

Court-System Transparency

Lynn M. LoPucki*

ABSTRACT: Over the past decade, the federal courts became the world's most transparent court system by switching from paper to electronic filing, resolving daunting privacy problems, and posting their case files on the Internet. Now they are embarking on a second, equally important transformation—the use of relational forms from which court data can be extracted automatically. This Article describes the technology and seeks to project and evaluate the effects of that second transformation.

If it occurs, the second transformation would create millions of windows into the courts at virtually no cost to the government. Policymakers, litigants, and the public would be able to see and understand the patterns of judicial decisionmaking—who wins what and how often. That would provide policy makers the feedback needed to fine tune the system, lawyers the ability to predict the outcomes of their cases, and the public the ability to see what courts actually do. All could also see whether the precautions they take for supposed legal reasons are the right ones.

Opponents argue that court-record transparency (1) would expose parties and witnesses to the risk of identity theft and other harms, (2) would invade privacy by making previously-difficult-to-obtain public-record information about individuals readily available, and (3) would pressure judges in ways that deprive them of judicial independence. This Article argues that none of those objections is well-founded.

* Lynn M. LoPucki is the Security Pacific Bank Professor of Law at the UCLA School of Law, the annual Bruce W. Nichols Visiting Professor of Law at the Harvard Law School, and the former Senior Advisor to the National Bankruptcy Review Commission with respect to the Commission's Data Study Project. He can be contacted at lopucki@law.ucla.edu. Thanks are due to Drew LoPucki for his assistance in developing the ideas presented in this paper; to Shahab Asghar for assistance with research; and to Frances Foster, Elizabeth Gibson, Melissa Jacoby, Peter Joy, Bob Lawless, Tony LoPucki, Aaron Moshiahwili, Katie Porter, Hillary Sale, Walter Weyrauch, and workshop participants at The University of Iowa College of Law, the University of North Carolina School of Law, the University of Illinois College of Law, the Washington University School of Law, and the Rand Corporation for comments on earlier drafts.

I. INTRODUCTION.....	483
II. THE TECHNOLOGICAL CHALLENGE.....	485
A. <i>THE MECHANISMS OF TRANSPARENCY</i>	485
B. <i>REMOVAL OF THE BARRIERS TO TRANSPARENCY</i>	488
C. <i>THE EFFECTS OF REMOVAL</i>	491
III. TRANSPARENCY'S BENEFITS.....	494
A. <i>EXPOSURE AND REDUCTION OF CORRUPTION</i>	494
B. <i>ENHANCEMENT OF LEGISLATIVE CONTROL OVER THE COURTS</i>	495
C. <i>POPULARIZATION OF THE LAW</i>	497
D. <i>PREDICTION OF LITIGATION OUTCOMES</i>	498
1. Can Lawyers Predict the Outcomes of Cases?.....	499
2. Can Statistical Analysis Predict the Outcomes of Cases?.....	501
3. The Inadequacy of Court-Opinion Data.....	504
4. Prediction's Systemic Effects.....	506
E. <i>ENHANCEMENT OF THE INFORMATION INFRASTRUCTURE</i>	510
F. <i>REDUCTION OF LAWYER AND LITIGANT ERROR</i>	510
G. <i>AUTOMATION OF DOCUMENT SERVICE AND FILE MAINTENANCE</i>	513
IV. OBJECTIONS TO TRANSPARENCY.....	513
A. <i>COST</i>	514
B. <i>PERSONAL IDENTIFICATION AND PRIVACY</i>	516
1. The Need for Identification in Court Records.....	518
2. The Difficulty of De-Identifying Court Records.....	521
3. The Privacy Objections.....	522
a. <i>Identity Theft and Other Specific Harms</i>	522
b. <i>Practical Obscurity</i>	523
c. <i>Personal-Data Aggregation</i>	526
d. <i>Contract Restrictions to Prevent Commercial Use</i>	529
C. <i>JUDICIAL INDEPENDENCE</i>	531
D. <i>COPYING LAWYERS' WORK PRODUCT</i>	533
E. <i>FLIGHT TO PRIVATE ADJUDICATION</i>	535
V. CONCLUSIONS.....	537

I. INTRODUCTION

Computer algorithms capable of predicting the outcomes of legal disputes may soon be a reality. These algorithms would base their predictions not on legal analysis of statutes, regulations, and court opinions,¹ but on statistical analysis of case outcomes. Because the algorithms do not yet exist, the case characteristics from which they would make their predictions remain largely a mystery.

The characteristics would certainly include the identities of the judge² and the lawyers³ in a particular case. Those two pieces of information alone would make the algorithms highly provocative, because neither is supposed to play any role at all in case outcomes. Through the lens of such algorithms, observers would see the courts as they actually operate, instead of as they are supposed to operate.

Court transparency requires such algorithms. A court system is “transparent” for the purposes of this Article when all relevant aspects of its operation are revealed to policymakers, litigants, and the public in forms that they can readily comprehend. For reasons that this Article explains,

1. The sad history of efforts to devise computer algorithms capable of predicting case outcomes based on analysis of issues and factors in legal opinions is summarized in Kevin D. Ashley & Stefanie Bruninghaus, *Computer Models for Legal Prediction*, 46 JURIMETRICS J. 309, 311–26 (2006).

2. See, e.g., Robert G. Dixon, Jr., *The Welfare State and Mass Justice: A Warning from the Social Security Disability Program*, 1972 DUKE L.J. 681, 717 (finding a “striking and disturbing” disparity in the rates at which social-security-disability hearing examiners reversed awards); Lynn M. LoPucki & Joseph W. Doherty, *Why Are Delaware and New York Bankruptcy Reorganizations Failing?*, 55 VAND. L. REV. 1933, 1939–45 (2002) (finding that large, public companies that were reorganized in the Delaware and New York bankruptcy courts from 1991 to 1996 failed at several times the rate of companies reorganized in all other courts, despite the lack of any apparent differences in the companies choosing those courts); Jaya Ramji-Nogales, Andrew I. Schoenholtz & Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295, 332 (2007) (finding that the levels at which particular asylum officers grant asylum frequently differ from the office averages by more than fifty percent); Paul Elias & Rinat Fried, *A Failure to Execute*, RECORDER (Cal.), Dec. 15, 1999, at 1 (finding huge variations from judge to judge in the lengths and dispositions of death-penalty cases); Tom Chang and Antoinette Schoar, *Judge Specific Differences in Chapter 11 and Firm Outcomes* (2008) (unpublished manuscript), available at <http://econ-www.mit.edu/files/3590> (5000-case empirical study documenting the existence of persistently pro-debtor and pro-creditor judges in business bankruptcies).

3. See, e.g., David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145, 1145 (2007) (“[A] veteran public defender with ten years of experience reduces the average length of incarceration by 17 percent relative to a public defender in her first year.”); Catherine T. Harris et al., *Who Are Those Guys? An Empirical Examination of Medical Malpractice Plaintiffs’ Attorneys*, 58 SMU L. REV. 225 *passim* (2005) (finding large differences in medical-malpractice-litigation outcomes based on attorney experience); Richard J. Lazarus, *Advocacy Matters Before and Within the Supreme Court: Transforming the Court by Transforming the Bar*, 96 GEO. L.J. 1487, 1487 (2008) (finding that the U.S. Supreme Court “grants the petitions filed by the expert members of the Bar at a significantly higher rate, and they also prevail on the merits more frequently”).

court systems can become transparent only when court files are maintained in relational electronic formats and the public has free, technologically unfettered access to their contents. Relational formats are the familiar formats used in data-management and spreadsheet programs. Each piece of data is tagged as the value of a characteristic of an object. Empiricists refer to relational data as “coded.” Statistics programs can process such data into statistics, tables, and graphs.

The federal courts—which already maintain court files in electronic format and make them available over the Internet—are on the brink of introducing relationally formatted forms. Users of these forms “code” the data as they create it, by entering it into fields (boxes) in specified formats—essentially the way customers fill out order forms on the Internet. The Judicial Conference of the United States (the “Judicial Conference”), the regulator of the federal courts, has approved a data-enabled PDF format for use, and the United States Trustee, a division of the Justice Department, has recommended to software vendors that they begin using it for nineteen documents commonly filed in bankruptcy cases.⁴

The federal courts are not introducing relationally formatted forms merely to achieve transparency. Indeed, transparency may not even be among the courts’ objectives. Data-enabled forms would improve court administration by facilitating its automation. Once the technology is in place, however, the courts will be transparent unless regulators erect or maintain barriers to prevent public use. So far, the Judicial Conference has permitted a steady but cautious advance toward transparency.

This Article takes as its starting point the current state of the world’s most transparent court system—the United States Courts as accessible through Public Access to Court Electronic Records (“PACER”).⁵ Part II describes the physical changes necessary for that system to achieve transparency.

The remaining Parts then use systems–strategic analysis⁶ to project the consequences. Part III explores the benefits of transparency. Transparency would expose and reduce corruption. It would expand the power of citizens

4. U.S. Tr. Program, U.S. Dep’t of Justice, Data Enabled Form Standard, <http://www.usdoj.gov/ust/eo/bapcpa/defs/index.htm> (last visited Sept. 23, 2008) (showing nineteen data-enabled bankruptcy forms qualified for use under the standard).

5. See, e.g., Peter W. Martin, *Online Access to Court Records—From Documents to Data, Particulars to Patterns* 15 (Cornell Law Sch., Cornell Legal Studies Research Paper No. 08-003, 2008), available at <http://ssrn.com/abstract=1107412> (“Due to a cluster of mutually reinforcing factors, state court systems have been far slower and less coordinated in making [the transition from paper to electronic media].”).

6. For an explanation of the systems/strategic analysis of law-related systems, see Lynn M. LoPucki, *The Systems Approach to Law*, 82 CORNELL L. REV. 479 *passim* (1997) (explaining the systems approach); Lynn M. LoPucki, *Legal Culture, Legal Strategy, and the Law in Lawyers’ Heads*, 90 NW. U. L. REV. 1498, 1545–49 (1996) [hereinafter LoPucki, *Legal Culture*] (explaining strategic analysis).

and legislators over the courts and make the actual rules that govern society visible to the public. Transparency would promote the settlement of the most predictable cases and improve overall efficiency by focusing the system's resources on the least predictable cases. Transparency would reduce or prevent many kinds of lawyer malpractice and litigant error.

Part IV examines and refutes the arguments generally raised in opposition to court-system transparency. The costs of transparency to the public would be minimal, because academic institutions, nongovernmental organizations, individuals, and the private sector would fund and do the cheap, relatively simple data processing required. On careful examination, the variety of privacy-based arguments generally raised in opposition to transparency all fail—principally because the records to be made transparent are already public and widely available. Transparency would enhance rather than impair judicial independence because it would provide a sounder basis for the public's evaluation of judicial performance. Transparency would enable widespread copying of legal documents. But that too turns out to be a benefit rather than a detriment when analyzed in terms of system function. This Article concludes in Part V that the effects of court-system transparency would be overwhelmingly positive.

The analysis proceeds on two assumptions. First, the government chooses to release all recorded court-system data except those sealed by court order or redacted pursuant to recently adopted federal privacy standards. Second, the government provides only the present level of funding for data analysis. In accord with the systems-strategic method, the analysis then projects the consequences by taking the perspectives of various court-system participants, imagining the strategies that each would pursue in response to the data release, and speculating on the interaction among those strategies.

II. THE TECHNOLOGICAL CHALLENGE

This Part briefly describes the current system for access to federal case files, the preconditions to transparency, and the mechanisms by which courts would achieve transparency. It describes some issues that would arise in removing the limitations that now prevent transparency and the effects of removing those limitations.

