

WRITTEN STATEMENT

of

William H. J. Hubbard
Assistant Professor of Law
University of Chicago Law School

for the hearing entitled

“The Costs and Burdens of Civil Discovery”

on

December 13, 2011

before the

Subcommittee on the Constitution
Committee on the Judiciary
United States House of Representatives

I. INTRODUCTION

As part of a larger reexamination of the Federal Rules of Civil Procedure (“Rules”), the Advisory Committee on Civil Rules is considering the possibility of amendments to the Rules that would govern discovery in civil litigation, and in particular the preservation of documents and electronically stored information (ESI).¹ This activity comes amid a widespread call for rules reform arising out of frustration with the cost of discovery and the patchwork of federal case law on preservation obligations. Many companies, generally companies who frequently find themselves defendants in federal court, argue that uncertainty over preservation obligations forces them to “over-preserve”—i.e., preserve more than a proper cost-benefit analysis would otherwise require. Over-preservation involves potentially large and otherwise unnecessary costs.

How serious are the problems of discovery costs and over-preservation? There is a wealth of anecdotal evidence that these costs can be enormous in some cases, and that the cost of discovery (including preservation) outweighs its benefit. Until recently, however, there was virtually no empirical data on the costs of discovery and preservation. New empirical work has begun to provide essential information on the nature and scope of discovery costs, including the costs of preservation.

The key studies on discovery costs are:

- Lee and Willging, *Civil Rules Survey* (2009),²
- Lawyers for Civil Justice, et al., *Litigation Cost Survey* (2010).³

With respect to preservation, I am leading the first major study of preservation costs, the Preservation Costs Survey, which was commissioned by the Civil Justice Reform Group.⁴ Preliminary results from this study have already yielded important findings, which I will describe below. I report these findings in greater detail in:

¹ Note that herein I will use “documents” and “information” interchangeably to refer both the paper records and ESI.

² Emery G. Lee III and Thomas E. Willging, *National, Case-Based Civil Rules Survey* 35 (Federal Judicial Center 2009). I will refer to this study throughout as the “*Civil Rules Survey*.”

³ Lawyers for Civil Justice, Civil Justice Reform Group, and U.S. Chamber Institute for Legal Reform, *Litigation Cost Survey of Major Companies* (Searle Center on Law, Regulation, and Economic Growth 2010) (herein, “*Litigation Cost Survey*”).

⁴ 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.

- Hubbard, *Preliminary Report* (2011),⁵
- Hubbard, *Letter to Judge Campbell* (2011),⁶
- Hubbard, *Preservation under the Federal Rules* (2011).⁷

These three documents are attached as Appendices to this Written Statement.

The goal of this Written Statement is to summarize the central findings of the Civil Rules Survey, the Litigation Costs Survey, and the Preservation Costs Survey, and provide analysis of how this data should inform the Rulemaking process. Part II provides a short overview of the discovery process to frame the following discussion. Part III discusses the data on discovery costs. Importantly, existing studies of discovery costs do not capture the costs of preservation. Part IV discusses the data on preservation costs. Part V addresses some policy implications for Rulemaking with respect to three aspects of preservation: trigger, scope, and sanctions. It then provides estimates of potential cost savings from new Rules and explains why some proposed alternatives to new Rules will not work. The analysis in this Written Statement will necessarily be brief; more detailed discussion can be found in the Appendices.

