that sanctions of dismissal and default judgment require a showing of “bad faith”).

negligence or bad faith,⁴⁷ while others will make such an inference only from bad faith,⁴⁸ or perhaps not at all.⁴⁹

II. THE PRESERVATION COSTS SURVEY

The Preservation Costs Survey is the first systematic, quantitative study of preservation costs across a spectrum of companies that are engaged in preservation activities. Part II.A discusses the key constraints that drove the design of the Survey. Prior to this study, these factors had combined to prevent any systematic collection of preservation costs. Part II.B describes the Survey methodology, and Part II.C describes the sampled companies.⁵⁰

A. Obstacles to Empirical Work on Preservation

In order to measure the costs associated with preservation obligations, this Survey had to overcome a number of challenges that prevented prior research from determining the nature and scale of preservation costs.⁵¹ Indeed, a prerequisite to gathering any quantitative data was identifying *which* costs of preservation are even susceptible to practical measurement. Thus, the first phase of the survey design focused on in-depth interviews with personnel at a pilot group of companies. These interviews sought to identify which aspects of the costs of preservation are most amenable to study and which would be difficult, or as a practical matter impossible, to estimate. Not surprisingly, every company

47. *Pension Comm.*, 685 F.Supp.2d at 467 (“Relevance and prejudice may be presumed when the spoliating party acted in bad faith or in a grossly negligent manner.”).

48. *See, e.g., D’Onofrio v. SFX Sports Grp., Inc.*, No. 06-687(JDB/JMF), 2010 WL 3324964, at *10 (D.D.C. Aug. 24, 2010). The court held:

When a party, for example, has acted negligently and lost evidence, [an adverse] inference does not flow naturally from the facts. When a person purposefully destroys evidence, it is reasonable to infer that he did so to keep it from being used against him [When the action was negligent or reckless,] a court cannot logically infer the intent of what a party did from its behavior because its behavior was unthinking.

Id.

49. *Orbit One Commc’ns v. Numerex Corp.*, 271 F.R.D. 429, 440 (S.D.N.Y. 2010) (holding that before addressing culpability “a court considering a sanctions motion must make a threshold determination whether any material that has been destroyed was likely relevant even for purposes of discovery”).

50. As noted above, earlier versions of portions of this article were submitted as a public comment to the Advisory Committee on Civil Rules in the form of the *Preservation Costs Survey Report*. The *Preservation Costs Survey Report* contains details on a number of Survey results not discussed in this article. This article develops doctrinal and prescriptive analysis that was beyond the scope of the Report submitted to the Advisory Committee. The discussion of the background and methodology of the Survey, however, is largely unchanged in this article from the earlier version in the Report, although in some places, the Report goes into more detail on the finer points of the methodology and data. For this reason, the article will include notes directing the reader to relevant portions of the *Preservation Costs Survey Report* that contain details related to the discussion in the text herein.

51. *See Preservation Costs Survey Report, supra* note 19, at 9–13.

interviewed for the Survey expressed that estimating the costs of preservation is difficult.⁵²

There are several reasons for this difficulty. First, identifying systems and their cost requires time-consuming, individualized investigation of each company.⁵³ Each company has different computer systems, different internal business flow, and different technology needs. While an “off-the-shelf” solution from an outside vendor comes with an invoiced price, the full cost of that solution includes company time and resources for project bidding, implementation, and maintenance over time. Systems that are developed in-house are even harder to price.

Second, individualized investigation is required to ensure that the costs being measured are properly attributable to preservation obligations, rather than other motivations. To address this concern, the author relied on detailed, in-depth interviews with companies to accurately identify specific systems whose *sole* purpose was compliance with preservation obligations. As a consequence of this approach, the Survey data on these costs generates a conservative estimate of the total costs of technologies adopted in response to preservation burdens.

Third, the human cost of preservation-related activity in terms of lost work time has never before been measured. One major cost of preservation obligations is the lost employee time spent complying with duties imposed through the issuance of litigation-hold notices. Because the cost of compliance with litigation-hold notices is dispersed throughout a company, and because the cost primarily takes the form of lost time rather than monetary payments, measuring the magnitude of this cost is difficult.⁵⁴ The time and energy that employees must divert towards preservation is never recorded or compensated, unlike the time spent by dedicated lawyers, such as outside counsel.⁵⁵

The strategy to measure these costs was to collect detailed information on the number of matters with litigation holds and the number of employees subject to each litigation hold at a sample of companies. The author combined these counts of employees subject to litigation holds with estimates of time lost per employee and the hourly cost of employee time, to quantify in dollar terms the value of employee time that is diverted from business purposes to compliance with preservation obligations.

52. See also *Where the Money Goes*, *supra* note 5, at xix (“Most interviewees did not hesitate to confess that their preservation costs had not been systematically tracked in any way and that they were unclear as to how such tracking might be accomplished.”).

53. *Id.* at 85.

54. See *id.* at xviii.

55. See *id.* at 85.

Part of the reason for a lack of existing information in this area appears to be that much of preservation involves expenditures incurred internally, such as the costs of IT staff time, law department attorney and paralegal time, other employees’ time (such as the effort required of custodians to comply with legal-hold notices), and purchases and licensing of applications and hardware to handle preservation.

Id.

Fourth, companies are unable or reluctant to share sensitive and confidential information about litigation-related costs.⁵⁶ In many cases, companies simply do not have the information or cannot gather it at reasonable cost.⁵⁷ This reluctance is also due to companies' concern that disclosing information about their litigation experiences and expenses could be used strategically against them in litigation.⁵⁸ For this reason, all information collected for the Preservation Costs Survey was gathered subject to assurances of strict confidentiality and anonymity for each survey participant.⁵⁹

Fifth, many costs associated with preservation are diffuse and cannot be directly measured. For example, while the lost time of affected employees can be measured, other costs remain unmeasured, such as delays in basic business processes like rolling out new computers to employees due to concerns about the preservation of data stored on due-to-be-retired hard drives.⁶⁰ Thus, the preservation costs measured by the Survey do not exhaust the universe of costs imposed by preservation obligations.

B. Survey Methodology

The Survey involved three phases, lasting from late 2011 through early 2014.⁶¹ The Survey was done with the support and assistance of the Civil Justice Reform Group (CJRG), a group of in-house counsel at large, U.S. corporations.⁶² CJRG asked a number of large companies to participate in the Survey and coordinated with other business associations (including small and medium-sized businesses) to request that their members participate in the Survey. This provided unprecedented access to information about companies' experiences with preservation and discovery; as noted above, it is usually impossible to

56. *See id.* at 4.

57. Only fourteen percent of Survey respondents stated that they track the costs of their litigation holds. *See Preservation Costs Survey Report, supra* note 19, at 12 n.18.

58. *See Where the Money Goes, supra* note 5, at 4.

59. *See Preservation Costs Survey Report, supra* note 19, at 13 (explaining further the steps to protect anonymity and data integrity). For example, in some cases, exact numbers are rounded or topcoded (e.g., employee counts larger than 100,000 are reported as "> 100,000") to protect anonymity. *Id.*

60. *Where the Money Goes, supra* note 5, at 86. The report explained:

[T]here may be economic impacts resulting from a decision not to adopt certain IT products (such as instant messaging or social-networking platforms) that might present significant difficulties when preserving information, from not implementing more-efficient data systems due to the need to maintain older legacy platforms and processes, from slower computer-system performance caused by halting the routine deletion of obsolete information in transactional databases, or from a reduced ability to recover lost but nevertheless important data due to a shift from a long-term data backup process to a short-term disaster-recovery system primarily because of preservation concerns.

Id.

61. *See Preservation Costs Survey Report, supra* note 19, at 13–16 (discussing details on the Survey methodology).

62. *Id.* at 6 n.10.

collect information on litigation-related costs from companies. Indeed, even with CJRG helping to convince companies to participate, a major component of survey design and promotion was to provide detailed, credible assurances of confidentiality and anonymity to respondents.

While the sponsorship of CJRG was essential to the viability of this project, there is no question that CJRG is an advocacy organization, and the author was compensated for his time and expenses associated with designing the survey, interviewing respondents, and processing response data.⁶³ Because of this, the methodology involved steps taken to protect the independence of the research and insulate the survey results from any outside influence. CJRG agreed not to participate in the design of the survey questions or access the data collected in the course of the survey. Nor was CJRG involved in the analysis of the data. Further, CJRG retained no interest in or oversight of the use or publication of results in this article. Thus, to be absolutely clear, all of the arguments and conclusions herein are the author's alone.

Given the complexity of the topic, and the largely unprecedented nature of a study focused on preservation costs, the Survey utilized a three-phase design. Phase I involved a set of four, in-depth case studies of large companies. These case studies involved both qualitative interviews and requests for quantitative data to be used for statistical analysis. One important aspect of Phase I was developing the survey instrument. The development process began with an extensive written survey coupled with follow-up interviews to obtain feedback on the clarity and practicability of each question. This information was used to draft the survey instruments used with larger samples of companies during Phases II and III.