A. THE MECHANISMS OF TRANSPARENCY

PACER is a system operated by the federal courts. Since about 1997, PACER has made federal-court case files, including the dockets, publicly available over the Internet.⁷ A case file generally includes all documents

7. Pub. Access to Court Elec. Records, Overview, <http://pacer.psc.uscourts.gov/pacerdesc.html> (last visited Sept. 23, 2008) (providing an overview of and links into the PACER system).

filed with the clerk of the court by the parties or the judge in connection with a case.⁸ The “docket” for a case is a list of all documents contained in the case file. Prior to PACER, the public could access case files only by traveling to the courthouse, ordering records by mail, or buying copies from service-company intermediaries.⁹ With only minor exceptions, PACER now makes district-court and bankruptcy-court case files—including hearing transcripts—publicly available over the Internet. A pilot project is underway to include digital audio recordings of hearings.¹⁰ The dockets are in HTML format; the documents are in PDF format.

PACER users must register online and pay eight cents per page to view or download each document, up to a maximum single-document charge of \$2.40.¹¹ In addition, PACER’s report function provides a limited ability to search for cases nationwide (by party or, in the district court, by nature of case) or in a specific court (by several criteria).¹²

For a court system to be transparent, up-to-date research must be available to answer millions of questions regarding the patterns of cases and case outcomes. Already hundreds, if not thousands, of researchers use data extracted from PACER. The work is labor intensive. A researcher can sometimes use PACER’s report function to identify a group of cases for study, but more often researchers must identify the cases from outside sources. The researcher must select and download the documents (and incur costs) before the researcher can word-search them. Many PACER documents cannot be word-searched at all, because the documents are scanned rather than text-based PDFs, and the scans are of insufficient

8. “Case file” is used as a term of art:

The term “case file” (whether electronic or paper) means the collection of documents officially filed by the litigants or the court in the context of litigation, the docket entries that catalog such filings, and transcripts of judicial proceedings. The case file generally does not include several other types of information, including non-filed discovery material, trial exhibits that have not been admitted into evidence, drafts or notes by judges or court staff, and various documents that are sometimes known as “left-side” file material. Sealed material, although part of the case file, is accessible only by court order.

JUDICIAL CONFERENCE COMM. ON COURT ADMIN. & CASE MGMT., REPORT ON PRIVACY AND PUBLIC ACCESS TO ELECTRONIC CASE FILES A-4 (2001) [hereinafter JUDICIAL CONFERENCE PRIVACY REPORT], available at http://www.uscourts.gov/Press_Releases/att81501.pdf.

9. Lynn M. LoPucki, *The Politics of Research Access to Federal Court Data*, 80 TEX. L. REV. 2161, 2166–67 (2002) (describing how scholars accessed court files before PACER).

10. *Pilot Project Will Post Digital Audio Recordings Online*, THE THIRD BRANCH (Admin. Office of the U.S. Courts, Wash., D.C.), June 2007, available at <http://www.uscourts.gov/ttb/2007-06/pilot/index.html>.

11. Admin. Office of the U.S. Courts, CM/ECF Frequently Asked Questions, <http://pacer.psc.uscourts.gov/cmecf/ecffaq.html#GE5> (last visited Sept. 23, 2008).

12. Admin. Office of the U.S. Courts, CM/ECF Frequently Asked Questions, <http://pacer.psc.uscourts.gov/cmecf/ecffaq.html#CR116> (last visited Sept. 23, 2008) (describing the use of the report function).

quality to rescan them to text. Researchers typically print hard copies, mark the data to be extracted, and convert the data to relational form by coding and entering them into spreadsheets or databases by hand. The researchers then analyze the data statistically and present them in tables, graphs, and journal articles. Probably no more than a few hundred studies to date used PACER, and only a handful of those are kept up-to-date.

Researchers would be able to generate and maintain millions of studies only if they could entirely automate the processes of data extraction and analysis. The required automation includes four processes.

First, a researcher must be able to conduct a word search of the entire court system to identify cases for study. What prevents such a search today is that (1) many of the PDF forms are scanned rather than text-based and thus not readily word-searchable, and (2) the researcher must pay for each document before searching it. The solution to the first problem is requiring future filings to be in machine-readable text formats. The solution to the second is ending user financing of PACER. If PACER documents were available for download without charge, users could download and search the entire court system or just the part that was of interest. PACER could also support online word searches of the entire system.

Second, the researcher must be able to download the relevant documents automatically. Automated Access to Court Electronic Records ("AACER") and Bloomberg are both private firms that identify relevant PACER documents by case and docket searches, and download millions of them for paying clients.¹³ Those firms' existence demonstrates that automatic downloads are feasible even without improvements in PACER. Identification of the documents to be downloaded is the problem, and the system-wide word search is the likely solution.

Third, the researcher must be able to move data from downloaded documents into spreadsheets, data-management programs, and statistics programs automatically. Researchers cannot do that today because court documents do not provide the information in sufficiently regular forms and formats. The solution is to require those who create the information (principally lawyers and judges) to create it in sufficiently regular forms and formats. Fortunately, the demands of automated case management are already driving the courts in this direction.¹⁴ The Judicial Conference has

13. AACER: Automated Access to Court Electronic Records, <http://www.aacer.com> (last visited Sept. 23, 2008) (describing AACER); Bloomberg Home Page, <http://www.bloomberg.com> (last visited Sept. 23, 2008) (follow "Log In/Register" hyperlink for subscription information).

14. Gregory M. Silverman, *Rise of the Machines: Justice Information Systems and the Question of Public Access to Court Records over the Internet*, 79 WASH. L. REV. 175, 198 (2004) (noting that courts are already integrating their case-management information systems with electronic-docket and document-management systems, which gives them the ability to grant public access without additional development costs).

authorized and may soon require the use of data-enabled PDF forms for nineteen of the most commonly filed documents in bankruptcy cases.¹⁵ The lawyer or judge who completes a data-enabled PDF form answers specific questions by checking boxes or filling in blanks. For example, a judge entering a judgment might check a box indicating it to be a money judgment, enter the dollar amount in response to one prompt, fill in the date from which interest would accrue in response to a second, and state the rate at which interest would accrue in response to a third. If the judgment is for an injunction, the judge might select, from a menu of choices, standard language identifying the acts to be enjoined, or might enter the judge's own customized language. The data would be stored in fields. The Adobe Acrobat program is already capable of automatically transferring the data from data-enabled forms into Excel spreadsheets.¹⁶

Fourth, the researcher must be able automatically to compile, display, and statistically analyze the data. Software already exists for performing these tasks.

After full implementation of these changes, researchers would still have to design studies and program their implementation. But once a particular study was up and running, the process of updating that study would be trivial. For most studies, the researchers would not need outside funding and so would not need to make grant applications. Each researcher would be able to do many times the number of studies that a researcher can do today. Studies that produced useful views of the system would be automatically updated and thus always current. Some researchers might hide their work as "proprietary" or make it available only for a fee. The current pattern, however, is for researchers to post their data and findings to the Internet and make them available without charge. Users would find both the fee and free sites through search engines such as Google.

B. REMOVAL OF THE BARRIERS TO TRANSPARENCY

The federal courts would need to make three changes to achieve transparency: (1) eliminate user fees, (2) specify and require relational forms for common documents, and (3) disclaim present and future restrictions on data retransfer.

Elimination of PACER fees would cost the federal government its current PACER revenue stream of about \$60 million a year.¹⁷ Requiring the use of relational forms would require the government to design those forms. The form-design project could advance form by form and might take many

15. See U.S. Tr. Program, *supra* note 4 (explaining data-enabled forms and providing links to the nineteen forms).

16. Acrobat Professional 7.0 and 8.0 support data-enabled forms. In 7.0, users select "Forms" and then "Create spreadsheet from data files."

17. ADMIN. OFFICE OF THE U.S. COURTS, ANNUAL REPORT OF THE DIRECTOR 2006, at 25 (2006) (stating that PACER revenues were \$58 million in 2006).

years to complete. If so, the court system would only gradually become transparent.

The government currently imposes only one restriction on data transfer—researchers working under PACER fee exemptions cannot retransfer data. If the government eliminated user fees, researchers would no longer obtain exemptions and the restriction would be irrelevant. Researchers contemplating substantial investments in programming projects might, however, still be concerned that the government would later impose restrictions. Those researchers might need assurances that the government would not.¹⁸

A decision to remove the limits on transparency, without more, would leave three political issues unresolved. The first would be the extent to which privacy and commercial secrecy claims should limit court transparency.¹⁹

The second remaining issue would be the extent to which the system should require litigants to furnish, for transparency purposes, information beyond what courts need to process cases or administer the court system. Federal courts already require litigants to furnish some data “for statistical purposes.” Despite serious problems with the accuracy of those data,²⁰ some—particularly the case classifications and outcomes—have proven to be valuable in research. Requiring litigants to furnish demographic data such as gender, age, marital status, or educational level would add to the power of the data.

As a practical matter, the courts are unlikely to require any additional information at all. Doing so would burden the right to litigate by adding expense and forcing litigants to disclose otherwise private information. Nor is additional information needed. Provided that the courts do not also de-identify the case files, researchers would be able to link the court files to outside sources of such information and thus supplement the case file data.²¹

The third issue—actually thousands of issues—would be the design of the forms that would elicit relational data from system participants. The precise questions and prompts included on the forms would determine the form of the resulting data and ultimately what studies will be possible. Researchers may be able to overcome some kinds of shortcomings in the data by reformatting them. Reformatting, however, would always be difficult and expensive, and sometimes it would be impossible. For all practical purposes, the formats in which parties, judges, and others would submit relational data, or the formats to which those data could automatically be converted, would be the formats in which they would be used.

18. See *infra* Part IV.B.3.d (discussing proposals for restrictions on data use).

19. Part IV.B addresses this issue.

20. See *infra* note 130 and accompanying text (describing some of those problems).

21. See *infra* Part IV.B (discussing de-identification and its effects).

Some questions and prompts would clearly be superior to others. Prompts requiring more specificity in the data—without requiring so much that litigants cannot reliably provide it—would maximize transparency. To illustrate, prompting a litigant to indicate the litigant’s “street address” would provide more transparency than prompting the litigant to indicate its “address.” Responses to the latter prompt would include post-office boxes, leaving some debtors’ geographical locations unspecified. Similarly, asking for the dollar amount in controversy would provide more transparency than asking within which of several ranges the amount in controversy falls.²² Researchers can easily categorize the data themselves if they want ranges. Requiring the “market value of assets” or the “book value of assets,” rather than the “value of assets,” would also increase the utility of the information furnished.

Specificity comes at a cost. The forms that litigants must file become longer and more difficult to complete. At some point, the more specific prompts confuse litigants and elicit incorrect answers. Some experimentation may be necessary to establish ideal levels.

By their choices, the form drafters will determine the questions that researchers ultimately can and cannot answer. The long-running controversy over the number of business bankruptcies illustrates the problem.²³ One might consider a bankruptcy to be a “business” bankruptcy (1) only if the debtor is currently engaged in business; (2) if the debtor is no longer in business, but the business caused the bankruptcy; (3) if the debtor is no longer in business, but the business contributed to the bankruptcy; or (4) if the debtor was never in business, but the bankruptcy resulted from the debtor’s guarantee of the debtor’s corporation’s debts. The current system leaves the choice among these definitions to each person entering data. The resulting data are useless. To generate useful data, a transparent court system would have to choose among the four definitions. To require debtors to categorize their cases by all four standards would be unreasonably burdensome.