II. THE STAGES OF DISCOVERY

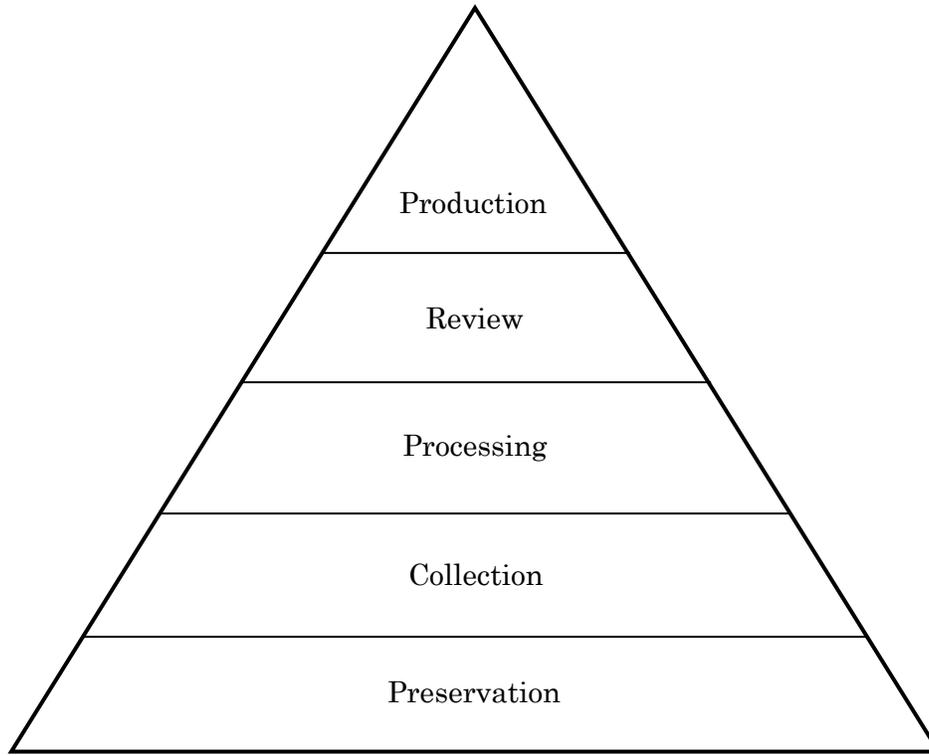
Discovery—the legal processes by which the parties unearth information to be used in a case—is typically divided into five stages. These stages are illustrated in the “discovery pyramid” in Figure 1 below. The discovery process begins with the *preservation* of information that may be relevant to ongoing 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 data. This renders data amenable to review and reduces redundancies and other costs further downstream. *Review* is the work of lawyers to determine relevance and privilege of the documents in discovery, and *production* turns over to the other side the relevant, non-privileged materials within the scope of discovery. The ultimate goal of discovery, of course, is the use in litigation of information valuable to the finder of fact.

⁵ William H. J. Hubbard, *Preliminary Report on the Preservation Costs Survey of Major Companies* (Civil Justice Reform Group Sept. 8, 2011) (Attached as Appendix A) (herein *Preliminary Report*).

⁶ William H. J. Hubbard, *Letter to the Hon. David G. Campbell* (Nov. 3, 2011) (Attached as Appendix B).

⁷ William H. J. Hubbard, “Preservation under the Federal Rules: Accounting for the Fog, the Pyramid, and the Sombrero,” unpublished working paper (Dec. 2, 2011) (Attached as Appendix C).

FIGURE 1: THE DISCOVERY PYRAMID



The pyramid shape is deliberately chosen and will be familiar to many practitioners. It indicates that not everything that is preserved is collected, and not everything that is collected is processed and reviewed, and so on. The question for the policymaker is whether the Rules governing discovery unnecessarily expand the base of the pyramid in a way that increases costs out of proportion to any benefit. I will return to this question shortly. First, I will summarize some data on the costs of discovery and preservation in particular.