Phase II broadened the sample of companies to thirteen and continued to employ an in-depth, case-study approach. A revised questionnaire was combined with interviews and the collection of matter-level datasets of preservation activity in order to create as complete as possible a picture of the sources and amounts of preservation costs for large companies. As used in this article, "matter-level datasets" are datasets in which information on the number of litigation holds is provided for each individual litigation matter. Often, a "matter" is a lawsuit, but not always. Litigation matters include both filed and anticipated lawsuits. For this reason, this article uses the term "matter" rather than "case." In addition to survey and interview responses, Phase II yielded six unique databases of matter- and employee-level preservation activity within specific companies. These databases of preservation activity were provided on a strictly confidential, anonymous basis. These datasets together provided information on over 3,600 separate litigation matters involving over 770,000

63. As noted above, CJRG's interest in sponsoring this research was to respond to calls for empirical data on preservation costs from members of the Advisory Committee on Civil Rules, who were considering proposals to amend the Rules to address preservation. The *Preservation Costs Survey Report* took no position on specific proposals, but did conclude that preservation costs were large enough to merit attention from the rulemakers. *Id.* at 47.

litigation hold notices issued to individual employees in individual matters. They are the first large samples of case-specific preservation activity data ever compiled for research purposes.

Phase III involved a shortened survey questionnaire and no interviews or requests for data. This Phase was deliberately designed to be distributed to a larger number of companies, which would allow them to respond with a much smaller investment in human resources. The goal of Phase III was to obtain survey responses from a large sample of companies, including small and medium-sized businesses, in order to draw inferences about preservation activity in a broader cross-section of civil litigants. Phase III was publicized to companies through groups such as the National Association of Manufacturers, Lawyers for Civil Justice, and the Association of Corporate Counsel. The surveys could be completed on a printable form or by an online survey instrument hosted on research.net. The Phase III survey was open from October 2013 to January 2014.⁶⁴ By the conclusion of Phase III, a total of 128 unique companies had completed survey questionnaires.⁶⁵

Although this study is by far the most rigorous survey of preservation costs ever conducted, this study's methodology, by its very nature, cannot guarantee a representative sample of all companies with preservation obligations.⁶⁶ As with any survey, this study could include only those who were willing and able to respond. Nonetheless, the Survey results provide several indications that the sample may be representative of the larger population of companies.

First, the results from each phase of the Survey are remarkably consistent with each other, despite substantial differences in the process by which companies were solicited for participation and the degree of effort required by the companies to complete their participation. This suggests that the amount of effort required to participate is not strongly correlated with the characteristics of the company.

Second, many of the patterns that one would predict to see in the data based on strong *a priori* justifications do, in fact, appear in the data. For example, smaller companies have very few (often zero) litigation attorneys and report

64. Two surveys were returned in February 2014. They are included in the results reported below. Excluding them has little effect on the reported results. *See id.* at 15 n.21.

65. The Phase III questionnaire is reproduced in the Appendix of the *Preservation Costs Survey Report*. *See id.* at 55–59.

66. *Compare Where the Money Goes*, *supra* note 5, at xiii–xiv (“We asked participants to choose a minimum of five cases in which they produced data and electronic documents to another party as part of an e-discovery request. . . . Because the participating companies and cases do not constitute a representative sample of corporations and litigation, we cannot draw generalizations from our findings that apply to all corporate litigants or all discovery productions.”), with *Preservation Costs Survey Report*, *supra* note 19, at 6 (“The Survey ultimately collected information from 128 companies from a wide spectrum of industries. These companies vary from small companies without in-house litigation counsel to Fortune 100 companies who have entire staffs of attorneys and other professionals devoted full-time to preservation work.”).

dramatically fewer active cases.⁶⁷ This pattern might not emerge if only the most sophisticated (or most embroiled in litigation) smaller companies participated in the Survey.

Third, unlike prior studies that also depended on the willingness of companies to provide data on discovery costs (*Litigation Cost Survey*) or to provide interview responses on preservation (*Where the Money Goes*), the Preservation Costs Survey did not allow participating companies to select specific cases for inclusion in the sample. Rather, the questionnaire asked only for information about cases in the aggregate, and the requests for databases of preservation activity included all litigation matters with litigation holds (excluding asbestos matters). Thus, the Preservation Costs Survey provides analysis of the first *truly representative* samples of the within-company distribution of litigation activity.

C. Sample Characteristics

The 128 survey respondents represent a broad cross-section of companies in the United States.⁶⁸ The participating companies come from a wide variety of industries.⁶⁹ The most heavily represented categories were health care, insurance, technology, and conglomerate, each with at least ten respondents.

The number of people employed worldwide by each company ranges from 18 to over 100,000.⁷⁰ Importantly, although large companies were the focus of Phases I and II, smaller companies are well represented in the sample. About a quarter of all respondents (twenty-four percent) have 1,000 or fewer employees worldwide; the same proportion have 500 or fewer U.S. employees, the threshold usually used to define a small or medium-sized enterprise (SME).⁷¹ The largest companies, those with over 100,000 employees worldwide, make up about one-sixth (sixteen percent) of the sample.⁷²

67. See *infra* note 94.

68. See *Preservation Costs Survey Report*, *supra* note 19, at 17–19 (discussing the survey respondents and data collected in further detail).

69. The categories are: Automobiles & Parts, Banks, Chemicals, Conglomerate, Financial Services, Food & Beverage, Health Care, Industrial Goods & Services, Insurance, Media, Oil & Gas, Other, Personal & Household Goods, Retail, Technology, Telecommunications, Travel & Leisure, and Utilities.

70. In order to protect the anonymity of some respondents, exact employee counts above 100,000 are not reported.

71. This article refers to companies with 1,000 or fewer employees as “smaller companies.”

72. Herein, the author will occasionally refer to companies with close to or more than 100,000 employees worldwide as “large companies.” Companies with 1,001–10,000 employees made up twenty-nine percent of the sample; companies with 10,001–100,000 employees made up thirty percent of the sample.

TABLE 1: SURVEY DESCRIPTIVE STATISTICS

Panel A: Employees, Lawsuits, and Litigation Hold Matters				
	Mean	Median	Min	Max
Total employees	43,454	8,000	18	> 100,000
U.S. employees	21,678	6,100	0	> 100,000
In-house litigation attorneys	12	4	0	> 50
Active suits	1,399	33	0	> 10,000
Open matters with holds	686	33	0	> 10,000
Panel B: Share with Preservation Resources or Practices				
Issues litigation holds notices				100%
Has formal preservation policies				84%
Tracks litigation holds and notices				63%
Has e-discovery team				40%
Has legal IT group				31%

The volume of litigation varies widely across these companies; the number of suits currently active varies from 0 to over 10,000.⁷³ Asbestos litigation was specifically excluded from the Survey.⁷⁴ There is also great variation in the number of litigation holds that companies report as active.⁷⁵ The number of in-

73. In order to protect the anonymity of some respondents, exact counts of lawsuits and litigation holds above 10,000 are not reported. Five companies did not report numbers of suits, and seven companies did not report numbers of matters with holds.

74. While asbestos litigation remains an important part of the federal civil docket, it is *sui generis* with respect to preservation; at this point in the history of asbestos litigation, virtually every document in the possession of a company defendant that could possibly be relevant to asbestos claims has long ago been preserved and produced.

75. As the duty to preserve may arise before a lawsuit is filed, the number of matters subject to litigation holds may be greater than the number of lawsuits. Conversely, a single litigation hold may suffice for a number of related lawsuits, and thus a company may have fewer litigation holds than lawsuits.

house litigation attorneys ranges from zero to over fifty.⁷⁶ Most in-house litigation teams are small—the median is four, and seventeen out of the 128 companies have no in-house litigation counsel. See Table 1.⁷⁷

Some basic Survey results are unsurprising.⁷⁸ Consistent with the great weight of anecdotal evidence and prior qualitative studies, surveyed companies generally reported significant preservation burdens, although some reported little or none at all.⁷⁹ In interviews, companies expressed that they are deliberately “overinclusive” or “overpreserve” to protect themselves against the great uncertainty associated with the current law of preservation.⁸⁰ Government investigations, rather than private lawsuits, ranked first in terms of preservation-related problems. This is unsurprising, given the often sweeping scope and indefinite duration of government investigations, which may entail incredibly broad requests for information.

III. THREE STYLIZED FACTS ABOUT LITIGATION

A. *The Discovery Sombrero*

Perhaps the most basic finding of the Survey was that most respondents did not know the extent of their preservation activity or what fraction of data that is put on litigation hold is ever collected, let alone reviewed or used, in the course of discovery. Those that did reported on average that perhaps half (fifty-one percent) of all data that is preserved is never processed and reviewed.⁸¹ This result is consistent with a recent survey by an e-discovery vendor, which found that for most companies, legal holds proceed to collection less than half the time.⁸²

For larger companies, the drop-off from preservation to collection, processing, and review is even steeper. Figure 3 presents data from a large company on the number of custodians involved in three stages of discovery:

76. In order to protect the anonymity of some respondents, exact counts of litigation attorneys above fifty are not reported. Three companies did not report the number of litigation attorneys.

77. For Total Employees and U.S. Employees, $N = 126$. Median numbers of employees are rounded by up to one percent to protect respondent anonymity.

78. In addition to the details noted here, the *Preservation Costs Survey Report* provides many additional results. See *Preservation Costs Survey Report*, *supra* note 19, at 20–43.