The choice is political because it determines how many bankruptcies will be categorized as “business.” That matters because commentators associate business bankruptcies with the politically positive image of

22. The petition form currently used in bankruptcy cases requires the debtor to provide “estimated” numbers of creditors, assets, and debts by choosing one of several ranges for each. U.S. Bankr. Court, Official Form 1, Voluntary Petition, available at http://www.uscourts.gov/rules/BK_Forms_08_Official/B_001_0108f.pdf. Users of this data are stuck with the categories chosen by the government. If the petition form asked for the numbers of creditors, assets, and debts—allowing the debtor to furnish an estimated number if the debtor did not know the precise number—users could construct whatever categories they preferred.

23. E.g., Robert M. Lawless & Elizabeth Warren, *The Myth of the Disappearing Business Bankruptcy*, 93 CAL. L. REV. 743, 757–64 (2005) (describing the history of Administrative Office problems in counting business bankruptcies).

entrepreneurial risk taking, while they associate consumer bankruptcies with the negative image of profligate spending.

Officials who work in the court system would be tempted to promote analyses that put them in a good light and to discourage those that would do the opposite. They could accomplish that by manipulating court information processing at its root—the relational formats in which litigants furnish it. To prevent that from occurring, officials from outside the system, academic researchers, and the public should participate in determining the relational formats.

C. THE EFFECTS OF REMOVAL

As previously noted, removing the limits on transparency would cost the government \$60 million per year in PACER revenues, plus the cost of developing and promulgating new, relational forms. Some or all of the costs in the latter category would be incurred even absent public transparency and, to that extent, should not count as transparency costs. Standardized, data-enabled PDF forms would reduce the costs of court administration by rendering court-file data transparent to judges and court administrators. That is not part of transparency as defined for the purpose of this Article. That work would go forward even if the courts did not intend to make court files transparent to the public. In fact, courts currently may be pursuing public transparency only as a means of financing internal transparency.

The government would incur no other costs. The private vendors of litigation-support software and their customers would bear the cost of implementing the new forms. This Article assumes that the government neither conducts nor funds any of the data processing and analysis that would render the court system transparent to the public.

Once relational data were available cost-free, both the public and private sectors would devote substantial resources to process them into usable forms.²⁴ Hundreds of U.S. law-school researchers are already engaged in the collection and analysis of court-system data.²⁵ Their numbers would multiply in the new, data-rich environment.

24. NAT'L BANKR. REVIEW COMM'N, BANKRUPTCY: THE NEXT TWENTY YEARS 933–34 (1997), [hereinafter NAT'L BANKR. REVIEW COMM'N], available at <http://govinfo.library.unt.edu/nbrcreport/21bdata.pdf> (predicting that an in-bulk release of bankruptcy data would lead to extensive processing at nongovernment expense); *id.* at 926 (predicting that the effect would be to “aid in the development of bankruptcy policy, the allocation of bankruptcy resources, and formulation of bankruptcy legislation”).

25. For example, scholars submitted more than 300 papers presenting the results of empirical studies to the Second Annual Conference on Empirical Legal Studies. E-mail from Jennifer Arlen, Norma Z. Paige Professor of Law, NYU Sch. of Law, to Lynn M. LoPucki, Security Pac. Bank Professor of Law, UCLA Sch. of Law (July 12, 2007, 2:05 p.m. PST) (on file with author). Courses and seminars in which students design and execute empirical studies of the legal system are now common. A group of scholars at the University of Illinois College of Law is preparing a set of teaching materials for those courses and seminars. *See generally* Robert

The history of securities and stock-price disclosures supports the idea that bulk data release induces privately funded data processing. U.S. securities laws require extensive disclosure of company information in highly structured, but not strictly relational, formats. Standard & Poor's converts that structured text into relational data and sells access to it through the Compustat database²⁶ and a second-generation product called "Capital IQ."²⁷ The University of Chicago's Graduate School of Business has done much the same thing with stock prices.²⁸ Today, thousands of business-school faculty members devote substantial portions of their working days to analyzing data from Capital IQ, the Center for Research in Securities Prices ("CRSP"), and hundreds of related proprietary databases, and publishing their findings in scholarly journals. The operation of public companies and the pricing of their securities are not yet fully transparent, but that may be only because Capital IQ and CRSP charge for their data and prohibit re-transfer. Websites that automatically convert stock-price data into user-adjustable tables and charts are ubiquitous.²⁹

The single set of data released from a transparent court system would be applied to answer a potentially infinite number of questions. As a result, numerous processing paths would flow from that single source. Those paths would have to be easy to create and modify, and capable of linking court data to data from other sources.

This need for flexibility suggests that the subsystems would be modular. That is, subsystems would arise to process particular kinds of data through particular stages for particular purposes. For example, one subsystem might specialize in converting non-relational, text-based data from the district courts into relational data. Another might link relational court data to census data. A third might specialize in providing journalists with a window on state and federal Freedom of Information Act cases.

The Bankruptcy Research Database ("BRD") illustrates this modular approach.³⁰ Through the BRD, I specialize in converting structured text

M. Lawless, Jennifer K. Robbennolt & Thomas S. Ulen, *Empirical Methods in Law* (2008) (unpublished manuscript, on file with the Iowa Law Review).

26. Univ. of Pa., Wharton Sch., Databases in WRDS, <http://wrds.wharton.upenn.edu/demo/databaselist.shtml> (last visited Jan. 27, 2009) (describing various business databases, including Compustat).

27. Capital IQ, Implementation—Data Feeds, <https://www.capitaliq.com/main/data/feeds.asp> (last visited Sept. 23, 2008).

28. Ctr. for Research in Sec. Prices, History, <http://www.crsp.com/crsp/about/history.html> (last visited Sept. 23, 2008) (describing how the Center for Research in Securities Prices ("CRSP") database developed).

29. See, e.g., Tradingcharts.com, Stock Price Charts, <http://stocks.tradingcharts.com/stocks/charts/> (last visited Sept. 23, 2008) (offering free, customized stock-price charts).

30. One version of the database is available online. Bankr. Research Database, Web BRD: A Window on the World of Big-Case Bankruptcy, <http://lopucki.law.ucla.edu> (last visited Jan. 27, 2009).

from PACER and Securities and Exchange Commission filings into relational data for the set of all large public-company bankruptcies.³¹ With the help of AACER, I monitor the cases to their ultimate dispositions. I make data relational by collecting each field according to BRD protocols and exercising the judgment sometimes necessary to do that. The BRD contains fields for linking BRD data to other business databases and for linking BRD data to PACER.³²

I make the complete BRD available free to legal scholars in Excel spreadsheet form. The scholars combine it with data from other sources, sometimes including data that they obtain from Capital IQ or directly from PACER, and use it to answer questions of interest to them. Professor Janis Sarra of the University of British Columbia is constructing a database of Canadian public-company bankruptcies that she designed to work with the BRD to facilitate cross-border comparisons.

The maintenance of databases like the BRD is increasingly common. Professor Elizabeth Warren's Consumer Bankruptcy Project combines PACER data with survey data to provide the basis for numerous research projects by independent scholars.³³ The Stanford Securities Litigation Clearinghouse collects documents from state- and federal-court files in securities class-action litigation and makes them available for free online.³⁴ Professor Margo Schlanger maintains a similar database with respect to civil-rights litigation.³⁵

Networks of these kinds of projects can serve as the foundations for a fully transparent court system. Today, these projects are labor intensive and hence limited in scope. In full transparency, they would be tied seamlessly together and the resulting system would process data automatically from court files to user displays.

Once the technology necessary to achieve transparency was in place and visible, it would pose a purely political issue. Should transparency's tremendous power be unleashed? The remainder of this Article assumes that the federal courts (1) eliminate all PACER user fees; (2) promulgate standardized, data-enabled PDF forms for all types of litigation and require lawyers and judges to use them; (3) continue to post the publicly available portions of case files to the Internet; and (4) impose no new barriers to

31. Compustat's conversion of the SEC data is inadequate because Compustat coverage excludes many large, public-company bankruptcies.

32. The link to Compustat is by the Standard & Poor's company identifier, GVKEY. The link to PACER is by court district and case number for the "lead" case in each corporate group.

33. The project is described generally in Elizabeth Warren, *Financial Collapse and Class Status: Who Goes Bankrupt?*, 41 OSGOODE HALL L.J. 115, 115 n.* (2003) (describing the Consumer Bankruptcy Project).

34. Stanford Law Sch., Securities Class Action Clearinghouse, <http://securities.stanford.edu> (last visited Sept. 23, 2008).

35. Wash. Univ. Sch. of Law, The Civil Rights Litigation Clearinghouse, <http://clearinghouse.wustl.edu> (last visited Sept. 23, 2008).

downloading or using the data. Based on those assumptions, the next two Parts seek to project and evaluate the consequences.

III. TRANSPARENCY'S BENEFITS

Transparency would provide an array of benefits. They include exposing and reducing corruption and impropriety, enhancing legislative control over the courts, apprising the public of the real rules by which they are governed, enabling lawyers and parties to predict the outcomes of their cases, providing a substantial new source of general knowledge, reducing legal malpractice, and increasing court-system efficiency.

A. EXPOSURE AND REDUCTION OF CORRUPTION

Government transparency is widely recognized as a deterrent to corruption.³⁶ The theory seems to be that if the government records and makes public its transactions, (1) the public is more likely to discover corrupt transactions and (2) the threat of discovery deters corrupt transactions. Discovery of a corrupt transaction is more likely in a transparent system, because the record is permanently available. Any member of the public can discover and report the transaction's corrupt nature to the government or the media.

A single agent—no matter how diligent or trustworthy—could not substitute for the public in this role. The corrupt nature of some transactions is apparent only to observers with other, privately held, information. For example, court records might show an otherwise unobjectionable decision by Judge *X* in favor of Party *Y*. Only a few members of the public might know that Party *Y* does business with *Z*, a relative or close

36. *E.g.*, *NBC Subsidiary (KNBC-TV), Inc. v. Superior Court*, 980 P.2d 337, 360 n.28 (Cal. 1999). The court noted:

[T]he public has a legitimate interest in access to . . . court documents. . . . If public court business is conducted in private, it becomes impossible to expose corruption, incompetence, inefficiency, prejudice, and favoritism. For this reason traditional Anglo-American jurisprudence distrusts secrecy in judicial proceedings and favors a policy of maximum public access to proceedings and records of judicial tribunals.

Id.; LOUIS D. BRANDEIS, *OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT* 92 (1933) ("Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman."); Ronald D. Lunau et al., *The Federal Accountability Act: Changes to Procurement and Contracting in Canada*, 42 *PROCUREMENT L.* 5, 6 (2007) ("The basic philosophy underlying the [Canadian Financial Administration Act] is that transparency will reduce government corruption."). Transparency International is an organization devoted to fighting corruption by making systems transparent. *E.g.*, TRANSPARENCY INT'L, *STRATEGIC FRAMEWORK 2008–2010*, at 7 (2007), available at http://www.transparency.org/about_us/strategy_2010 (follow "Transparency International Strategic Framework 2008–2010" hyperlink) ("Throughout our efforts to develop research and tools, our aim is to build and strengthen integrity systems, providing long-term improvement in transparency and accountability for societies.").

associate of Judge *X*. In a court system policed by the public, Judge *X* would be wise to recuse him or herself from the case, because Judge *X* could have no assurance that the relationship would not come to light. In a court system policed by a single agent, Judge *X* might feel free to decide the case corruptly, because the agent would be unlikely to know of the relationship between Judge *X* and Party *Y*.³⁷

Even when the evidence is inadequate to demonstrate corruption in any particular case, the evidence may be adequate to demonstrate corruption within a group of cases. The latter kind of evidence consists of patterns in outcomes that are unlikely to occur in the absence of corruption. Examples of such patterns include (1) patterns in the significant digits of particular numbers on tax returns that do not conform to Benford's Law;³⁸ (2) patterns in Sumo-wrestling match outcomes in which wrestlers badly needing particular wins get them and then lose the following matches to the same opponents;³⁹ and (3) patterns in standardized-test results that plausibly could only be generated by test-administrator cheating.⁴⁰

Researchers can apply such methods to judicial decisionmaking. In *The Super Crunchers*, Professor Ian Ayres points out that the random assignment of cases to the judges of a given court sets up a natural experiment.⁴¹ In such a system, judge-to-judge differences in outcomes must be attributable to one of two causes: the judges themselves or random differences in the cases assigned to the judges. Statistical methods can calculate the likelihood that a given judge-to-judge difference in case outcomes occurred randomly and thus the likelihood that it resulted from differences in the judges.