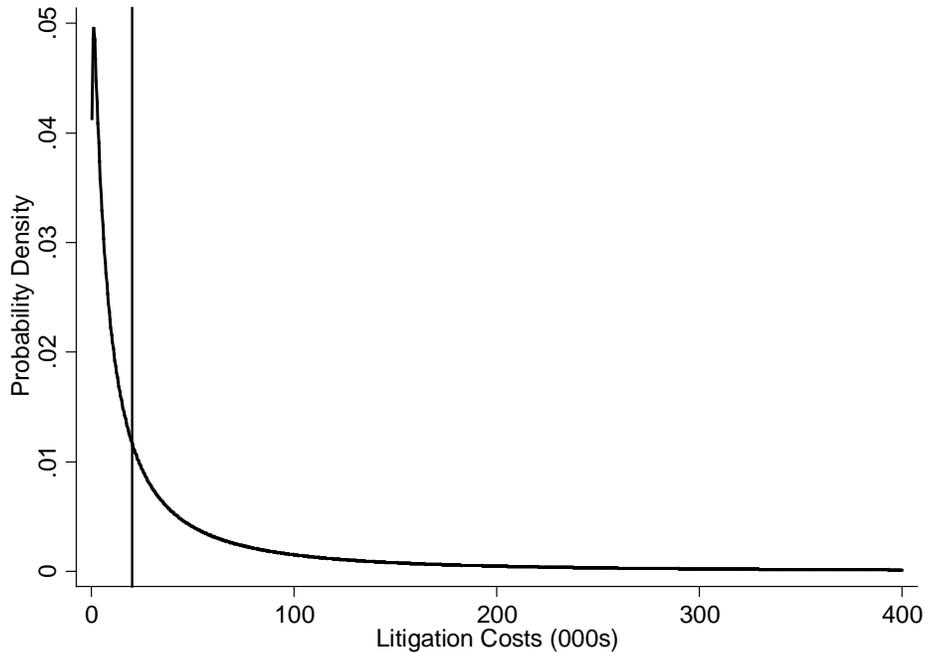
III. COMMENTS ON THE COSTS OF DISCOVERY

Two studies on litigation costs, the Civil Rules Survey and the Litigation Cost Survey, shed light on the role of discovery in the costs of litigation. The Civil Rules Survey covers a large sample of outside counsel from a broad cross-section of federal cases. Perhaps its most salient finding is that the median per-case cost of litigation to defendants in cases with discovery is \$20,000. (The median cost for plaintiffs is \$15,000.) And in this sample, the cost of discovery is maybe 30 percent of litigation costs. This result seems to suggest that discovery costs may not be a major problem.

The Litigation Cost Survey covers a different sample: in-house counsel at large companies were asked about the costs of their largest

lawsuits. The per-case average cost of discovery in this sample was over \$700,000. This result suggests that improving the efficiency of discovery could lead to cost savings for the economy.

FIGURE 2: LITIGATION COSTS WITH MEDIAN OF \$20,000,
GIVEN A LOG NORMAL DISTRIBUTION OF COSTS



While the studies' results appear to contradict each other, a careful analysis shows otherwise. A closer look at the Civil Rules Survey shows why: the 10th percentile of defendants' litigation cost, \$5,000, is one-fourth the median, but the 95th percentile, \$300,000, is fifteen times the median! In other words, it appears that the cost of litigation, and of discovery in particular, is a "long tail" phenomenon. The distribution of costs has many cases close to zero, but also a long tail of extreme, and extremely important, outliers. To illustrate this, I fit the data from the Civil Rules Survey to the log-normal distribution, which is commonly used by economists and fits the Civil Rules Survey data remarkably well. See Figure 2. The median (\$20,000) is marked with a vertical line. Perhaps contrary to our intuition, the median case is hardly representative. There is a bulk of cases with low costs, and then there is a "long tail" of extremely costly cases far above the median. These latter cases are exactly the kinds of cases that the Litigation Cost Survey addresses and investigates in greater detail.⁸

⁸ For further discussion of discovery costs, see Appendix A, pp. 5-7.

How important is this “long tail”? Consider the following: in the distribution illustrated above, *the top 5 percent of cases accounts for 60 percent of all litigation costs.*

IV. COMMENTS ON THE COSTS OF PRESERVATION

Although preservation is a stage of discovery, it is important to recognize that the studies cited above do *not* include the costs of preservation in their estimates. Studies such as the Civil Rules Survey obtain their cost data from outside counsel in litigated cases, and consequently cannot measure costs, such as preservation and collection, that are borne by the parties themselves. Hence, the estimates of discovery costs in the Civil Rules Survey are really estimates of processing, review, and production costs. The Litigation Cost Survey, on the other hand, did collect data from in-house counsel, and thus was able measure the cost of collection in addition to the costs of processing, review, and production. Even the Litigation Cost Survey, however, did not capture preservation costs. This is for two reasons: First, many of the costs of preservation, such as the costs of committing IT resources and infrastructure and the lost time of employees who must comply with “litigation holds,” are not observable by in-house counsel; they don’t appear on any legal department’s budget. Second, many of the disputes that currently impose preservation costs on companies are not lawsuits—they are *potential* lawsuits, and may never develop into litigation. The costs of preservation in these disputes are totally invisible to the court system, but are very real to the parties that must bear them.