79. See *id.* at 20–21 (finding that over seventy-nine percent (102 of 128) of respondents reported a “great extent” or “moderate extent” of burdens from preservation activity.). See also *Where the Money Goes*, *supra* note 5, at xix (“All interviewees reported that preservation had evolved into a significant portion of their companies’ total e-discovery expenditures.”).

80. *Preservation Costs Survey Report*, *supra* note 19, at 46. See also *Where the Money Goes*, *supra* note 5, at 92 (“If there was one consistent theme in what we heard, it revolved around complaints of a lack of understandable legal authority and guidance that could be comfortably relied on when making preservation decisions.”).

81. See *Preservation Costs Survey Report*, *supra* note 19, at 43–44.

82. LEGAL HOLD AND DATA PRESERVATION BENCHMARK SURVEY 2013 16 (2013) (finding that for “64 percent of respondents, legal holds progress to collection less than half the time”).

preservation, collection, and processing. Out of over 5,000 custodians placed on litigation hold, and thus subject to preservation obligations, fewer than ten percent ultimately see their data collected, let alone processed.

FIGURE 3: NUMBER OF CUSTODIANS SUBJECT TO PRESERVATION, COLLECTION, AND PROCESSING OF AN ANONYMOUS LARGE COMPANY

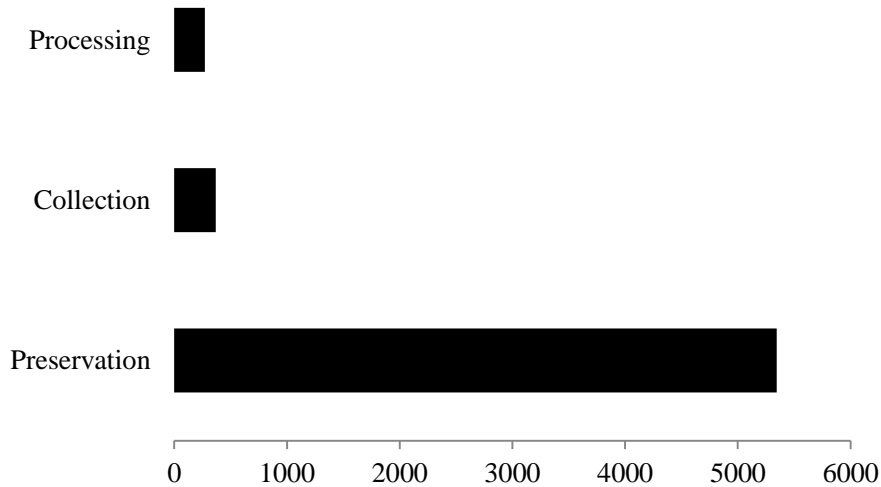
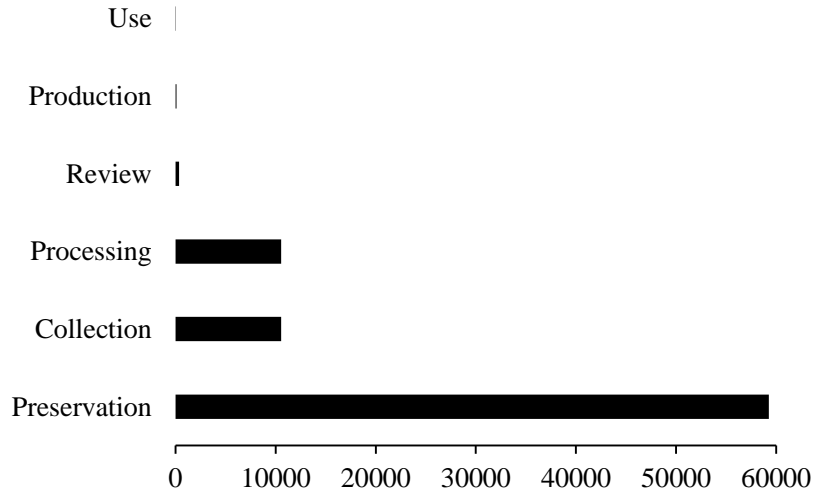


Figure 4 presents a similar picture with non-anonymous data provided in public testimony on behalf of Microsoft. In Figure 4, the unit of measurement is the quantity of data preserved, collected, and processed rather than the number of custodians subject to those activities. The Microsoft data also illustrates how little data, relative to the quantity preserved, is ever used in litigation.⁸³ From this, the shape of the discovery sombrero, illustrated above in Figure 2, is apparent: a wide “brim” of preservation, and a much narrower, tapering set of documents subject to collection, processing, and so on.⁸⁴

83. Testimony of David M. Howard on behalf of Microsoft Corp., *Transcript of Public Hearings on Proposed Amendments to the Federal Rules of Civil Procedure* 79–80 (Jan. 9, 2014), <http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/public-hearings/civil-hearing-transcript-2014-01-09.pdf>.

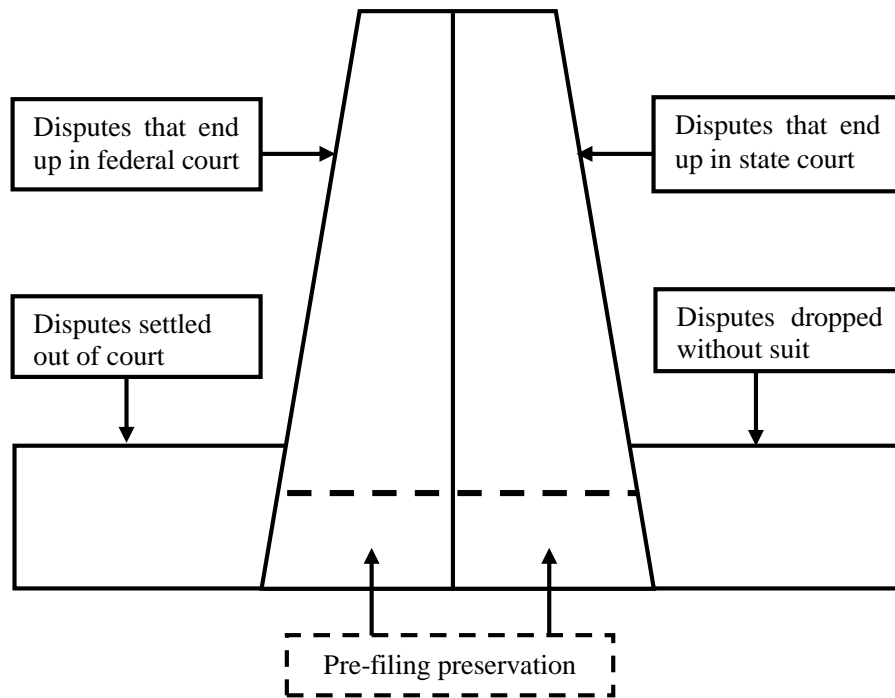
84. See *supra* Figure 2. From Figure 4, it is clear that Figure 2 is not to scale. If this were drawn to scale, the brim would be even wider, and the top would be very narrow.

FIGURE 4: NUMBER OF PAGES OR DATA EQUIVALENT (IN 1000S) PRESERVED, COLLECTED, AND PROCESSED FOR MICROSOFT CORPORATION



As the disproportionate bulk associated with preservation becomes apparent, the sense of urgency for new Rules governing preservation becomes obvious. But why is the “discovery sombrero” a sombrero? Why such a wide base?

FIGURE 5: THE DISCOVERY SOMBRERO, WITH LOCUS OF DISPUTE



The answer requires one to approach the question of discovery from the perspective of the preserving party. This is an *ex ante* perspective, in which the preserving party must make decisions before any uncertainty about the legal claim is resolved.⁸⁵ From this perspective, preservation is not part of litigation at all, and at this point in time the preserving party is not dealing with “lawsuits,” but with “disputes.” These disputes may or may not turn into lawsuits, let alone federal lawsuits.⁸⁶ To illustrate this idea, Figure 5 divides the discovery sombrero by *where the dispute ends up*, rather than by stage of discovery.⁸⁷

85. See Easterbrook, *supra* note 1, at 639; Paul Stancil, *Balancing the Pleading Equation*, 61 BAYLOR L. REV. 90, 100 (2009).

86. See Spencer, *supra* note 10, at 2007–08 (discussing the level of foreseeability needed to trigger the duty to preserve).

87. See *supra* Figure 5. Note that the dashed line indicates that, among matters that end up in state or federal court, some preservation occurs before the matter becomes a filed lawsuit, and some preservation occurs after. The remaining stages all occur after filing, of course. Note, too, that as before, the sections of the sombrero are not to scale.

Survey respondents generally confirmed that a substantial portion of preservation activity is conducted in the absence of a filed lawsuit.⁸⁸ In contrast, collection, processing, review, and production will usually occur in the context of litigation.⁸⁹ This is the crucial difference between stages of discovery that federal courts have long regulated under the Rules, generally with success,⁹⁰ and preservation, a phase of discovery that currently vexes courts and litigants alike.

B. *The Preservation Iceberg*

This section focuses on the costs associated with preservation specifically. The costs of preservation fall into two broad categories, which will be referred to as *fixed costs* and *variable costs*.