Not all judge-specific differences result from corruption. But once the statisticians have identified particular differences in outcomes as judge-specific, focused regulators can more easily find the evidence needed to prove particular outcomes to be corrupt.

B. ENHANCEMENT OF LEGISLATIVE CONTROL OVER THE COURTS

The United States has a republican form of government. Citizens elect representatives to legislatures. Those legislatures make the laws. Courts also make law, but they are supposed to do so only in situations where the

37. *Ex parte* Capital U-Drive-It, Inc., 630 S.E.2d 464, 469 (S.C. 2006) ("Public access discourages perjury and encourages bringing the truth to light because participants are less likely to testify falsely in a sunlit courtroom before their neighbors than in a private room before court officials.").

38. Theodore P. Hill, *The Difficulty of Faking Data*, 12 CHANCE 27, 31 (1999) ("Nigrini had substantial evidence that in most fabricated tax data, however, the significant digits are *not* close to Benford, and his article describes a goodness-of-fit-to-Benford test to help identify fraudulent financial data.").

39. STEVEN D. LEVITT & STEPHEN J. DUBNER, FREAKONOMICS: A ROGUE ECONOMIST EXPLORES THE HIDDEN SIDE OF EVERYTHING 40–45 (2005).

40. *Id.* at 25–38.

41. IAN AYRES, THE SUPER CRUNCHERS 71–72 (2007).

legislatures have not manifested the public will. When legislation, in the form of a constitution or a statute, conflicts with judge-made law, the system requires the judges to yield. As Professor Todd Peterson put it:

Congress . . . has the power to define the substantive law that the courts apply in the cases that come before them. Congress may not, of course, overturn the Supreme Court's interpretation of the Constitution, unless it amends the Constitution through the process specified in Article V. With respect to all other law, however, Congress has the final say. If the Congress disagrees with a Supreme Court decision on the scope of a common-law issue, Congress may enact a statute to reverse the effect of the decision. In the same way, if Congress disagrees with the Supreme Court's interpretation of a statute, it may amend the statute to overrule the Court's interpretation, even if it does so with respect to a specific case that is still pending in the courts. About the only limitation on Congress's power over substantive law is that it may not change the result in a particular case once the Court has issued a final decree in that case.⁴²

The same is true with respect to the interpretation of a statute. Assuming that a statute is constitutional, the job of the courts is to vindicate the statutory scheme enacted by the legislature. Pure textualists think that judges should be bound by the precise language of the statutes before them, while more nuanced interpretivists think that judges should seek to implement the legislature intent. Both, however, agree that the courts are not free to ignore statutes.⁴³

Despite this theory, legislatures often find it difficult to control the courts, especially when the nature of the legislative policy requires the delegation of discretion to the courts. Particular judges may be hostile to a given policy and deliberately make decisions that frustrate it. Other judges may not understand the policy, yielding the same outcome. As a result, judges often fail to implement the legislature's policies.

Court-system transparency could show legislatures and the public how the courts implement laws. If Congress made changes in the bankruptcy system designed to increase creditors' recoveries, Congress would know whether creditors' recoveries increased. If it made changes intended to reduce the rate of recidivism by sex offenders, it would know the extent to which recidivism decreased.⁴⁴ This feedback would alert the legislatures to

42. Todd David Peterson, *Congressional Investigations of Federal Judges*, 90 IOWA L. REV. 1, 36 (2004).

43. Pamela S. Karlen, *Two Concepts of Judicial Independence*, 72 S. CAL. L. REV. 535, 544 (1999).

44. Such information is not available today. The government published its only report on sex-offender recidivism in 2003. That report examined sex offenders released in 1994 and

change statutory language when necessary to specify the legislators' intentions more clearly. The process would be benign because it would occur in full public view.

Feedback would also assist judges in their efforts to faithfully implement laws and policies. For example, the Bankruptcy Code requires judges to find, as a pre-condition to chapter 11 plan confirmation, that "[c]onfirmation of the plan is not likely to be followed by the liquidation, or the need for further financial reorganization, of the debtor"⁴⁵ From 1991 through 1996, the Delaware bankruptcy judges made such a finding in each large public-company bankruptcy. Those findings were systematically incorrect. Within five years after confirmation, fifty-four percent of the reorganized companies had filed a second bankruptcy or liquidated.⁴⁶ The judges remained unaware of their failure for years after their decisions, because no mechanism existed to identify the failure or bring it to their attention.

Court-system transparency would provide such a mechanism. Feedback from researchers would enable individual judges to adjust their future rulings to comply with statutes. In this example, the researchers could, in addition to highlighting the judges' noncompliance, identify the subgroups of judges or cases in which noncompliance was highest. The judges could then give greater scrutiny to the cases in those subgroups, and by doing so, achieve compliance.

C. POPULARIZATION OF THE LAW

Law is often thought of merely as a mechanism for resolving disputes. But an even more important function of law is to apprise citizens of the rules that govern them. As every first-year law student learns, the law in its current form—cases and statutes—does a poor job of that. The “rules” are not rules at all, but merely vague, generalized standards. Those standards usually take on concrete meaning only when the courts implement them in large numbers of cases.

By revealing the patterns in those cases to the public, court-system transparency would, for the first time, actually apprise citizens of the rules that govern them. The information would reach the public through a series of steps. For example, the first step might be regression analysis by a legal academic on the body of cases challenging clickwrap agreements. That analysis might discover the circumstances in which courts enforce particular

provided only three-year recidivism rates. Thus, the data were final more than five years before the government released the report. U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, RECIDIVISM OF SEX OFFENDERS RELEASED FROM PRISON IN 1994, at 5–6 (2003), *available at* <http://www.ojp.usdoj.gov/bjs/pub/pdf/rsorp94.pdf> (defining re-arrest and reconviction as occurring within three years). The researchers did not attempt to evaluate any changes in law.

45. 11 U.S.C. § 1129(a)(11) (1994).

46. LYNN M. LOPUCKI, *COURTING FAILURE: HOW COMPETITION FOR BIG CASES IS CORRUPTING THE BANKRUPTCY COURTS* 113 tbl.6 (2005).

contract terms and the statistical likelihood of such enforcement. In a second step, the analyst might publish conclusions in an academic journal, provide them to the media, display them on a website, or embed them in software to make them easily available to users when needed. The information would enable users to more intelligently decide whether or in what circumstances to enter into click-wrap agreements.

If, for example, the data showed that that courts generally enforce oppressive clickwrap agreements, users might be reluctant to sign them but still lack practical alternatives. If that were the case, however, consumer advocates could press for a change in the law. With data, the advocates would be more likely to succeed. They could prove that a problem exists. Members of the public would be aware of the problem and could join in requesting a solution.

The revelations of a transparent court system would not be completely unbiased. Researchers may have their own agendas. But in a system that made research easy, researchers would compete for public attention by seeking to establish credibility. Researchers would do that by making both their data and data processing transparent. Users want accurate, unbiased views. In open, transparent contests between more and less accurate views, the more accurate views generally would prevail.

D. PREDICTION OF LITIGATION OUTCOMES

The practice of law consists principally of advising, planning, and litigating. Each of these three activities requires lawyers to predict what courts will do. The centrality of prediction in the legal process is captured in Oliver Wendell Holmes's definition of law. Over a century ago, he wrote that "[t]he prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law."⁴⁷

Holmes advocated an essentially empirical approach to predicting legal outcomes.⁴⁸ The data sources he contemplated—court opinions, statutes, regulations, and commentary—continue to dominate legal prediction.⁴⁹ In

47. Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 461 (1897).

48. Catharine Pierce Wells, *Holmes on Legal Method: The Predictive Theory of Law as an Instance of Scientific Method*, 18 S. ILL. U. L.J. 329, 335–42 (1994) (arguing that Holmes's conception of law is empirical); *id.* at 341 (stating that Holmes's conception "must be recognized as an empirical theory of law").

49. *Id.* at 341. ("[Holmes's] case method begins with legal cases. They are the data that must be explained."). Holmes himself wrote:

The means of the study are a body of reports, of treatises, and of statutes, in this country and in England, extending back for six hundred years, and now increasing annually by hundreds. In these sibylline leaves are gathered the scattered prophecies of the past upon the cases in which the axe will fall. These are what properly have been called the oracles of the law. Far the most important and pretty nearly the whole meaning of every new effort of legal thought is to make these

addition, lawyers rely heavily on their personal knowledge of judges and opinion-less decisions, and on other lawyers' knowledge shared through personal conversations, continuing-legal-education programs, and published commentary.

With respect to court opinions, the processing method that Holmes recommended still dominates. The lawyer finds and reads the opinions most relevant to the case at hand, extracts a governing principle by inductive reasoning, and then applies the principle deductively to the case at hand. The result is the lawyer's prediction of what a court will do.

To the extent that opinions interpreting statutes, regulations, and ordinances exist, the same method applies. To the extent that they do not exist, the lawyer works from the statute's text. The lawyer reads the text, determines whether the facts of the case at issue are within its scope, and if so, the lawyer applies the statute deductively to reach a prediction.

1. Can Lawyers Predict the Outcomes of Cases?

Lawyers often assume that their legal expertise equips them to predict outcomes better than a computer program running mechanically extractable relational data. For at least some kinds of legal problems, however, data-based analysis can already outperform legal experts. A recent experiment in predicting Supreme Court decisions illustrates the point.⁵⁰ The experiment pitted the aggregate predictions of eighty-three law professors and appellate-attorney experts against a regression analysis that required only six easily available pieces of data for each case: "(1) circuit of origin; (2) issue area of the case; (3) type of petitioner (e.g., the United States, an employer, etc.); (4) type of respondents; (5) ideological direction (liberal or conservative) of the lower-court ruling; (6) whether the petitioner argued that a law or practice is unconstitutional."⁵¹ The experts accurately predicted the Supreme Court's decision in only fifty-nine percent of the sixty-seven cases; the regression analysis accurately predicted the Court's decision in seventy-five percent of those cases.⁵²

Lawyers also assume that they can see the patterns in the cases they litigate. In fact, their ability to do so is quite limited. The cases that a given lawyer knows about are likely too small a number to exhibit the general pattern. Even if the important patterns occur in the cases that the lawyer knows about, those patterns may be too subtle to notice. For example, a

prophecies more precise, and to generalize them into a thoroughly connected system.

Holmes, *supra* note 47, at 457.

50. Theodore Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150 *passim* (2004).

51. *Id.* at 1154 n.19.