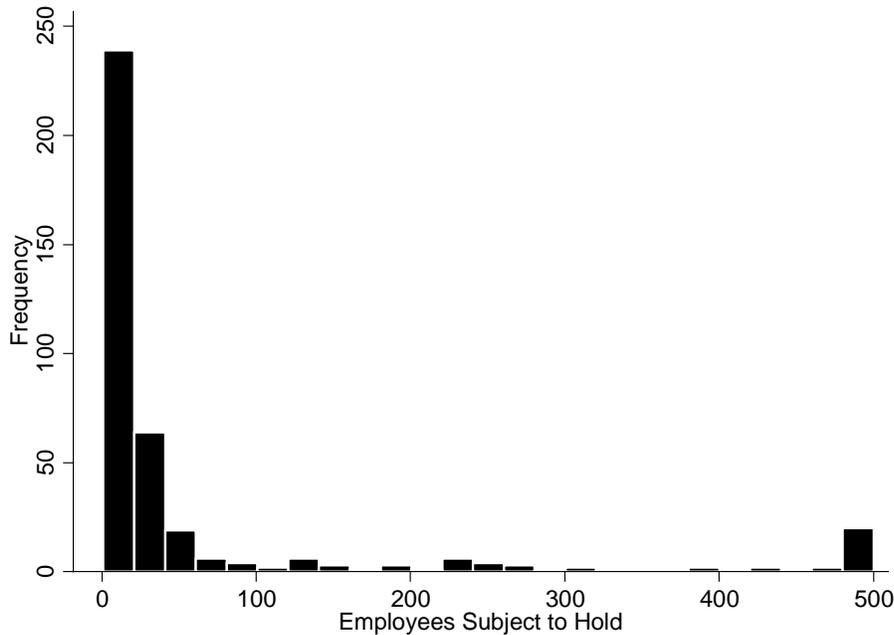
In this Part, I will show that the distribution of preservation costs has a “long tail,” similar to other discovery costs. I will then briefly discuss some of the “fixed costs of preservation” that can’t be detected when looking at individual cases. I will then examine the volume of data subject to preservation obligations relative to the amount of data involved in other stages of litigation.

A. The Long Tail in Preservation

Preliminary results from the Preservation Costs Survey suggest that there is a “long tail” of preservation costs. The data displayed in Figure 3 are for a sample of 390 distinct matters involving 43,011 litigation holds issued at a large company over a five year period.⁹

⁹ In Figure 3, note that for graphical clarity, matters with more than 500 employees subject to hold have been included in the category for 500 employees subject to holds.

FIGURE 3: DISTRIBUTION OF NUMBER OF EMPLOYEES ON LITIGATION HOLD PER MATTER



This preliminary data suggests that preservation costs, like litigation costs, are highly skewed, with a long tail in which a small number of highly complex and burdensome cases accounts for a large share of the total costs borne by individuals subject to holds.¹⁰

B. The “Fixed Costs” of Preservation

While many costs of preservation, such as the cost of responding to litigation holds, accrue on a per-case basis, other preservation costs are not tied to a particular matter. They instead reflect the costs for a company to create internal systems to handle preservation across all cases. These “fixed costs” include expensive investments in technology that companies make in order to control what would otherwise be even higher per-case preservation costs.

Preliminary results from the Preservation Costs Survey provide examples of these fixed costs. One fixed cost is the cost of systems to handle litigation hold notices. Two companies report that implementing such systems cost approximately \$800,000 to \$900,000, with upkeep and maintenance costs of \$150,000 per year. Other examples include a tool for collecting data to be preserved separately that cost \$4,800,000 to implement. One company’s data vault system cost \$12,000,000 to implement and maintain in 2010.

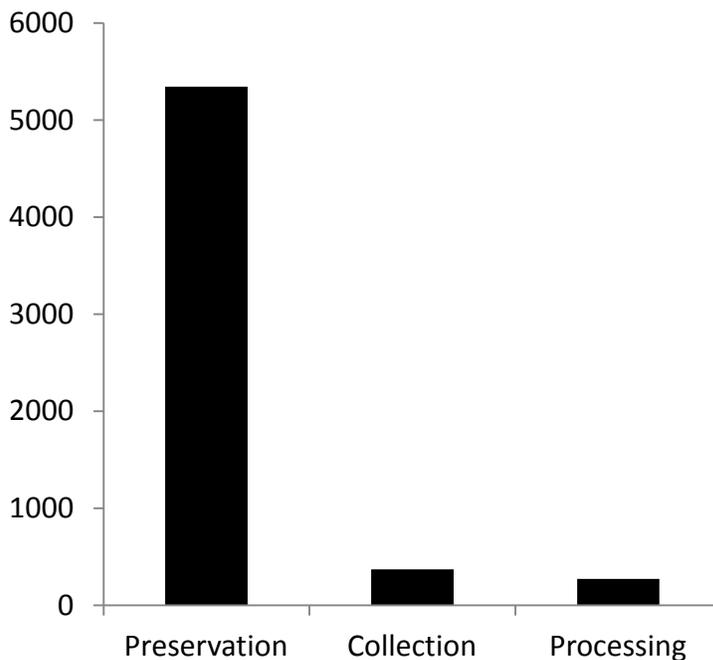
¹⁰ For further evidence and discussion, see Appendix A, pp. 8-10.