Fixed costs are costs that do not depend on the volume of preservation activity or the number of cases a company faces.⁹¹ These costs are “fixed” because they do not arise in the context of individual litigation matters, but represent a company’s ongoing expenses. For example, the costs of developing a repository for e-mails preserved in anticipation of litigation will exist whether the company faces 100 lawsuits or 1000 lawsuits; the specific scope of preservation rules or the number of holds that will have to be issued will have little effect on this cost.⁹² Fixed costs include the costs of maintaining a staff of attorneys, IT specialists, and other professionals devoted to preservation activity, as well as the costs of automated systems to manage litigation holds and preserve data.⁹³ Virtually all prior reported information on the costs of preservation reflect only the fixed costs of preservation-related technology.

Variable costs of preservation are costs that arise in the context of individual litigation matters, and thus vary with the volume of preservation activity. The primary variable cost is the time that non-legal employees subject to litigation-hold obligations must divert from business activities to compliance with a litigation hold. This lost time is a variable cost because the time spent by an employee on litigation holds increases as either the number of holds rises or as the complexity of each hold rises.

The Preservation Costs Survey sought to quantify both fixed and variable costs of preservation. While collecting specific, quantitative estimates of fixed

88. Most companies do not track these numbers, but a few companies did provide such data. These reports ranged from forty-four to seventy-seven percent of holds not being associated with active litigation. *Preservation Costs Survey Report*, *supra* note 19, at 43.

89. *Id.*

90. See *Civil Rules Survey*, *supra* note 2, at 69–70.

91. See *Preservation Costs Survey Report*, *supra* note 19, at 7–11 (discussing in more detail fixed and variable costs).

92. See *Where the Money Goes*, *supra* note 5, at 86 (“[P]reservation responsibilities can sometimes involve enterprise-level costs, such as would be incurred with the implementation of an automatic legal-hold tool. Such applications are certainly costly and have an observable price tag, but the expenditures are spread across all of the company’s present and future preservation needs.”).

93. See *id.* at 85.

costs was infeasible for Phase III of the Survey, the author collected information on fixed costs in interviews with a number of large companies in Phases I and II. These interviews provided detailed information on several categories of fixed costs.

With respect to fixed costs associated with personnel, these companies all had legal IT or e-discovery groups with attorneys and paralegals devoted full- or part-time to preservation activity. On average, the companies had two attorneys and four paralegals or other legal professionals working full-time in a dedicated legal IT or e-discovery group.

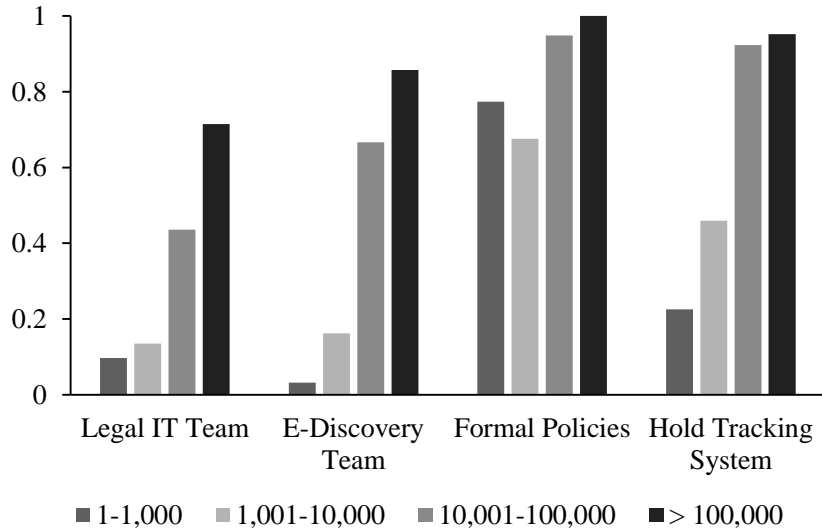
With respect to the fixed costs of technology, these companies also provided estimates of the costs of automated preservation systems that ranged from hundreds of thousands to tens of millions of dollars *per system*. One important type of system in this area is the automated litigation-hold management system. These systems automate the process of distributing, tracking, and monitoring litigation-hold notices that are created by in-house counsel. The largest fixed costs, however, are associated with the preservation of data itself. Every large company surveyed has a diverse set of systems used to address preservation obligations. Such an array of systems is necessary due to the large variety of types of ESI, many of which have distinct business purposes and are used and stored in different ways on a company's computer systems.

These reports suggest that for large companies, the per-year fixed costs associated with preservation activity run into the millions of dollars. Table 2 presents a rough but conservative calculation based on Survey results. The total (measurable) fixed costs of preservation for a single, large company exceed \$2.5 million per year. For smaller companies, though, fixed costs could be essentially zero. Most smaller companies do not report having a dedicated e-discovery team, legal IT function, or automated litigation hold system. See Figure 6.

TABLE 2: APPROXIMATE FIXED COSTS OF PRESERVATION FOR A HYPOTHETICAL LARGE COMPANY

Preservation Solution	Per Year Fixed Cost
A legal IT and/or e-discovery team	\$1,000,000
Litigation-hold management system (implementation cost amortized over a 5 year expected life)	\$160,000
Maintenance of litigation-hold management system	\$150,000
Automated data preservation system (implementation cost amortized over a 5 year expected life)	\$1,000,000
Maintenance of automated data preservation system	\$200,000
Total	\$2,510,000

FIGURE 6: SHARE REPORTING PRESERVATION PRACTICES, BY COMPANY SIZE



This striking difference is likely due to the simple fact that smaller companies face fewer lawsuits.⁹⁴ The fixed investments of many large companies benefit from large economies of scale.⁹⁵ By leveraging legal and technical expertise and automation, investment in these fixed costs lowers the per-matter (variable) cost of preservation activities. However, the high up-front investment is only justified by a large volume of litigation. As Figure 6 shows, while the very largest companies almost uniformly use in-house preservation experts for managing litigation holds, virtually none of the smallest companies have separate legal IT or e-discovery staff. Similarly, automated litigation-hold tracking software is virtually standard practice among large companies but is uncommon among smaller companies.

As high as the fixed costs of preservation may be, the largest share of preservation costs is variable: the costs in human time and effort to address preservation obligations on a case-by-case basis. Individual employees placed on hold or otherwise asked to engage in preservation activities must divert time and attention away from normal business activities. In this respect, a day spent responding to litigation-hold notices is just as significant a drain on worker productivity as a sick day.

Prior to this Survey, the magnitude of this aspect of preservation costs was unknown.⁹⁶ The following calculation is used to determine the magnitude of these costs: multiply the number of litigation matters per year,⁹⁷ times the number of employees on litigation hold per matter,⁹⁸ times the number of hours

94. Companies with 1–1,000 employees reported an average of 27 active lawsuits (scaled proportionally to company size) at the time of the survey. *Preservation Costs Survey Report*, *supra* note 19, at 33 tbl.8. Companies with 1,001–10,000 employees reported an average of 228 lawsuits; companies with 10,001–100,000 employees reported an average of 2,563; and companies with more than 100,000 employees reported an average of 3,404. *Id.*

95. For example, one interviewed company spent around one million dollars to implement and maintain software to assist in indexing and searching preserved data, but the interviewee saw this cost as a fraction of the savings it has generated. *Id.* at 34 n.52.

96. See *Preservation Costs Survey Report*, *supra* note 19, at 28–31 (discussing in more detail the calculation of the variable costs of preservation).

97. This number was taken directly from data reported in responses to Phase III of the Survey.

98. This figure was derived from Phase I and Phase II companies that provided detailed data based on litigation-hold tracking software. Based on the companies that provided data for this calculation, about 0.12% of employees on average are subject to a hold for each litigation matter. This estimate was applied to all companies, with a minimum of five employees per hold for smaller companies. As a check on this latter figure, an analysis was conducted of employment cases, defined in the Survey as employment discrimination and retaliation cases, to test the validity of extrapolation from larger companies to smaller companies. Unlike certain categories of litigation that uniquely affect large companies (such as antitrust), employment litigation is a risk for companies of all sizes, and typical employment suits at large companies look very much like employment suits at smaller companies in terms of stakes and numbers of “key players.” Focusing only on employment litigation data indicates that the assumption that smaller companies have an average of five employees on hold per litigation matter is actually a conservative estimate. *Id.* at 36–37.

per year a worker will spend on the litigation hold,⁹⁹ times the average salary per hour for employees in management occupations,¹⁰⁰ per Expression 1.

EXPRESSION 1:

$$\left(\frac{\$}{\text{Hr.}}\right) (\text{Hours}/\text{Worker}) (\text{Wkrs.}/\text{Hold}) (\text{Holds}/\text{Yr.}) = \$/\text{Yr.}$$

The estimates, broken down by company size, appear in Table 3. For companies of all sizes, the costs in lost employee time are significant. For the smallest companies in the Survey, the costs average over \$12,000 per company per year. The estimate of costs for the largest companies exceeds \$38.6 million per company per year. In sum, the two or three million dollars that a large company might spend in a year on preservation-related systems may be no more than the tip of the preservation iceberg.

99. This could include time spent reading, confirming receipt, and asking questions about a litigation-hold notice; time spent changing personal device settings and other work practices to comply with the litigation-hold notice; and time spent reviewing electronic and paper files to mark, copy, or set aside files for preservation. The calculation uses an estimate of three hours per employee per year spent on each litigation hold based on estimates reported by interviewed companies; however, no surveyed company had a precise estimate. As the Survey found, few companies track this type of cost.