52. *Id.* at 1150 ("The model predicted 75% of the Court's affirm/reverse results correctly, while the experts collectively got 59.1% right.").

prominent bankruptcy practitioner recently noted that “[i]t was once believed that some insolvent enterprises would be difficult to sell for full value,” but claimed that “the size of the distressed business no longer operates as a constraint on sale . . . and . . . buyers have become less likely to impose an ‘insolvency’ discount in connection with such a sale.”⁵³ Almost simultaneously with that publication, Joseph Doherty and I released empirical results showing that insolvent enterprises, on average, sold for only half of what they would have been worth in reorganization.⁵⁴

The recent discovery of high refiling and failure rates among Delaware-reorganized large, public companies provides a second example. By historical accident, the law entitles bankrupt large, public companies to choose their bankruptcy courts.⁵⁵ In the early 1990s, companies suddenly began choosing the relatively inexperienced, one-judge court in Wilmington, Delaware. The Delaware court’s market share increased steadily—from zero in the decade prior to 1990 to eighty-seven percent of all large, public-company bankruptcies filed in the United States in 1996 (thirteen of fifteen cases).⁵⁶ Bankruptcy professionals and academics lauded the Delaware court as the most sophisticated in the nation, and other courts began emulating its methods.⁵⁷

In March 2000, Sara Kalin added data on the ultimate fate of the companies reorganized from 1990 through 1996 to the Bankruptcy Research Database. Purely by accident of juxtaposition, we noticed that the companies filing in Delaware and New York (the second most popular court) failed at rates two to seven times as high as those filing in all other courts.⁵⁸ The lawyers thought the Delaware and New York courts were doing the best job of reorganizing companies when in fact they were doing the worst.⁵⁹ The

53. Donald S. Bernstein, *U.S. Chapter 11 Today: A Funny Thing Happened on the Way to the Courthouse*, in THE INTERNATIONAL COMPARATIVE LEGAL GUIDE TO: CORPORATE RECOVERY AND INSOLVENCY 2007, at 6 (2007). Bernstein was hardly the only one to misperceive the pattern. Lynn M. LoPucki & Joseph W. Doherty, *Bankruptcy Fire Sales*, 106 MICH. L. REV. 1, 10–11 (2007) (quoting leading bankruptcy academics to the same effect).

54. LoPucki & Doherty, *supra* note 53, at 24.

55. LOPUCKI, *supra* note 46, at 30–39 (describing the application of bankruptcy venue rules).

56. *Id.* at 49–50.

57. *Id.* at 123–28.

58. For example, forty-two percent of the companies reorganized in the Delaware court from 1991 to 1996 were back in bankruptcy within five years. The corresponding rate for companies reorganized in the New York court was nineteen percent. The corresponding rate for companies reorganized in all other courts was six percent.

59. Ayotte and Skeel have offered a spirited defense of the Delaware courts’ poor performance. Kenneth Ayotte & David A. Skeel, Jr., *An Efficiency-Based Explanation for Current Corporate Reorganization Practice*, 73 U. CHI. L. REV. 425 *passim* (2006). We disagree. Lynn M. LoPucki & Joseph W. Doherty, *Delaware Bankruptcy: Failure in the Ascendancy*, 73 U. CHI. L. REV. 1387 *passim* (2006). The point made here, however, does not depend on the resolution of that

difference in the performances of the three sets of courts during that period was statistically significant at the .001 level.⁶⁰

The discovery of this pattern did not require statistical analysis, but it did require data. That is, the pattern was obvious if one looked at a list that showed the name of the emerging company, the reorganization court, and whether the company refiled. Nearly every refiler on the list was among the minority of companies that had been reorganized in Delaware or New York. The lawyers did not, however, have such a list. None noticed the pattern until we announced our findings in June 2000.⁶¹ Lawyers may become so absorbed in the specific case that even the most prominent pattern in case outcomes can be invisible to them.

2. Can Statistical Analysis Predict the Outcomes of Cases?

The social sciences employ a predictive method far more powerful than a lawyer's intuition: regression analysis. The researcher who seeks to predict a judicial decision begins by identifying past cases in which similar decisions were made. The researcher forms a hypothesis regarding the identity of the factors that determine the outcomes and collects data regarding the factors and outcomes. The researcher then determines mathematically which factors correlate most highly with the outcomes. Based on those correlations, the researcher estimates a mathematical formula that determines the odds of a particular outcome under any combination of factor values.

The LoPucki–Doherty Professional Fees Calculator (the “Calculator”) illustrates the process concretely. Joseph Doherty and I conducted a study of court-awarded professional fees in large, public-company bankruptcies. Using the Bankruptcy Research Database, we identified a sample of 102 large, public-company bankruptcies. From the court files on PACER, we collected the amounts of professional fees awarded and the values of several case characteristics (“independent variables”) that we thought might determine the amounts of the fees and expenses (the “dependent variable”). Among the independent variables were several measures of the size of the debtor, several measures of the complexity of the case, several measures of the numbers of parties involved in the case, several measures of the duration of the case, the identities of the professionals and judges, and some characteristics of those professionals and judges.

Through a reiterative process—part computerized mathematics, part human judgment—we determined from the data what we considered to be the “best” predictive model for determining the total amount of professional

disagreement. Even if Ayotte and Skeel were right, the fact would remain that a distinct pattern existed in the outcomes of the lawyers' cases; yet, the lawyers did not know about it.

60. LoPucki & Doherty, *supra* note 2, at 1939–45.

61. No reports of higher refiling or refailure rates in Delaware were made prior to the release of the results of our studies.

fees and expenses that a large, public company would incur in bankruptcy. A regression model is essentially a binomial mathematical formula in which one term is the value of an independent variable and the other—the “coefficient”—indicates that variable’s level of importance in predicting the dependent variable. Here, the dependent variable is the amount of the professional fees and expenses in a case. By way of illustration, this is one of the models we estimated:

$$F = (A \times 0.694) + (D \times 0.161) + (N \times -0.091)$$

F = the log of the fees and expenses awarded to professionals in the case in dollars

A = the log of the debtor’s assets reported in dollars

D = whether the case is in the Delaware bankruptcy court, expressed as “1” or “0”

N = whether the case is in the New York bankruptcy court, expressed as “1” or “0”

The formula tells us that the log of fees and expenses is on average equal to .694 times the log of the debtor’s assets plus 0.161 if the case is in Delaware, or minus 0.091 if the case is in New York. From the assets and the court location, this formula predicts the fees and expenses.

What model—that is, formula—would predict fees and expenses “best” is a matter of opinion. But the accuracy with which a given model predicts them within a data set is a mathematically measurable fact. Add one more assumption—that the relationship of the independent variables to the dependent variable will be the same in future cases as in past cases—and what you have is mathematical prediction of the law, that is, “what officials do about disputes.”⁶²

The model that we concluded was best in predicting professional fees and expenses in large, public-company bankruptcies employs six variables: (1) the debtor’s assets as reported on the bankruptcy petition, (2) the length of the case from filing to plan confirmation, (3) the number of professional firms authorized to work in the case, (4) the year the case is filed, (5) the number of people the debtor employed before filing, and (6) whether the debtor forum-shops to a court away from its headquarters. Employed in the mathematical combination indicated by our analysis, these factors explained ninety-one percent of the case-to-case variance in fees and expenses in the 102 cases we studied.

The last step in the prediction process was to provide our discovery to the public in an easily usable form. The Calculator is a computer program that accepts values for the independent variables in our model, plugs them

62. KARL LLEWELLYN, *THE BRAMBLE BUSH* 12 (1951) (“*What these officials do about disputes is, to my mind, the law itself.*”).

into the mathematical formulas of our models, and returns the predicted dollar amount of fees and expenses, along with the eighty percent confidence levels for the prediction.⁶³ We posted the Calculator on a website.⁶⁴

If the court system were transparent, models such as these would be ubiquitous. Together, they would predict virtually every predictable aspect of judicial decisionmaking.

Regression analysis not only identifies the factors that cause a given set of outcomes, it also quantifies the relationship between the factors and the outcomes. The quantification of predictions—in the form of “odds” or chances—can be a crucial aid to legal planning.

A recent study of medical-malpractice recoveries illustrates the point.⁶⁵ Using court-file data, the researchers found that plaintiffs represented by experienced lawyers were more likely to recover money than plaintiffs represented by inexperienced lawyers. Anyone might have guessed that without the need for a study. The importance of the study was in its quantification of the effect. Plaintiffs with experienced lawyers facing inexperienced defense lawyers recovered in eighty percent of their cases, while plaintiffs with inexperienced lawyers facing experienced defense lawyers recovered in less than forty percent of their cases.⁶⁶

A transparent court system would not offer an answer to every legal question. Parties often settle cases and issues before resolution, and in most instances, the courts do not require the parties to reveal the terms of settlements.⁶⁷ A transparent court system would, however, provide far more answers than the current system provides, because it would take into account a far higher number of decisions and more relevant information about those decisions.

63. THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE 386 (4th ed. 2000) (defining “confidence interval” as “[a] statistical range with a specified probability that a given parameter lies within the range”). As applied to our fee calculator, there is an eighty-percent probability that the actual amount of fees and expenses for a case of the kind in the data set will be within the eighty-percent confidence level.

64. Bankr. Research Database, The LoPucki–Doherty Professional Fees Calculator, <http://lopucki.law.ucla.edu/feecalculator.asp> (last visited Sept. 23, 2008).

65. Harris et al., *supra* note 3, *passim*.

66. *Id.* at 243 tbl.6.

67. *E.g.*, *Goodyear Tire & Rubber Co. v. Chiles Power Supply, Inc.*, 332 F.3d 976, 980 (6th Cir. 2003) (“Secrecy of settlement terms . . . is a well-established American litigation practice.” (quoting *In re Franklin Nat’l Bank*, 92 F.R.D. 468, 472 (E.D.N.Y. 1981))); Scott A. Moss, *Illuminating Secrecy: A New Economic Analysis of Confidential Settlements*, 105 MICH. L. REV. 867, 869 (2007) (“Courts regularly allow confidentiality provisions; indeed, under existing law, they cannot force parties to disclose settlement terms they had agreed to keep confidential.”).

3. The Inadequacy of Court Opinion Data

The courts' reported opinions are already freely available. Researchers have done numerous regression analyses of case outcomes based on data collected from them.⁶⁸ Reported opinions are, however, of limited use in mathematically predicting outcomes. First, opinions are available for only a small minority of decisions.⁶⁹ Because researchers have less data to work with, analyses based on opinions are less probative than those based on decisions. Second, opinions are notoriously nonrepresentative of all decisions.⁷⁰ Third, opinion data are text-based. Researchers must convert the data to relational format for statistical analysis.⁷¹ That prevents automated analysis—at least until judges begin issuing their opinions in relational formats. It also means that variables of interest are often missing from the opinions, rendering the data less powerful.⁷² Fourth, a court opinion is often a biased description of the decided case. The writer may be unwilling to explain the real reasons for the decision or may even be unaware of them.

Professor Frederick Schauer suggests that the variables that are the true determinants of case outcomes might not even be disclosed in court opinions.

[I]f we were to undertake a statistical analysis of “the law” in order best to engage in the process of predicting future legal outcomes, we would, in some form or other, look to identify the variables that had the greatest predictive value. These variables might, as Holmes suspects, be the variables of legal doctrinal categorization. But whether the variables were in fact what Holmes suspected—and desired—would be an empirical question, and it might turn out, as

68. David Sherwyn et al., *Don't Train Your Employees and Cancel Your "1-800" Harassment Hotline: An Empirical Examination and Correction of the Flaws in the Affirmative Defense to Sexual Harassment Charges*, 69 *FORDHAM L. REV.* 1265, 1275 (2001) (acknowledging that their study from reported opinions “may or may not represent or closely resemble the entire legal universe”); Robert B. Thompson, *Piercing the Corporate Veil: An Empirical Study*, 76 *CORNELL L. REV.* 1036, 1046 (1991) (“These results are based on reported cases that may or may not be a representative sample”); James J. White, *Revising Article 9 to Reduce Wasteful Litigation*, 26 *LOY. L.A. L. REV.* 823, 838 (1993) (“[S]urely not all of the opinions were reported, and I suspect that my reported opinions are a small minority of the total.”).