Further research is necessary to provide estimates of the costs of preservation that better capture the full range of costs. These initial examples, however, indicate that compliance with broad preservation obligations can be a very expensive undertaking for many companies.¹¹ As I estimate in Part V, the total cost of current Rules is in the billions of dollars. Reducing these costs will potentially lower product prices and create jobs.

C. The Discovery Sombrero

How does the amount of data subject to preservation compare to the amount of data affected by other stages of discovery? Figure 4 presents data from a large company on the number of custodians involved in three stages of discovery: preservation, collection, and processing. Out of over 5000 custodians placed on litigation hold, and thus subject to preservation obligations, fewer than 10 percent ultimately see their data collected, let alone processed.

FIGURE 4: NUMBER OF CUSTODIANS SUBJECT TO PRESERVATION, COLLECTION, AND PROCESSING, FORTUNE 100 COMPANY

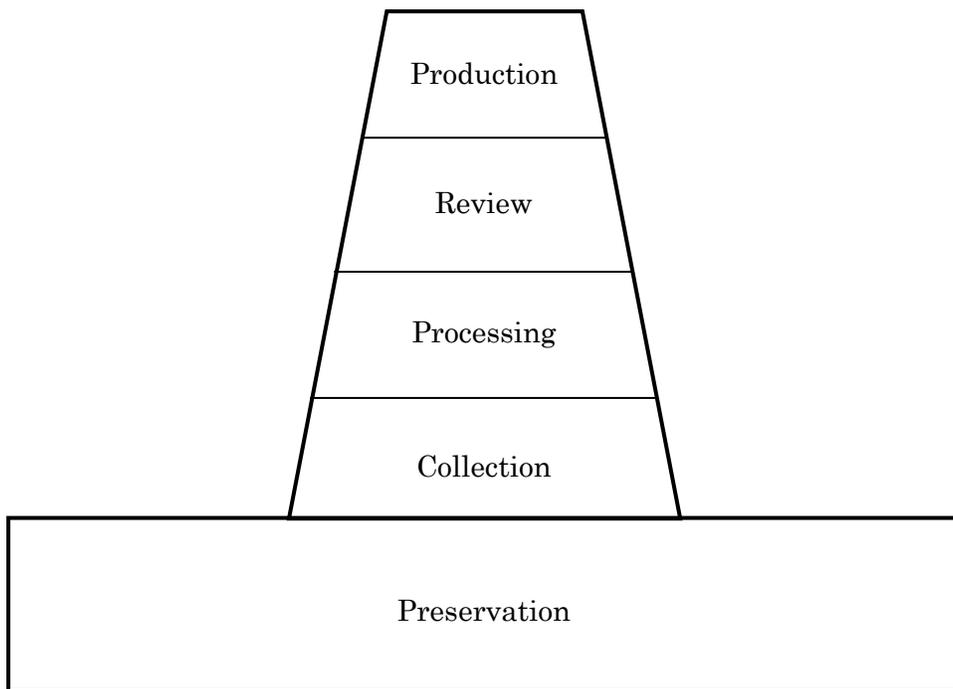


Source: Hubbard, *Preliminary Report*.

¹¹ For further discussion, see Appendix A, pp. 10-11.

Let's return to the discovery pyramid in Part II. In light of this data, a more accurate representation would be a sombrero: a wide "brim" of preservation, and a much narrower, tapering set of documents subject to collection, processing, and so on. See Figure 5.¹²

FIGURE 5: THE DISCOVERY SOMBRERO, WITH STAGES OF DISCOVERY



When we see the disproportionate bulk associated with preservation, we begin to understand the urgency for new Rules governing preservation. It is also important to understand *why* we see this sombrero shape. Preservation is unique in two important respects, both of which bear directly on the proper direction of Rules reform.