100. This figure, \$52.20/hour, is the average hourly wage of workers in management occupations (across all sectors and all business sizes) provided by the latest data from the Bureau of Labor Statistics. See *May 2012 National Occupational Employment and Wage Estimates*, BUREAU OF LABOR STATISTICS, http://www.bls.gov/oes/2012/may/oes_nat.htm#11-0000 (last visited April 19, 2015).

TABLE 3: ESTIMATED PER-COMPANY COSTS OF EMPLOYEE TIME LOST TO LITIGATION HOLDS, BY COMPANY SIZE¹⁰¹

(1)	(2)	(3)	(4)	(5)
Total Employees	Matters with Holds	Employees per Matter	Employee Hours per Year	Time Cost per Year
1–1,000	16	5	240	\$12,528
1,001–10,000	249	11	8,217	\$428,927
10,001–100,000	1,245	71	265,185	\$13,842,657
> 100,000	1,333	185	739,815	\$38,618,343

C. The Long Tail of Costs

One question that existing, yet entirely anecdotal, evidence cannot answer is whether the cases that have high preservation costs are typical or atypical. There have been many anecdotes suggesting that preservation burdens are large, and many anecdotes suggesting that they are not large.¹⁰² The Survey data reveals that these conflicting anecdotes do not pose a credibility contest between two contrary accounts. Instead, these divergent anecdotes on cost reflect different aspects of the *same* phenomenon—an enormous amount of preservation activity that is very unevenly distributed across litigation matters. This is the context in which individual experiences with the costs of preservation must be understood.

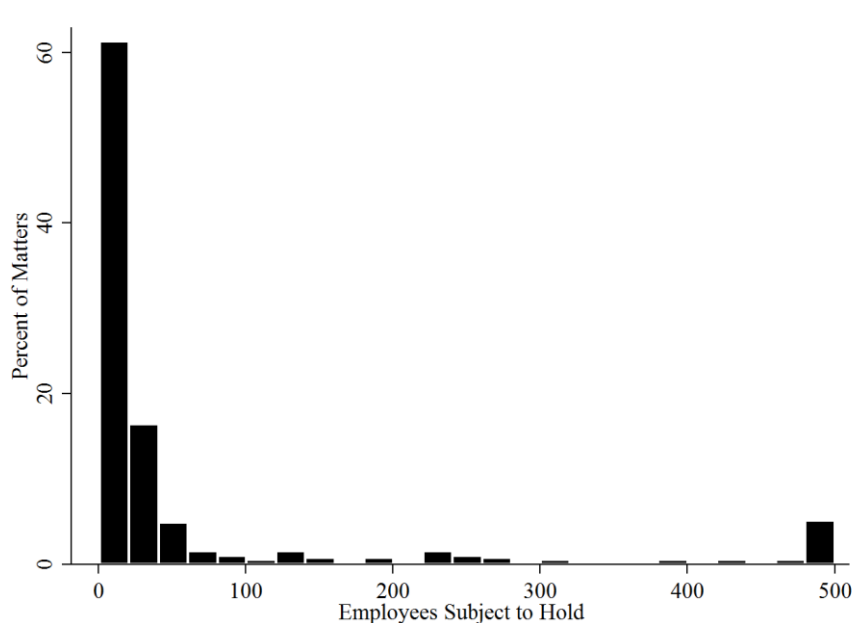
Detailed, case-by-case data was collected on litigation holds from six companies. These data sets together include information on over 3,600 separate litigation matters and over 770,000 individual litigation-hold notices. The data from a representative company appears in Figure 7; histograms for the remaining five companies look essentially the same. This company's dataset covers 390 distinct matters representing actual or anticipated civil litigation. For each matter, the dataset provides the number of individuals subject to a litigation hold.

101. Columns (2) and (3) are averages within the company size categories in column (1). Column (4) is the product of column (2) and column (3) times 3 hours per employee-hold. Column (5) is column (4) times \$52.20 per hour.

102. See *supra* notes 2–3.

Figure 7 shows the frequency with which litigation matters involve a given number of employees subject to holds. For example, the left-most vertical bar in Figure 7 represents the number of matters with twenty employees or fewer on hold, the next bar indicates the number of matters with twenty-one to forty employees on hold, and so on.¹⁰³

FIGURE 7: DISTRIBUTION OF EMPLOYEES ON HOLD PER MATTER OF A REPRESENTATIVE LARGE COMPANY, TOPCODED AT 500



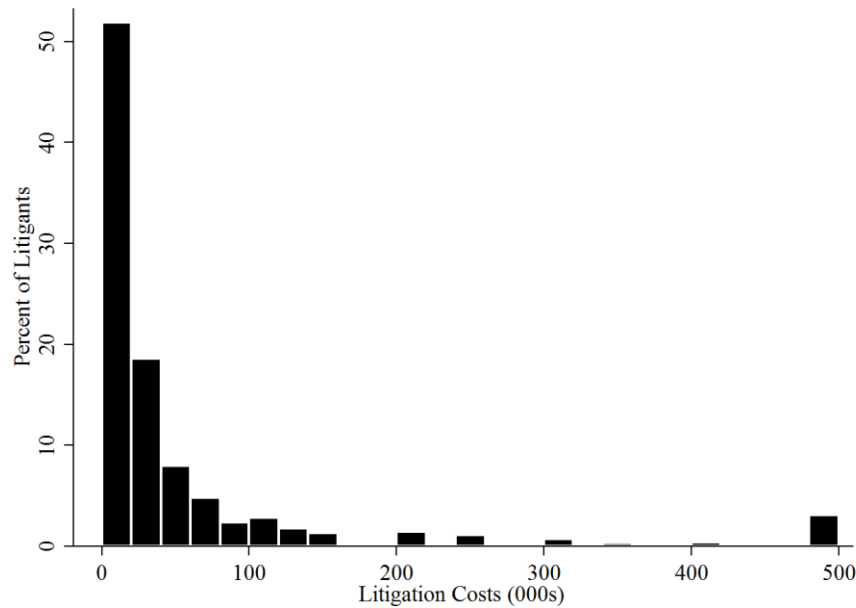
As Figure 7 shows, most litigation matters involve litigation holds affecting relatively few employees—well over half of the matters had twenty holds or fewer. Yet, the distribution of litigation holds across matters is highly skewed, and there is a “long tail” of matters in which huge numbers of employees are placed on hold in each case. This means that a small percentage of litigation matters can account for the bulk of all litigation-hold activity. Indeed, across sampled companies, *five percent of matters account for more than fifty-two percent of litigation-hold notices issued.*¹⁰⁴

103. For graphical clarity, the distribution of the number of employees on hold per matter in Figure 7 is topcoded at 500. Matters with more than 500 employees subject to hold are included in the right-most vertical bar.

104. See *Preservation Costs Survey Report*, *supra* note 19, at 40 tbl.11. This pattern also holds for employment litigation specifically. Across companies for which the data is available, the top five percent of employment-related matters account for almost exactly half (49.5%) of all employment-related litigation holds. See *id.* at 41 tbl.12.

Notably, the patterns that appear in the Preservation Costs Survey data resemble the patterns that other researchers have found in the context of outside counsel's litigation costs. In their *Civil Rules Survey*, Lee and Willging found that while the *median* case had relatively low litigation costs (\$15,000 for plaintiffs and \$20,000 for defendants), the 95th percentile case involved costs of approximately \$300,000 for each party.¹⁰⁵ Figure 8 presents the distribution of litigation costs from the *Civil Rules Survey*.

FIGURE 8: DISTRIBUTION OF LITIGATION COSTS PER PARTY, FJC CIVIL RULES SURVEY DATA



Strikingly, even though Figure 7 describes preservation costs borne by the client and includes matters not filed in court while Figure 8 assesses litigation costs incurred by outside counsel for filed lawsuits, the patterns are virtually identical. Also parallel is the fact that in the *Civil Rules Survey* data, the top five percent of cases accounted for more than half of all litigation costs.¹⁰⁶ Notably, this long tail of litigation costs is the product of *discovery* costs, not the product

105. *Civil Rules Survey*, *supra* note 2, at 35–37.

106. To be precise, 5% of all cases accounted for 59.4% of defendants' total litigation costs across all cases. Author's calculations (available upon request from author) are based on data from the *Civil Rules Survey*. The author thanks Emery G. Lee, III for sharing the *Civil Rules Survey* data with him.

of trials, motion practice, or the like. Federal Judicial research indicates that while most cases have little or no discovery costs, in the fraction of cases that do actively involve discovery, discovery often accounts for the vast majority of all costs.¹⁰⁷ This suggests that the long tail of costs is a phenomenon that broadly describes all phases of preservation and discovery.

In sum, these results reveal that preservation costs are not high in most cases, but the distribution is highly skewed, with a long tail in which a relatively small number of highly complex and burdensome cases account for a large share of the total costs. Thus, the “typical” case is no cause for concern, but the “average” case may have a very high level of preservation activity because a small but important number of cases substantially drive up the total costs of preservation.