69. Lynn M. LoPucki & Walter O. Weyrauch, *A Theory of Legal Strategy*, 49 *DUKE L.J.* 1405, 1438 n.170 (2000) (providing statistics supporting the conclusion that “[i]n the large majority of their cases, American judges write neither opinions nor explanatory orders”).

70. *E.g.*, Theodore Eisenberg & Stewart J. Schwab, *What Shapes Perceptions of the Federal Court System?*, 56 *U. CHI. L. REV.* 501, 539 (1989) (“Readers of published appellate opinions perceive a high percentage of successful constitutional tort cases. Those tracking the outcome of district court cases see a far different reality.”).

71. *E.g.*, Thompson, *supra* note 68, at 1044–45 (describing the coding of veil-piercing opinions).

72. Requiring courts to render opinions in relational formats could solve both these problems. I am unaware of any effort to impose such a requirement.

Llewellyn suspected to the contrary, that they were variables not likely to be identified from the opinions of the courts that reached those decisions.⁷³

The transparent court system responds to Schauer's concern by offering the entire case record for analysis rather than just the court's opinion. Using the entire record minimizes missing data and vastly multiplies the numbers of cases and variables that might be used in predictions.

Even these vast multiplications do not necessarily allay Schauer's fears. He continues:

But what if legal outcomes are not amenable to categorization? Although the possibility that legal outcomes may be totally random seems too remote to be taken seriously, it could still be the case that no variable had any substantial amount of predictive power, such that no single factor, and even no collection of factors, could provide with any confidence a prediction of a future legal outcome. Even the kind of empirical analysis that Llewellyn championed, and even when that analysis was done with the best possible tools of multiple regression, might simply yield the conclusion that no identifiable variable yielded a useful correlation with decisional outcomes.⁷⁴

Schauer acknowledges that the concern he raises is itself an empirical question. The answer to that question is already partially in. At least some kinds of legal outcomes are startlingly predictable. As previously mentioned, a simple computerized algorithm outperformed constitutional-law experts in predicting decisions of the U.S. Supreme Court.⁷⁵ In a study of punitive damages, a group of researchers found that "unless the case involves an intentional tort or a business-related tort (such as employment claims), punitive damages will almost never be awarded."⁷⁶ In a study of 1600 bankruptcy-court case files, Teresa Sullivan, Elizabeth Warren, and Jay Westbrook did not find a single case in which the court denied discharge.⁷⁷ In a study of 101 death-penalty case files, two journalists found that the identities of the judges explained the largest disparities in treatment.⁷⁸ Using a database of all federal trial and appeals, Kevin Clermont and Theodore Eisenberg found that defendants were nearly three times as likely

73. Frederick Schauer, *Prediction and Particularity*, 78 B.U. L. REV. 773, 783–84 (1998).

74. *Id.* at 786.

75. See *supra* notes 50–52 and accompanying text.

76. Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 J. LEGAL STUD. 623, 659 (1997).

77. TERESA A. SULLIVAN, ELIZABETH WARREN & JAY LAWRENCE WESTBROOK, AS WE FORGIVE OUR DEBTORS: BANKRUPTCY AND CONSUMER CREDIT IN AMERICA 265–66 n.11 (1989).

78. Elias & Fried, *supra* note 2, *passim* (finding huge variations from judge to judge in the length and disposition of death-penalty cases).

as plaintiffs to obtain reversals on appeal.⁷⁹ This list could easily be expanded.⁸⁰

4. Prediction's Systemic Effects

Greater predictability of litigation outcomes would have important impacts, not only on the court system, but on society as a whole. Some scholars argue that court-system predictability promotes private-sector economic planning and, ultimately, increases productivity.⁸¹ Those engaged in legal-planning activities of any kind would have sounder bases on which to proceed. Doctors deciding whether to practice “defensive medicine” could know the extent to which their efforts actually reduced liability risk.⁸² Parties to clickwrap agreements⁸³ or liability releases⁸⁴ could know the odds that those agreements or releases would be effective. Persons hesitating to become directors of corporations, to offer useful but risky products in the marketplace, or to agree to seemingly draconian contract terms would be able to quantify the liability threats that they would face. In short, those wanting to plan for legal liability would be able to do so more effectively.

Greater predictability would also impact case processing. Judges seeking to provide “horizontal equity” to litigants by deciding their cases consistently with the decisions of other judges⁸⁵ would have the necessary data. They could decide in accord with the predicted outcome.

Statistical analysis and data management would become important legal skills. Predictability would facilitate settlement by reducing the differences in

79. Kevin M. Clermont & Theodore Eisenberg, *Plaintiphobia in the Appellate Courts: Civil Rights Really Do Differ from Negotiable Instruments*, 2002 U. ILL. L. REV. 947, 947.

80. E.g., Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119 *passim* (2002) (surveying empirical studies of civil litigation).

81. Peter Boettke & J. Robert Subrick, *Rule of Law, Development, and Human Capabilities*, 10 SUP. CT. ECON. REV. 109, 111 (2002) (“[T]he rule of law provides us with the stability and predictability in economic affairs required for agents to engage in entrepreneurial action—both in terms of exploiting existing opportunities for profit through arbitrage and the discovery of new profit opportunities through innovation.”); Edward A. Morse, *Reflections on the Rule of Law and “Clear Reflection of Income”: What Constrains Discretion?*, 8 CORNELL J.L. & PUB. POL’Y 445, 457 n.52 (1999) (“Predictability may also increase productivity, to the extent that economic commitments can be made in reliance upon a stable legal structure.”).

82. Michelle M. Mello & Troyen A. Brennan, *Deterrence of Medical Errors: Theory and Evidence for Malpractice Reform*, 80 TEX. L. REV. 1595, 1606 (2002) (defining “defensive medicine” as “care provided solely (or mostly) to reduce the probability of litigation”).

83. William J. Condon, Jr., *Electronic Assent to Online Contracts: Do Courts Consistently Enforce Clickwrap Agreements?*, 16 REGENT U. L. REV. 433 *passim* (2003) (attempting to answer that question based on reported opinions).

84. Ryan S. Holcomb, *The Validity and Effectiveness of Pre-Injury Releases of Gross Negligence in Texas*, 50 BAYLOR L. REV. 233 *passim* (1998) (attempting to explain when one narrow category of releases will be effective based on published opinions).

85. See, e.g., *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 395 (2006) (reiterating the basic legal principle that “like cases should be decided alike”).

the parties' expectations as to outcomes.⁸⁶ To the extent that parties litigate in order to resolve outcome uncertainty, that litigation would be less necessary. Cost savings might result.

Despite this tendency toward fewer cases and lower costs, net reductions in caseloads and costs likely would be modest. Parties litigate for many reasons, and dispute resolution is only one of them. Litigation would continue in order to obtain discovery,⁸⁷ delay,⁸⁸ or control over adversaries.⁸⁹ If third parties such as lenders or rating agencies concluded that they reliably could predict litigation outcomes, financing litigation would become easier, and people might bring more claims.⁹⁰

Predictability would also benefit particular parties by informing their litigation strategies. A plaintiff that knew its likelihood of winning in both state and federal court could choose the "better" court from its perspective. Parties could do cost-benefit analyses to determine what issues were worth raising.

Predictability would enable parties who contemplated illegal action—Holmes's "bad man"—to calculate the likely consequences. If the legal system were well-designed, that predictability would not be a problem. The pattern of incentives would encourage socially productive conduct. At present, however, the legal system is not well-designed. Armed with outcome predictions, the Holmesian bad man might have a field day. The legislature that authorizes court-system transparency should stand by to fix the

86. See, e.g., Adam M. Samaha, *Judicial Transparency in an Age of Prediction*, 53 VILL. L. REV. (forthcoming 2008) (manuscript at 11, on file with the Iowa Law Review) (asserting that if the equation for prediction of legal outcomes "is available to all sides, the number of lawsuits might remain at zero").

87. See, e.g., *id.* ("Another hindrance to settlement might be that party A cannot evaluate the strength of her position without using the tools of court-supported discovery to extract information withheld by party B."). Parties also abuse discovery in order to obtain information for a variety of other purposes. Such abuses might continue in a transparent court system, but the likelihood of discovery and sanction probably would increase.

88. Defendants who know full well that they will lose their cases often continue to litigate them to obtain delay. This motive dominates defense of what generally are referred to as "debt collection" cases. Debtors may use the time to obtain money to pay or to judgment-proof themselves.

89. The existence of litigation alters one's relationship with adversaries. The control may be formal, as when a bankruptcy filing imposes an automatic stay, or informal, as when one's opponent "voluntarily" stops the conduct complained of in the lawsuit to avoid the possibility of offending a judge or jury pending a decision. Either kind of relationship can be viewed as a method of controlling the adversary.

90. Empirical studies indicate that the large majority of valid legal claims are never brought. PAUL C. WEILER, A MEASURE OF MALPRACTICE: MEDICAL INJURY, MALPRACTICE LITIGATION, AND PATIENT COMPENSATION 73 (1993) (estimating that only one in fifty patients who suffer negligent medical injury file malpractice claims).

numerous inconsistencies in the legal system that the bad men would discover through transparency and seek to exploit.⁹¹

A certain amount of embarrassment for public officials is inevitable. For example, when Paris Hilton went to jail in Los Angeles, she brought the media spotlight with her, creating her own court-transparency hotspot. I, like most Los Angeles residents, learned for the first time that a sentence of forty-five days in jail actually means three days in jail.⁹² The apparent purpose of the courts in using “forty-five” to mean “three” was to be able to give short sentences while appearing to the public to give long ones. The media spotlight thwarted that purpose and exposed the officials involved to criticism.

In a transparent court system, such embarrassments would occur more frequently because the media could more easily discover and document inconsistencies. That, however, is a positive attribute of a transparent court system. Embarrassment and the threat of embarrassment are the very mechanisms by which transparency leads to improvement.

Because we know so little about what actually happens in the courts, predictions of the levels of embarrassment that would result from transparency are difficult to make. My guess is that levels initially would be high. Two embarrassing factors—the identity of the judge and the quality of legal representation—probably would prove to be the most powerful predictors of who wins in court.⁹³ In addition, researchers have already documented huge differences from city to city in the way that courts apply the same laws.⁹⁴

In some circumstances, the law requires courts to predict outcomes in other courts. For example, in diversity cases, the law requires federal courts to predict what state courts would do.⁹⁵ There are numerous other

91. Samaha, *supra* note 86, at 24 (“Perhaps these actors would use the information [from court transparency] to bend the courts to an injurious political will, or to ‘game the system.’”).

92. Los Angeles attorney Robert Shapiro explained the practice on CNN: “Somebody gets a 45-day sentence. They go in. They either get booked and released immediately or serve a maximum of three days. That’s reality. I’m not saying it’s right. I’m not saying it’s wrong. That’s what happens in almost every single case.” *Anderson Cooper 360 Degrees: Paris Hilton Back in Jail* (CNN television broadcast June 8, 2007), transcript available at <http://transcripts.cnn.com/TRANSCRIPTS/0706/08/acd.01.html>.