First, preservation is governed by Rules that are broad, unclear, and lacking uniformity. This leads preserving parties to "over-preserve" because the boundaries of their obligations, and therefore their risk of sanctions, are uncertain.

Second, under current law, preservation is an obligation imposed not only on parties *in* federal court, but also on parties *outside* of federal court, i.e., parties who are not involved in litigation at all. The other stages of discovery all occur *in* litigation, and the Federal Rules

¹² The illustration, of course, is not quite to scale. If this were drawn to scale, the brim would be even wider, and the top of the sombrero would be extremely narrow!

governing discovery apply only to cases in *federal* court. But the current federal law on preservation places affirmative obligations on parties to undertake costly preservation activity before litigation commences. As a consequence, companies have to incur the costs of preservation in matters that never become lawsuits (and also matters that end up in state, not federal, court).

Of course, the conduct of companies and individuals outside of federal court is regulated by federal law all the time. But it is usually substantive federal law, enacted by Congress. The federal courts should be cautious before using the Federal Rules or their inherent power to create affirmative obligations on individuals and businesses in matters that never cross the threshold of a federal courthouse.¹³

With this in mind, I now turn to question of whether new Rules could improve the law governing preservation and generate significant cost savings for the economy.

V. POLICY ANALYSIS

A. Trigger, Scope, and Sanctions

There are three major subjects that preservation Rules must address. I provide a detailed analysis of these three subjects in my paper, “Preservation under the Rules,” which is attached as Appendix C. I summarize my arguments here.

The first question that Rules governing preservation must address is when an obligation to preserve is triggered. Currently, the trigger is the onset of litigation or the “reasonable expectation” of litigation. As shown above, such a rule has the unintended consequence of federal courts regulating the activity of parties in disputes that never end up in federal court. Hence, **the Rules should limit the triggering event for the affirmative duty of preservation under the Rules and under the federal court’s inherent power to the initiation of proceedings in federal court.**

Such a Rule would be good policy; in the next section I estimate the cost savings from such a Rule. It also respects the myriad state and federal laws that govern data preservation outside the context of federal court, and it avoids the concern that the federal courts might be imposing too much (in the way of affirmative duties) on parties not in federal court.

Importantly, too, triggering the duty to preserve at the onset of federal litigation would *not* limit the ability of federal courts to police attempts to destroy incriminating information. The *duty to preserve* should not place affirmative obligations on parties who are not in litigation to change their normal business activities and undertake

¹³ For further discussion, see Appendix C, pp. 11-19.

costly actions to set aside documents and data. But parties out of court have never been allowed to—and must never be allowed to—deliberately destroy evidence with the purpose of preventing its use in future litigation. This latter rule, which I call the *duty not to spoliare*, would not be affected by a Rule governing trigger for the duty to preserve. Unlike the duty to preserve, the duty not to spoliare does not impose costly obligations on parties: it requires that businesses and individuals *not* interrupt their usual activities; i.e., *not* override their usual records management activities in order to dispose of incriminating materials. The Rules will not affect any prohibition against changing one’s usual activities in order to destroy incriminating data.

The next major issue is the scope of the preservation obligation, once triggered. Here, **the Rules should set presumptive limits on the scope of preservation**, for example by setting a limit of 15 custodians subject to a litigation hold in a case. Presumptive limits not only reduce the costs of preservation; they give parties the incentive to meaningfully negotiate over the scope of preservation. Currently, with no presumptive limits on preservation, plaintiffs’ lawyers have the incentive to ask that “everything” be preserved, and defense lawyers have no incentive to involve the other side in the preservation process.

Third, the Rules should address the standard for imposing sanctions. Here, I will simply note that the reality of modern discovery is that it often takes the form of a search for the needle in the haystack. As such, when data is lost, the overwhelming likelihood (absent evidence of bad faith) is that the lost data was neither relevant nor prejudicial, if only because the vast majority of data is never relevant or material. **Rules governing sanctions should take care to protect parties from sanctions that rest on presumptions that any lost data is relevant and material.**

B. Estimating Cost Savings from New Rules

Well-designed Rules governing preservation should generate substantial savings in the costs of preservation and litigation. Many of the costs savings are hard to project, given the limitations of current data and the fact that some benefits of improved Rules will be indirect, such as more meaningful and productive negotiations between parties. To estimate one set of cost savings from improved Rules, I will focus on two specific suggestions for Rules changes, one regarding trigger and one regarding scope. (I do not separately quantify the effect of a new Rule addressing sanctions, but my estimates for the effects of Rules addressing trigger and scope assume that a Rule appropriately clarifying sanctions is put in place.) These are rough, “back of the envelope” calculations.