IV. THREE IMPLICATIONS FOR LAW AND RULEMAKING

A. *The Sombrero and the Erie/Hanna Boundary*

As noted above, the Federal Rules have never mentioned the duty to preserve nor placed any requirements on what data litigants retain—only what they produce. Nonetheless, current federal case law has created rules governing preservation that impose preservation obligations on parties even before a suit is filed.¹⁰⁸ This case law is almost exclusively the product of the district courts, as discovery-related orders and most orders imposing sanctions for failure to preserve are interlocutory and, therefore, not appealable.¹⁰⁹ This raises the doctrinal question of whether a wholly judge-made federal common law of preservation is consistent with the *Erie* doctrine.

Under the *Erie* doctrine, federal courts cannot create rules of decision through federal common law, but must decide cases by interpreting and applying state substantive law or codified federal law.¹¹⁰ Underlying this decision was the recognition of the principle, embodied in the Rules of Decision Act, that “[e]xcept in matters governed by the Federal Constitution or by acts of Congress,

107. Stancil, *supra* note 85, at 100 (“According to one recent study, discovery consumes approximately 50% of all federal litigation expenditures; moreover, that study noted that discovery can account for ‘as much as 90% of the litigation costs in the cases where discovery is actively employed.’”) (citing Memorandum from Paul V. Niemeyer, Chair, Advisory Committee on Civil Rules, to Hon. Anthony J. Scirica, Chair, Committee on Rules of Practice and Procedure (May 11, 1999) (192 F.R.D. 340, 357)).

108. See Spencer, *supra* note 10, at 2005–08 (discussing varying approaches to pre-litigation preservation obligations among district and circuit courts and their impact on potential, not actual, litigants).

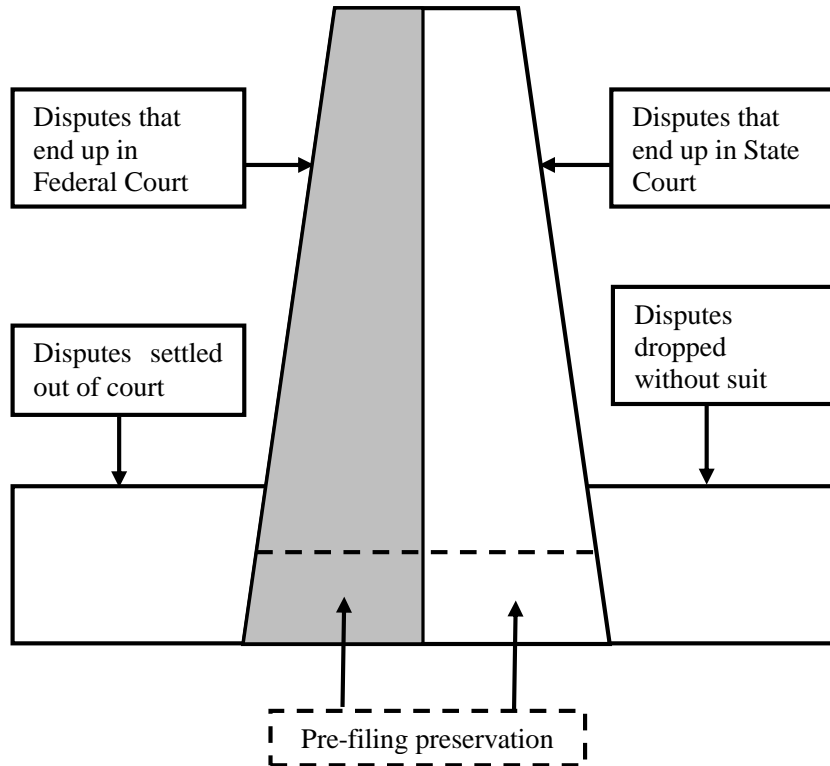
109. See Cassandra Burke Robertson, *Appellate Review of Discovery Orders in Federal Court: A Suggested Approach for Handling Privilege Claims*, 81 WASH. L. REV. 733, 735 (2006) (describing the various approaches courts take toward permitting review of orders for the discovery of allegedly privileged information).

110. *Erie R.R. Co. v. Tompkins*, 304 U.S. 64, 78 (1938).

the law to be applied in any case is the law of the state. . . . There is no federal general common law.”¹¹¹

The notion that a federal common law of preservation might trample upon the domain of state substantive law, thereby running afoul of the *Erie* doctrine, is hardly obvious. Preservation seems procedural in nature, and every court has inherent power to ensure the integrity of its proceedings and judgments.¹¹² Therefore, a federal judge would naturally assume she can invoke judge-made federal law to regulate the pre-filing preservation activity of litigants in federal court.¹¹³ It would seem almost tautological that if a federal judge is asked to enforce preservation law, then the subjects of that law are parties in federal court whose actions are governed by federal procedures for conduct in court. The shaded area in Figure 9 illustrates this view of the scope of the federal law on preservation.

FIGURE 9: THE PERCEIVED SCOPE OF THE DUTY TO PRESERVE



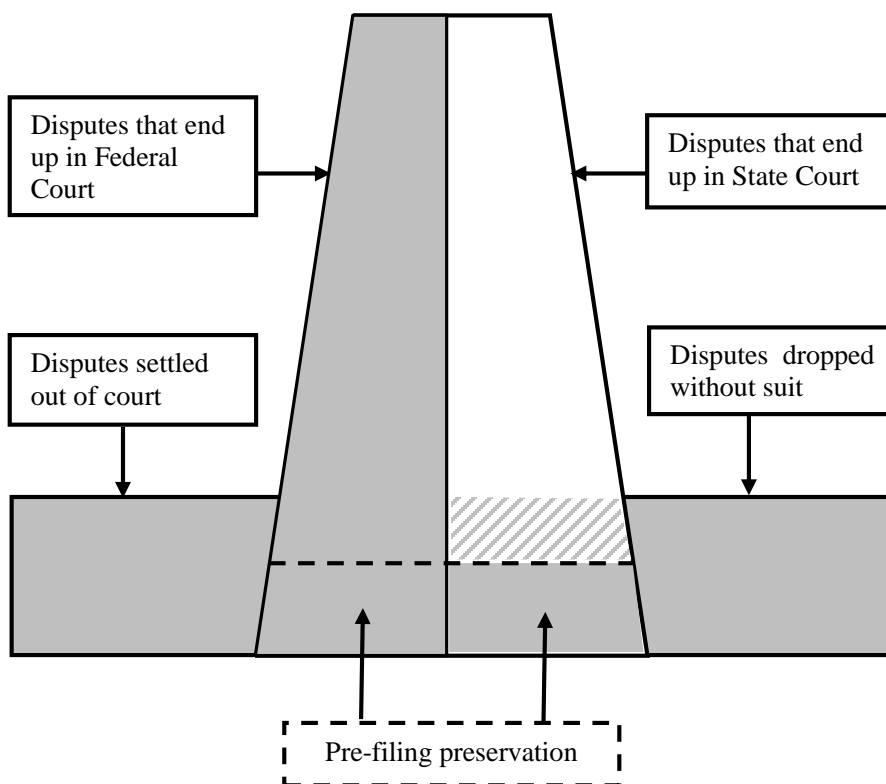
111. *Id.* See also Rules of Decision Act, 28 U.S.C. § 1652 (2012).

112. Amy Coney Barrett, *Procedural Common Law*, 94 VA. L. REV. 813, 833–35 (2008) (describing differing views about the judicial branch’s inherent control over judicial procedure relative to that of Congress).

113. *Cf. id.* at 834 n.65.

The problem is that preservation decisions are often made *ex ante*, i.e., before a lawsuit is filed.¹¹⁴ If federal rules governing preservation apply to pre-filing conduct, then the federal rules govern the conduct of parties in *every* dispute that *could* end up in federal court, including those disputes that ultimately end up in state court or are never litigated at all. The shaded area in Figure 10 represents the extent to which federal preservation law could affect cases. (The partially shaded area denotes the possibility of federal preservation duties affecting cases filed in state court while there is still a possibility of removal to federal court).

FIGURE 10: THE EFFECTIVE SCOPE OF THE DUTY TO PRESERVE



Of course, federal law regulates the conduct of companies and individuals outside of federal court in many instances, including behavior that affects preservation and discovery.¹¹⁵ For example, laws define what happens when a

114. See, e.g., Spencer, *supra* note 10, at 2006–11 (exploring the standards courts use to determine whether parties’ duties to preserve have been triggered, which are triggered prior to filing the lawsuit).