93. See authorities cited *supra* notes 2–3.

94. LoPucki, *Legal Culture*, *supra* note 6, at 1504–08 (1996) (documenting sharp geographical differences in subordination of insider debt in chapter 11 cases); Teresa A. Sullivan, Elizabeth Warren & Jay Lawrence Westbrook, *The Persistence of Local Legal Culture: Twenty Years of Experience from the Federal Bankruptcy Courts*, 17 HARV. J.L. & PUB. POL’Y 801, 822–26 (1994) (documenting sharp geographical differences in debtors’ exercise of the choice between chapter 7 and chapter 13).

95. *Erie R.R. Co. v. Tompkins*, 304 U.S. 64, 78 (1938) (holding that a federal court sitting in diversity must apply the substantive law of the relevant state); Evan H. Caminker, *Precedent and Prediction: The Forward-Looking Aspects of Inferior Court Decisionmaking*, 73 TEX. L. REV. 1, 5 (1994) (“In [diversity] cases, the courts’ task is to try to *predict* how the highest court of that

examples.⁹⁶ These requirements facially suggest that predictions based on court-file data might be helpful.

Unfortunately, they would not be. As Professor Michael Dorf convincingly argued, the prediction contemplated by those requirements is fundamentally different from the prediction under discussion here.⁹⁷ To borrow Schauer's example, federal courts in West Virginia should not rule in favor of coal companies, because coal companies always win in that state's highest court.⁹⁸ The obligation to predict what another court would do should be interpreted to require the predicting court to work solely from "impersonal principles" of law.⁹⁹ In other words, the courts should deliberately reject the best predictors—the actual pattern of past outcomes—and continue to use the legal materials recommended by Holmes.

Transparency's contribution to those required predictions would be to expose differences in results from the two kinds of analyses. If regression analysis showed that the coal-company variable is the best predictor of victory in the West Virginia Supreme Court, that finding would prove the West Virginia Supreme Court deficient.¹⁰⁰ If, on the other hand, the factors identified by impersonal principles of law caused legal outcomes, those factors would dominate regression models. If nonlegal factors were not persistently significant in the models, that would show that the legal system was working as it was supposed to work.

In sum, greater predictability may or may not significantly reduce litigation. But even if it does not, predictability would tend to increase the social utility of the litigation that persists. Transparency would tend to reform the court system in accord with policymakers' preferences. That

State would decide the question." (quoting *Salve Regina Coll. v. Russell*, 499 U.S. 225, 241 (1991))).

96. See Michael C. Dorf, *Prediction and the Rule of Law*, 42 UCLA L. REV. 651, 690–95 (1995) (discussing the obligation of a single U.S. Supreme Court justice to predict whether four justices will vote to grant certiorari).

97. *Id.*

98. See Schauer, *supra* note 74, at 783. Schauer compares two predictors of victory in the West Virginia Supreme Court:

For if one actually looks at the cases dealing with injunctions decided by the West Virginia Supreme Court of Appeals from 1920 to 1954, one would likely discover that the principle "the coal company wins" has substantially more predictive power than the principle "a party who delays claiming its rights to the detrimental reliance of another party is precluded from obtaining an injunction."

Id.

99. Dorf, *supra* note 97, at 686 (referring to "the requirement that judges justify their decisions according to impersonal principles").

100. I assume that variables representing the factors that legally should be controlling the outcomes have been included in the regression analysis and found to be less important.

would strengthen incentives to bring the types of litigation that serve the public interest.

E. ENHANCEMENT OF THE INFORMATION INFRASTRUCTURE

Court-system transparency would also provide the public with a valuable source of general information. In our society, the most valuable kinds of information are not for sale to the public. They include information about specific individuals, the many kinds of lifestyles available in society, and how people enter or succeed in an occupation or business, to name just a few.¹⁰¹

Litigated matters, from the wills of the rich and famous¹⁰² to the bankruptcies of large public companies,¹⁰³ are forced into public view. In the process, otherwise private information becomes public. The fact-rich adventure stories of the litigants may not always be the truth, but the adversarial process through which the facts emerge does as much as humanly possible to test them. Legal education and law practice have always been valued for the unique window they offer into a variety of lives, occupations, and activities. By offering direct access to court records, a transparent court system would vastly multiply the capacity and effectiveness of this window.

Litigation is when the facts come out. Litigation over police tactics makes police tactics visible. Litigation over the safety of a drug or product reveals all of the research on that drug or product—whether published or “proprietary.” Litigation over an asset-securitization transaction may bring to light how parties constructed the transaction. Litigation over organized crime exposes its inner workings. Experts on every conceivable scientific issue and a vast array of unscientific ones testify as witnesses in litigation. A transparent court system would make all of that knowledge readily available to those who want or need it. That knowledge would break up information monopolies, promote competition, and ultimately increase economic activity. Because the courts are among the most information-rich institutions in society, transparent courts would be among society’s principal suppliers of knowledge.

F. REDUCTION OF LAWYER AND LITIGANT ERROR

Two of the most common kinds of lawyer malpractice are (1) missing a deadline and (2) omitting necessary information from a filed document.

101. Ersatz versions are, of course, readily available in the form of newspapers, magazines, websites, books, and films, but those come with scant means for assessing their accuracy.

102. *E.g.*, Frances H. Foster, *Trust Privacy*, 93 CORNELL L. REV. 555, 559–67 (2008) (describing the distinction between wills, which are public, and trusts, which are private, while including a few details from the wills of Jacqueline Kennedy Onassis and Doris Duke).

103. *E.g.*, LOPUCKI, *supra* note 47, *passim* (using numerous examples from court files to describe the large-public-company-bankruptcy process).

Court-system transparency could almost completely eliminate both kinds of errors.

Consider, for example, the rule that a trustee in a chapter 7 bankruptcy case must assume an executory contract within sixty days after the order for relief.¹⁰⁴ Under current practice, it is the responsibility of the trustee to know of this deadline and comply with it. In a transparent court system, the entry of an order for relief, the appointment of a trustee, and the filing of a schedule of executory contracts could, in combination, automatically trigger notice to the trustee of the deadline's existence and a series of increasingly insistent reminders as the deadline approached. The reminders would cease if the trustee filed an election to assume, or indicated his or her intention not to assume, each executory contract listed.

In the current system, lawyers commonly file documents that omit necessary information. A transparent court system could alert the filing lawyer of the insufficiency of the filed document in the same way that the order form on a commercial website alerts buyers that they have not furnished a required email address or a sixteen-digit credit card number. The system would immediately afford the lawyer the opportunity to correct the error.

Newly instituted federal-court privacy rules prohibit the inclusion of Social Security numbers and a few other kinds of personal information in court files. Commentators urge lawyers to vet the documents they file for violations. In a transparent court system, the courts' computers could automatically check that vetting before the system accepted each document for filing.

Automation would also provide the foundation necessary for the application of artificial intelligence to litigation. That is, based on the status of the case as gleaned from the relational data in the court file, an "expert" algorithm could make suggestions to the lawyers. For example, the algorithm might suggest the possibility of bankruptcy or Truth in Lending Act claims to the defendant in a mortgage-foreclosure action.¹⁰⁵ When a party seeks to pierce a corporate veil the algorithm might suggest consideration of the related doctrines of agency and direct action. In the foreclosure of a security interest in personal property, the algorithm might

104. 11 U.S.C. § 365(d)(1) (2000).

105. Katherine Porter, *Owning Up: Homeowners in Bankruptcy* (2007) (unpublished manuscript, on file with the Iowa Law Review) (reporting a Hale & Dorr Legal Services Center study finding that six of twenty-two mortgagees (twenty-seven percent) over-claimed in bankruptcy and that defenses to the mortgages or proofs of claim existed in fifty-four percent of the cases).

remind the plaintiff of the necessity to continue the filing that perfects the security interest.¹⁰⁶

These kinds of systems can be, and sometimes are, built into software used by one of the parties or the court. Running them in the court files would add three dimensions to their capability. First, the software could use data generated by opponents and courts as well as data generated by the user to formulate its suggestions. Second, the software could use data from other cases within the same court system. Third, the court could operate the system, making it possible to code rather than promulgate court procedures and policies. For example, if the rule required particular recitals in a motion, the automated system could refuse to accept a motion that did not contain them. By coding court rules, the courts can make it literally impossible to violate them.

Transparency can also reduce other kinds of errors by enabling clients to identify and hire the most effective lawyers. The publication of comparative-provider statistics in the field of medicine established that those who did procedures most frequently were best at doing them.¹⁰⁷ The same is true in law.¹⁰⁸ In a transparent court system, prospective clients could hire from among the lawyers most active or most successful in the particular kind of case.

The press and public already evaluate judges on the basis of the percentage of their decisions that are reversed on appeal.¹⁰⁹ In a transparent court system, evaluators would have more information with which to work.¹¹⁰ They could also evaluate on the basis of reversals in unreported opinions or on the basis of the frequency of appeals from a judge's decision.

106. See U.C.C. § 9-515 cmt. 4 (2005) ("Subsection (c) . . . imposes a new burden on the secured party: to be sure that a financing statement does not lapse during the debtor's bankruptcy.").

107. MICHAEL A. MILLENSON, DEMANDING MEDICAL EXCELLENCE 192 (1997) Millenson states:

Among heart surgeons performing at least one hundred procedures a year, the risk-adjusted death rate ranged from zero to 11 percent. For those doing fewer than one hundred operations, by contrast, the death rates of individual surgeons ran as high as 82 percent for a doctor doing only nine cases.

Id.

108. Harris et al., *supra* note 3, at 250 (finding that malpractice attorneys with more experience get better results).

109. Bronson D. Bills, *A Penny for the Court's Thoughts? The High Price of Judicial Elections*, 3 NW. J.L. & SOC. POL'Y 29, 50, 53, 55 (2008) (criticizing particular judges as "the most reversed"); Nina Bernstein, *Immigration Judges Facing Yearly Performance Reviews*, N.Y. TIMES, Aug. 10, 2006, at A14 (reporting on the Attorney General's announcement of "annual performance evaluations" for immigration judges, with particular attention to judges with high reversal rates, backlogs, or complaints).

110. Courts write opinions in only a small proportion of the cases they decide. See *supra* note 69 and accompanying text.

In many cases, particularly small claims and consumer bankruptcies, parties proceed without lawyers. These pro se litigants face the same problems that lawyers face, but without the knowledge and resources that the lawyers have. They are in even greater need of the guidance that a partially automated court system could provide.

In theory, courts could provide these kinds of services without becoming transparent. But to provide the services, the courts would have to adopt transparency's prerequisite infrastructure. That is, the courts would have to cease charging user fees and require parties to file machine-readable documents containing information in relational formats. Almost inevitably, transparency would be a byproduct of those changes.

G. AUTOMATION OF DOCUMENT SERVICE AND FILE MAINTENANCE

Eliminating the restrictions on access to court-file documents would also improve court-system efficiency in more mundane ways. Parties and their attorneys need constantly to be aware of what is occurring in litigation. Today, they achieve that awareness through a complex process in which lawyers "serve" filed documents on other parties to the litigation and all parties forward copies to their clients. Because a transparent court system would not charge user fees, it could automatically and immediately serve all filed documents by email to anyone who had indicated a desire to receive them. Each recipient could control what documents it elected to receive and the email address at which it elected to receive them. Neither the lawyers nor the clients would need to maintain their own copies of the courts' files because all could access the courts' copies as easily as they could access their own. Software could eliminate document boundaries to provide users with whole-case views.¹¹¹

IV. OBJECTIONS TO TRANSPARENCY

From a systems standpoint, transparency has virtually no drawbacks. Transparency would reveal so-called "personal" information about individuals. Personal information that is in court files, however, is already public record. In cases where the harm from court-file revelation of personal information might outweigh the benefits, the law already authorizes the courts to make the information secret. The courts accomplish that by "sealing" all or part of a record for cause.¹¹²

111. Silverman, *supra* note 14, at 198 ("Imagine, for example, being able to display simultaneously the conflicting factual claims contained in a plaintiff's complaint and a defendant's answer, or an argument and its critique culled from one side's memorandum in support of a motion and the other side's memorandum in opposition.").