To estimate some of the cost savings from a Rule regarding trigger, I will consider a Rule that places the trigger for the duty to preserve at the filing of a lawsuit in federal court, which I have argued is the appropriate trigger point. Such a Rule would have two direct effects on costs: First, it would reduce the number of matters subject to litigation holds, because many or even most litigation holds at large companies are for matters for which there is no filed lawsuit. Second, it would reduce the average scope of any given litigation hold. Why? Because litigation holds would be implemented at, rather than (in some cases) long before the onset of litigation, meaning that preserving parties will have the benefit of a complaint (or subpoena) that clarifies the proper scope of preservation. This allows parties to design litigation holds to more efficiently capture the data most likely to be useful to the parties.

To estimate some of the cost savings from a Rule regarding scope, I will focus solely on a potential provision that would set a presumptive, numerical limit for how many custodians would be subject to the duty to preserve. In my *Letter to Judge Campbell*, I use the example of a presumptive limit of 15 custodians.

Taken together, and using conservative estimates of the impact of the new Rules, the total effect is a reduction in litigation hold costs of 63 percent. Focusing only on the lost employee productivity caused by litigation holds, **I would estimate the dollar value of these savings for a single, large company to be about \$2 million. This is intended as a lower bound estimate**, because it includes only the cost savings in employee time, and no other cost savings. Given the thousands of large companies that face significant preservation costs, one can extrapolate from this number to estimate that **the savings for all companies would be in the billions of dollars.**¹⁴

C. Why Rulemaking Is Needed

Finally, I note that the dialogue on rulemaking has included two suggested alternatives to new Rules. First, there is the possibility that improved technology will render moot concerns about preservation costs (and maybe even discovery costs generally). Second, there is the possibility that continued development of legal rules on a case-by-case basis in the federal courts will eventually lead to the clarity that is needed. I do not believe that either of these alternatives is viable.

Technology. There is no doubt that in some ways, technology has reduced the costs of preservation. The per-unit-of-data cost of storage has fallen exponentially for decades, and continues to fall. If the cost

¹⁴ For further discussion, and the details of the calculations underlying these figures, see Appendix B at pp. 1-4.

of preservation were simply the cost of storing a fixed amount of data, the cost of preservation would have ceased to be a live issue long ago.

But as discussed above, the main costs of preservation are not storage costs. There is the human cost in terms of workers diverted away from productive, business activity to preservation obligations. There is also the cost of adapting new systems to preservation obligations and managing the complexities of legacy data and outdated storage formats. The rapid advance of technology actually exacerbates these costs. More importantly, technology's full potential to offer low-cost solutions to preservation cannot be realized so long as legal obligations are amorphous and unclear. Computers and technology can help effectuate clear legal rules, but they can't make confusing rules clear.¹⁵

Case-by-case law making. It has been suggested that the continued evolution of the case law on preservation will lead to a gradual convergence of rules, reducing the current uncertainty and conflicting obligations imposed by the case law. I am skeptical that this process offers significant hope for national uniformity. The law on discovery, and preservation especially, is almost exclusively created at the district court level. (Virtually all discovery rulings are "non-final judgments" and thus not appealable.) District court opinions, of course, are not even binding precedent in their own district, and in the eight years since the landmark *Zubulake* case,¹⁶ there is not even uniformity *within* circuits, let alone across circuits.¹⁷

CONCLUSION

The Advisory Committee on the Civil Rules has undertaken a reexamination of the Federal Rules to consider amendments that would improve the efficiency and reduce the burdens of litigation. It is focusing first on the Rules governing discovery and preservation rules, and has requested empirical data on the costs of discovery and preservation. Although the empirical studies to date are preliminary, it appears that companies could save billions of dollars with new Rules clarifying the events triggering the duty to preserve, the scope of preservation, and the standards for sanctions.

¹⁵ For further discussion, see Appendix A at pp. 16-17.

¹⁶ *Zubulake v. UBS Warburg*, 220 F.R.D. 212 (S.D.N.Y. 2003).

¹⁷ For further discussion, see Appendix C at pp. 27-29.