115. See *id.* at 2006 n.7–8.

witness is murdered and exactly how long and on what kind of media one must store certain emails.¹¹⁶ But these are substantive laws enacted by Congress or regulations promulgated by administrative agencies pursuant to statute, in the same way that laws proscribing the murder of non-witnesses or regulations addressing the proper storage of meat are substantive laws that dictate conduct outside of litigation. In contrast, the federal law of preservation considered in this article is entirely a creature of common law, yet it governs the conduct of parties outside of federal court and even in the absence of any litigation, state or federal. The fact that a number of states have created independent causes of action for the tort of spoliation of evidence further reveals the extent to which the overarching federal common law of preservation conflicts with the *Erie* doctrine.¹¹⁷

Preservation decisions are, by and large, made prior to litigation when a potential defendant faces great uncertainty about where, if at all, it will be sued.¹¹⁸ Consequently, plaintiffs can exploit this legal variation to their advantage, choosing the forum “with the most demanding requirements of the toughest court to have spoken.”¹¹⁹ Nor can the process of appellate review iron out these stark and seemingly arbitrary variations in preservation standards because these standards are the product of non-appealable interlocutory rulings by district courts.¹²⁰ To use the idioms of *Erie* and *Hanna*, these variations across courts invite “forum-shopping” and the arbitrary and uncertain application of these conflicting precedents leads to “inequitable administration of the laws.”¹²¹

Still, preservation in some sense is procedural, and it would be absurd to deny federal courts any say in the preservation activity of litigants. With the benefit of a clearer picture of the reach of current preservation law beyond the federal courtroom, one sees that there remain two complementary ways forward that restrict federal courts’ inherent, common law powers to their proper domain without requiring more restraint than necessary when policing spoliation.

The first of these is through the federal rulemaking process. Federal judicial power under the Rules stems from the Rules Enabling Act, which authorizes the federal courts to enact rules of “practice and procedure” that do “not abridge, enlarge or modify any substantive right.”¹²² Importantly, because the Rules

116. See 18 U.S.C. § 1513 (2012) (murder of a witness); Commodities and Securities Exchanges, 17 C.F.R. § 240.17a-4 (2012) (recording retention obligations for securities exchange members, brokers, and dealers).

117. See MARGARET KOESEL, ET AL., *SPOILIATION OF EVIDENCE: SANCTIONS AND REMEDIES FOR DESTRUCTION OF EVIDENCE IN CIVIL LITIGATION* 50 (Daniel F. Gourash, ed., 2000).

118. See Spencer, *supra* note 10, at 2007 (explaining that a potential litigant’s duty to preserve is triggered “prior to the initiation of litigation”).

119. Victor Stanley, Inc. v. Creative Pipe, Inc., 269 F.R.D. 497, 523 (D. Md. 2010), *aff’d* No. 2012-1638, 2012 WL 616939 (Fed. Cir. Feb. 20, 2013).

120. See, e.g., Mohawk Industries, Inc. v. Carpenter, 558 U.S. 100 (2009).

121. *Hanna v. Plumer*, 380 U.S. 460, 468 (1965).

122. 28 U.S.C. § 2072 (2012).

Enabling Act and the Rules promulgated thereunder are *codified* federal law, *Erie* by its terms does not apply.¹²³ Rather, the Supreme Court has treated the Rules Enabling Act as creating a more forgiving standard for legitimate judicial lawmaking through the Rules.¹²⁴ *Sibbach* and *Hanna* merely ask whether the Federal Rule “really regulates procedure” regardless of whether it also affects substantive rights.¹²⁵ It is self-evident that the preservation law “really regulates procedure.”

The second method is to limit the scope of federal court regulation of pre-litigation preservation activity to solely pre-litigation activity that is *directed toward* the court. It would then become difficult to question using the inherent power to punish litigants who—in anticipation of a lawsuit—deliberately act to undermine the integrity of the court proceedings. This is exactly the sort of behavior against which the Supreme Court permitted the deployment of inherent power in *Chambers v. NASCO, Inc.*¹²⁶ While the misconduct in *Chambers* occurred out of court, and prior to the filing of a federal lawsuit, the Court emphasized the bad faith and intentional character of conduct whose *purpose* was to frustrate the anticipated federal proceedings.¹²⁷

As applied to preservation, this is the approach taken by the Southern District of Texas in *Rimkus Consulting Group, Inc. v. Cammarata*,¹²⁸ which held that bad faith was required for the court to issue sanctions for a pre-litigation failure to preserve.¹²⁹ But the law is sharply divided on this point, and a bad faith requirement is not applied consistently under current law.¹³⁰ A more widely cited case from the Southern District of New York, *Pension Committee of University of Montreal Pension Plan v. Banc of America Securities, LLC*,¹³¹ imposed sanctions for pre-litigation failure to preserve based on gross negligence.¹³²

123. See Spencer, *supra* note 10, at 2031–32.

124. See *Shady Grove Orthopedics Assocs. v. Allstate Ins. Co.*, 559 U.S. 393, 407 (2010).

125. *Hanna*, 380 U.S. at 464; *Sibbach v. Wilson & Co.*, 312 U.S. 1, 14 (1941). This rule was followed in *Shady Grove*, albeit only by a plurality of the court. See *Shady Grove*, 559 U.S. at 407 (2010).

126. 501 U.S. 32, 44 (1991).

127. *Id.* at 37, 50–51 (affirming the district court’s imposition of sanctions on a party whose “entire course of conduct . . . evidenced bad faith and an attempt to perpetuate a fraud on the court,” including contact that took place before the lawsuit was filed).

128. 688 F.Supp.2d 598 (S.D. Tex. 2010).

129. *Id.* at 614.

130. See Robert A. Weninger, *Electronic Discovery and Sanctions for Spoliation: Perspectives from the Classroom*, 61 CATH. U. L. REV. 775, 790 (2012) (explaining that the federal circuits differ by requiring either negligence, gross negligence, or bad faith before issuing sanctions).

131. 685 F.Supp.2d 456 (S.D.N.Y. 2010) (*abrogated in part* by *Chin v. Port Auth. of New York & New Jersey*, 685 F.3d 135, 162 (2d Cir. 2012) (holding that failure to institute a litigation hold is not gross negligence *per se*)).

132. *Id.* at 496.

B. The Iceberg and the Law/Technology Boundary

The preservation iceberg reveals that preservation is a much more expensive, higher-stakes stage of discovery than was previously understood. It also negates the belief that technology alone can reduce the costs and burdens of discovery, especially in the context of preservation. An irony of the debate on preservation costs is that the most often cited costs of preservation—technology costs—are only a small part of the problem, and for large companies, reducing the burdens of preservation will probably involve *further increasing* technology costs. Total burdens will fall as human costs are reduced, even as (and precisely because) the most salient costs of preservation—big ticket technology spending—will rise. But that is the good news.

The bad news is that the iceberg is primarily a large-company phenomenon. For smaller companies, there is no “tip of the iceberg.” All or nearly all of their preservation costs are human costs. Without the scale of litigation activity that justifies the fixed costs associated with litigation-hold management systems—data vaults, and legal IT staff—the costs borne by smaller businesses are not technology costs at all, but human costs.

This helps explain the Survey result that smaller companies and larger companies reported similar burdens. Precisely because they are smaller and face fewer lawsuits, it is not generally cost effective for smaller companies to make expensive, but beneficial, investments in sophisticated automated systems or in-house expertise. Thus, although preservation disputes or spoliation allegations are rarer for smaller companies, these companies are also far less equipped to handle these controversies.

This is true even for companies in the technology field itself; one respondent, a tech company with about 100 employees, explained:

We are a small company, but we are in a space where we need to protect our IP and also to prevent customers from eluding payment. We manage most of the process in house, but it is a huge burden on our IT. We are looking at vaulting solutions for e-mail, which should be a big help. But the costs are enormous, and vendors are unwilling to give us a good demo or trial vault.¹³³

Instead, smaller companies may have to rely on ad-hoc, outside assistance, which may be less efficient and more expensive on a per-case basis. One respondent, an industrial company with about 200 employees, explained:

Our company along with every other company in our industry is involved in several suits concerning one toxic tort-related issue. We are a very small player in this field. Yet, we have to produce the same documents as the big guys. In our case, our IT employee, our President, our Accountant, our Attorney, etc. [have] to devote all of

133. Survey Response of Respondent 2865509178 (Anonymous ID).

their time to answer discovery. We also employ an outside law firm at an hourly rate to help us. It is very costly.¹³⁴

Thus, technology is not a substitute for legal reform because technological solutions are only practical for the largest companies for which the economies of scale from high-tech solutions justify their high price tags. For smaller companies—those with hundreds of employees—technology plays a much smaller role in the preservation process. From this point of view, legal innovation, rather than technological innovation, may be the best hope for controlling the costs of individuals, small businesses, and NGOs, and virtually every potential litigant other than the largest and most sophisticated litigants.

C. *The Long Tail and Transsubstantivity/Tailoring Boundary*

The fact that most of the costs of preservation are generated by a small fraction of cases suggests that it may be productive to devise Rules to control preservation costs and focus those Rules on particular categories of large, information-intensive cases. This may require steps away from a commitment to transsubstantivity to which the Rules generally adhere. But there are ways to design discovery rules to address cost that neither sort cases into substantive categories nor require judges to do so. The “long tail” provides us with a way to do this. Because preservation and discovery costs are highly skewed, the Federal Rules can set presumptive, quantitative limits on the scale of preservation and discovery such that the “typical” case is unaffected but the court and the parties have levers for controlling litigation costs in particularly large or complex cases.

Parties should be provided tools to reduce discovery costs because active judicial oversight of discovery, although widely praised as highly effective, rarely occurs.¹³⁵ Close judicial oversight of preservation is even less feasible given that the duty to preserve may trigger before a suit is even filed.¹³⁶ Furthermore, the fog of litigation is greatest at the outset of a case; yet this is precisely when judicial supervision of preservation would be required.¹³⁷

Thus, it is essential that the parties have tools for negotiating the scope of preservation and discovery. It is not enough to rely on negotiations to arise

134. Survey Response of Respondent 2867300205 (Anonymous ID). Note that this company refers to its “IT employee” in the singular. *Id.*

135. *Cf.* Easterbrook, *supra* note 1, at 638–39 (finding the judicial oversight ineffective and impractical).

136. *See* Surowiec v. Capital Title Agency, Inc., 790 F.Supp.2d 997, 1005 (D. Ariz. 2011).

137. Judges may also have little incentive to monitor preservation and discovery costs; if high discovery costs induce settlement, as many models of litigation predict, and judges prefer leisure to effort, then active case management not only imposes the direct and immediate cost of effort on the part of the judge, but it increases the likelihood that the case will not settle, which requires further judicial effort. *See, e.g.*, LEE EPSTEIN ET AL., *THE BEHAVIOR OF FEDERAL JUDGES* 31 (Harvard Univ. Press 2013); William H.J. Hubbard, *Nuisance Litigation* 1–2 (April 1, 2014) (unpublished manuscript) (on file with Univ. of Cal.-Berkeley School of Law), http://scholarship.law.berkeley.edu/law_econ/Spring2014/Schedule/12/.

organically or to require negotiations by rule. In those cases where negotiating is productive, the parties have an incentive to negotiate regardless of any requirement to do so, and in those cases where it is not productive, requiring negotiations to occur will not make it so. The question remains—how can the Rules get parties to sit down and address preservation in a cost-effective way? To some extent, there is little that the Rules can do because for most disputes the duty to preserve attaches before a lawsuit is filed, and in many cases before an opposing party contacts the preserving party.¹³⁸ At least initially in these disputes, the scope, and therefore costs, of preservation are set without any opportunity for parties to work together to control these costs.¹³⁹

To the extent that it is feasible for the parties to work together to control preservation costs, the Rules can give the parties incentives to negotiate a scope of preservation that prioritizes important data while attending to cost concerns as well. Under the current Rules, the parties do not have any incentive to negotiate. Because courts lack the information to conduct a careful balancing of the costs and benefits envisioned by Rule 26, the default rule under current law is usually to place no firm limits on preservation or discovery.¹⁴⁰

How does this affect incentives? Consider the scenario where an individual plaintiff (or putative class action representative) sues a large company. The plaintiff has essentially no data relevant to the case, but the company has vast quantities of data, some of which may be relevant to the case and some of which may not. The plaintiff's attorney has little incentive to agree to reasonable limits on the scope of preservation in order to save costs, as the plaintiff will have no preservation costs in any event. The problem is that the parties do not have anything over which to negotiate.¹⁴¹

Compared to other stages of discovery in “asymmetrical” litigation, preservation is especially problematic. A plaintiff with no information still has an incentive to limit discovery requests for production because larger production increases the plaintiff's own costs of review. But a broad demand for preservation has no such self-correcting feature; the costs are borne entirely by the defendant.

If both sides to a dispute have similar preservation burdens, then there certainly is something to negotiate over. Each party can agree to preserve only the documents most likely to be relevant and not preserve the rest. Each then saves a lot of time, money, and aggravation at the cost of a small potential loss

138. See *Surowiec*, 790 F.Supp.2d at 1005.

139. DISCOVERY SUBCOMM., *supra* note 13, at 4. As one in-house counsel put it, “I can’t talk to opposing counsel because there is no opposing counsel.” *Id.*

140. See *supra* notes 35–39 and accompanying text.

141. There is a qualification to this: The defendant and the plaintiff *could* negotiate over how much the defendant has to pay the plaintiff in order to avoid the unconfined duty to preserve. But this sort of negotiation—usually referred to as a “nuisance settlement”—is definitely *not* the kind of negotiation that the author suspects the Rules aspire to encourage. See Hubbard, *supra* note 137, at 2.

in the number of relevant documents. This is exactly the sort of sound cost-benefit analysis that the Rules anticipate. The agreed-upon scope of preservation may be over- or under-inclusive, but the costs and benefits are symmetrical and, more importantly, agreed upon—and therefore settled and insulated from wasteful second-guessing down the road.

If this analysis is correct, then the majority of preservation headaches would arise in the context of “asymmetrical” litigation where one side has little or nothing to preserve, and the other has large quantities of data. Using individual-versus-company (as opposed to company-versus-company) litigation as a proxy for asymmetrical litigation, the Survey’s findings confirm this prediction.

TABLE 4: PRESERVATION-RELATED PROBLEMS BY OPPOSING PARTY TYPE (5 = “VERY OFTEN” AND 1 = “VERY RARELY”)

Configuration of Parties	Average Rating
Large, complex matters, individuals on other side	3.81
Large, complex matters, businesses on other side	3.45
Small, routine matters, individuals on other side	2.99
Small, routine matters, business on other side	2.59

Table 4 presents the results from the Survey.¹⁴² Respondents reported higher levels of preservation-related problems in litigation against individuals than in litigation against other businesses. Further, these differences are highly statistically significant.¹⁴³

The key, therefore, is to structure the Rules so that in cases where costs are likely to be large, *both* parties have something to lose and something to gain in preservation negotiations. The Rules governing production in discovery already do this. Rule 30 sets presumptive limits on the number and length of

142. A total of 122 respondents provided responses to this set of questions on the frequency of preservation-related problems based on the type of litigation and opposing party.

143. Using these results, the Survey tested two hypotheses using paired, two-tailed t-tests: (1) among large, complex matters, the means for cases against individuals and for cases against businesses are (statistically) the same, and (2) among small, routine matters, the means for cases against individuals and for cases against businesses are (statistically) the same. Both hypotheses are rejected at the one percent level.

depositions.¹⁴⁴ Rule 33 sets presumptive limits on the number of written interrogatories.¹⁴⁵ To avoid obvious injustices, Rule 26(b)(2)(A) permits the court to issue an order altering these presumptive limits.¹⁴⁶ In practice, however, most exceptions to these presumptions are negotiated by the parties and not determined by judicial order. Meaningful negotiation occurs in this context because the Rules, by construction, ensure that both sides of any discovery dispute have bargaining chips, as they can agree to more or fewer depositions.

Thus, in the preservation and document discovery context, a similar approach should be effective. Establishing a presumptive limit of fifteen to twenty custodians to be subject to litigation holds ensures that every party to a preservation dispute has bargaining chips. In most cases, this presumptive limit will be uncontroversial and not disturbed. In the cases where it is controversial, even a party with no preservation obligations itself will have an incentive to make meaningful rather than outlandish preservation demands because the other party now has a bargaining chip. A party can offer to preserve more in exchange for cost-justified concessions with respect to other aspects of discovery, or in exchange for cost-sharing between the parties.

The long tail of litigation costs indicates that Rules based on presumptive limits can be calibrated to leave the large numbers of cases with modest costs and few discovery disputes untouched, while directing parties' efforts, and potentially courts' attention, to the smaller set of cases with high costs. Presumptive limits can be set relatively low, but still impose no binding constraints on parties in most disputes. Existing Rules addressing depositions appear to do this already. For example, data collected by Emery Lee III and Thomas Willging indicates that in most cases, surveyed attorneys deposed five or fewer non-expert witnesses, well below the limit of ten set by Rule 30.¹⁴⁷

V. CONCLUSION

Until now, knowledge of the costs of preservation and discovery depended on anecdote and speculation. This research on preservation costs has made a first step toward a more rigorous, quantitative understanding of how preservation activity is distributed across cases and how its various costs stack up against each other. The discovery sombrero, the preservation iceberg, and the long tail of costs serve as basic, stylized facts in this regard.

These facts also serve to inform legal and policy debate. This includes, most immediately, currently pending amendments to the Federal Rules, which, if adopted, would expressly address preservation for the first time. As the

144. FED. R. CIV. P. 30(a)(2); FED. R. CIV. P. 30(d) (providing that each party may take no more than ten depositions, each of which may be no longer than seven hours).

145. FED. R. CIV. P. 33(a) (providing that each party may serve no more than twenty-five written interrogatories upon another party).

146. FED. R. CIV. P. 26(b)(2)(A).

147. *Civil Rules Survey*, *supra* note 2, at 10.

discussion of the discovery sombrero and *Erie* makes clear, moving the locus of federal lawmaking in this area from federal common law to rulemaking under the Rules Enabling Act is a welcome development. Additionally, as the discussion of the preservation iceberg makes clear, the decision of the rulemaking committees to act now rather than wait for technological solutions is good news for the vast majority of parties who cannot afford to use the high fixed costs of technology to control the costs of preservation.

Of course, the merits of exactly *how* the Rules address preservation is up for debate. In the discussion of the long tail of costs, this article advocates for an approach that would create clear, presumptive limits on the scope of preservation and discovery in order to encourage mutually beneficial bargaining that would define the proper scope of preservation and discovery in cost-intensive cases. Such an approach is admittedly far from perfect because it does little to address preservation costs that arise before the parties join issue in court, but it is an important potential path not yet taken by the rulemakers.

More importantly, though, any meaningful assessment of the merits of standards governing preservation must take into account both the costs and benefits of preservation. While this research has begun to quantify the costs, quantifying the benefits of preservation is as elusive today as quantifying costs once was.