112. *E.g.*, 11 U.S.C. § 107 (2000) (authorizing the courts to seal records in bankruptcy cases); *Kamakana v. City & County of Honolulu*, 447 F.3d 1172, 1180 (9th Cir. 2006) ("Those who seek to maintain the secrecy of documents attached to dispositive motions must meet the high threshold of showing that 'compelling reasons' support secrecy. A 'good cause' showing

Transparency would alter the balance of power in the courts. The courts are already transparent to those with the financial ability to access the data. For example, the insurance industry compiles, and shares internally, detailed information on the outcomes of claims against insurance companies.¹¹³ That information gives insurance companies an advantage over claimants in evaluating claims for settlement. The insurance industry selectively provides that information to researchers sympathetic to the industry's interests,¹¹⁴ thus biasing research in the industry's favor. By making information available to both sides in litigation, transparency would tend to level the playing field.

Those with greater financial ability might tend to reap greater advantage from transparency, because they could afford more sophisticated analyses. This tendency would be moderated, however, by the existence of more than 10,000 law professors—who are already paid to conduct and publish research and constitutionally inclined to side with the have-nots. Most do not perform empirical research today. But that is already changing, and if the data were available, it would change more rapidly.

Transparency's opponents can be expected to raise at least five kinds of arguments in opposition: (1) transparency would be costly; (2) it would harm litigants by exposing them to embarrassment, identity theft, or the aggregation of personal information; (3) it would invade their privacy by rendering "practically obscure" information discoverable; (4) it would pressure judges to decide cases badly; and (5) it would facilitate the copying of lawyers' work product. None of those objections is well taken. Each is considered here separately.

A. COST

Despite the efficiency of automated research, the aggregate cost of court-system transparency would be substantial. Nearly all of the cost, however, would be for research design time. The research designers would be volunteers, so the cost to the government would be negligible.

The government need only do two things. First, the government must require judges and parties to record as much court information as is practical in relational formats. The added expense would be negligible. Courts and lawyers have already discovered that the use of forms and templates saves effort and money. In the bankruptcy courts, for example, dozens of documents—including some court orders—are filed on standardized forms. Lawyers in every field use word-processing templates for

under Rule 26(c) will suffice to keep sealed records attached to nondispositive motions." (quoting Phillips *ex rel.* Estates of Byrd v. Gen. Motors Corp., 307 F.3d 1206, 1135–36 (2002)).

113. Tom Baker, *Transparency Through Insurance: Mandates Dominate Discretion*, in *TRANSPARENCY IN THE CIVIL JUSTICE SYSTEM* (Joseph Doherty & Robert T. Reville eds., forthcoming 2008).

114. *Id.*

the documents they file, changing only the portions of each document that are case specific. Those forms and templates could just as easily be data-enabled forms.

Second, for courts to be transparent, the government must release court information unconditionally, without charge. Recall that researchers must be able to automate the entire process of data acquisition, analysis, and display. Government restrictions on further dissemination of court data, such as those that Professor Daniel Solove would impose,¹¹⁵ would force researchers to either (1) disseminate their data one user at a time so that they could enforce the restrictions, or (2) refuse to disseminate their data at all. Many researchers could not afford to do the former. Even if they could, dissemination would be more difficult and fewer users would have access. Some researchers would refuse to disseminate their data at all, making their studies “black boxes” that other researchers could neither replicate nor verify. Others would decline to study court data because they could not disseminate their data in a manner that satisfied the data-dissemination standards of their discipline.¹¹⁶

If the government imposed even a small charge for data, the effect would be to reduce sharply the number of studies conducted and to change their nature from scientific to commercial. Consider, for example, PACER’s charge of eight cents a page for document downloads. Even a simple study that downloaded and compared data from two or three documents in each of one hundred cases would cost in excess of \$100. That amount is trivial for a law firm billing its clients at \$500 per hour of attorney time, but current PACER charges severely limit students conducting empirical research in law-school seminars. Any per-page charge, however small, for court data would prevent students—and many faculty members—from automatically updating their studies.

Granting fee exemptions to academic researchers would not solve the cost problem. The courts already grant such exemptions. One problem is that the courts may grant, deny, or condition them in ways that encourage researchers to portray the courts in a positive light.¹¹⁷ Another is that each bankruptcy or district court grants exemptions for only its own records. A researcher can conduct exempt nationwide research only by obtaining an exemption from each of the ninety-eight federal districts. Even if the application process were consolidated, the system would still have to distinguish and restrict exempt researchers. The minimum necessary restriction would be that the exempt researcher could not transfer data to

115. Professor Daniel Solove has proposed that government place restrictions on retransfer of court data as a condition of the data’s release. See *infra* text accompanying note 171–75.

116. Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 46–47 (2002) (describing political-science standards requiring data dissemination to facilitate replication).

117. For example, after I released research that was critical of the New York bankruptcy court, that court denied my request for an exemption.

nonexempt persons. The adverse consequences have already been discussed.¹¹⁸

B. PERSONAL IDENTIFICATION AND PRIVACY

Law is mainly about people. So are public court records. Those records reveal who committed crimes, filed lawsuits, lost lawsuits, perjured themselves, divorced, filed bankruptcy, or cheated their partners. Such information is tremendously valuable. Public court records are principal contributors to credit reports, background and employment investigations, and criminal-records checks. I will refer to the information that a searcher could obtain about a particular individual from a transparent court system as “reputation data.” The reputation-data benefits of transparency would come in addition to the statistical-analysis benefits.

Alan Westin defined “informational privacy” as “the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others.”¹¹⁹ That claim conflicts directly with the goal of court-system transparency.

By making court records public, the courts long ago rejected the privacy claim.¹²⁰ The rejection, however, was neither complete nor final. The courts have long recognized the need to treat some court records as private by “sealing” them. As part of the process of computerization and electronic release of court records, the courts must now grapple with the possible need for additional privacy restrictions.

Commissions and committees (hereinafter “public-access committees”) appointed by the state and federal governments are doing the grappling. Some have already reported.¹²¹ Others are still deliberating. In most

118. See *supra* text accompanying notes 116–17 (discussing the adverse consequences of restrictions on data dissemination).

119. ALAN F. WESTIN, *PRIVACY AND FREEDOM* 7 (1967).

120. See, e.g., *Doe v. Heitler*, 26 P.3d 539, 544 (Colo. App. 2001). In *Heitler*, the court noted:

A claim that a court file contains extremely personal, private, and confidential matters is generally insufficient to constitute a privacy interest warranting the sealing of the file. Likewise, prospective injury to reputation, an inherent risk in almost every civil lawsuit, is generally insufficient to overcome the strong presumption in favor of public access to court records.

Id.

121. See, e.g., COMM’N ON PUB. ACCESS TO COURT RECORDS, REPORT TO THE CHIEF JUDGE OF THE STATE OF NEW YORK (2004), available at http://www.nycourts.gov/ip/publicaccess/report_publicaccess_courtrecords.pdf (recommending that court records already deemed public should not be subject to additional restrictions when posted to the Internet); SUPREME COURT OF FLA., COMM. ON PRIVACY & COURT RECORDS, PRIVACY, ACCESS, AND COURT RECORDS (2005) [hereinafter COMM. ON PRIVACY & COURT RECORDS], available at http://www.flcourts.org/gen_public/stratplan/privacy.shtml (making a general policy statement in favor of electronic access).

jurisdictions, electronic release of court records is in its early stages and the governing privacy regulations are not yet settled.

A reasonably clear pattern already exists, however, in the recommendations. The predominant view is that court files should be available online to the same extent that they are available at the courthouse.¹²² As part of a “go slow” implementation policy, the committees recommend that courts begin online release, observe the effects, and then decide how and how far to continue.¹²³ Some contemplate permanently omitting the most sensitive kinds of cases from online release. Those kinds include domestic violence, child abuse, juvenile, social security, and criminal cases.¹²⁴

Privacy advocates generally seek to block the release of court records that contain “personal information.” “Personal information” has a variety of meanings. Under the narrowest definition, it refers to the human identifiers most commonly used by credit-reporting agencies as passwords to identify people: Social Security number, birth date, mother’s maiden name, and the like. Under the broadest definition, it includes information regarding an identifiable person, whether or not the information is sensitive. In Florida, the effect of a decision to block the release of records containing personal information was to almost completely eliminate public access to court records.¹²⁵

The recently adopted privacy policy of the federal courts requires removal or redaction of only four, narrowly defined types of personal information from case files prior to release. They are (1) Social Security numbers, except for the last four digits; (2) financial-account numbers, except for the last four digits; (3) birth dates, except for the year; and (4) the names of minors, except for the initials.¹²⁶

122. *E.g.*, CAL. R. CT. 2.503(b) (2007) (“A court that maintains the following records in electronic form must provide electronic access to them, both remotely and at the courthouse, to the extent it is feasible to do so: . . . (2) All records in civil cases, except those listed in (c) (1)–(6).”).

123. SUPREME COURT ADVISORY COMM. ON RULES OF PUB. ACCESS TO RECORDS OF THE JUDICIAL BRANCH, RECOMMENDATIONS OF THE MINNESOTA SUPREME COURT ADVISORY COMMITTEE ON RULES OF PUBLIC ACCESS TO RECORDS OF THE JUDICIAL BRANCH 47–48 (June 2004) [hereinafter RECOMMENDATIONS], available at <http://www.lawlibrary.state.mn.us/access/accessreport.htm> (“The advisory committee’s recommendations on Internet access should be viewed as the first step in a go-slow approach to providing more remote access to information.”).

124. *E.g.*, CAL. R. CT. 2.503(c) (excepting records in the following proceedings: (1) under the Family Code, (2) juvenile court, (3) guardianship or conservatorship, (4) mental health, (5) criminal, and (6) civil harassment).

125. COMM. ON PRIVACY & COURT RECORDS, *supra* note 121, at 33 (“After lengthy struggle, the Committee has therefore reluctantly reached the conclusion that implementation of a system that allows large volumes of court records to be released electronically cannot be achieved at this time.”).

126. FED. R. CIV. P. 5.2 (containing the redaction rules).

This Part considers the transparency implications of the privacy restrictions that courts adopt. The discussion proceeds in three steps. Section 1 explains the need for personally identifying information for statistical analyses and the extent to which the federal view might impair those analyses. Section 2 examines the difficulty of de-identifying court records in a world where court proceedings will remain public. Section 3 addresses four bases on which privacy advocates are pressing for further restrictions on access and the manner in which each threatens transparency.

1. The Need for Identification in Court Records

Privacy advocates generally seek to force removal of personally identifying information from court records prior to their release. Such removal is referred to as “de-identification.” To the extent that courts de-identify records, the user is unable to match them to, or associate them with, any particular person.

For example, the Health Insurance Portability and Accounting Act (“HIPAA”) privacy regulations require health-care providers to de-identify medical records before releasing them for research. HIPAA’s de-identification safe harbor permits the release of medical records after redaction of these data items: (1) names; (2) postal addresses, but not city, state, and zip code; (3) fax numbers; (4) e-mail addresses; (5) Social Security numbers; (6) medical-record and other account numbers; (7) certificate and license numbers, including automobile-license numbers; (8) device identifiers and serial numbers; (9) URLs; (10) IP addresses; (11) biometric identifiers, including finger and voice prints; and (12) full-face photographic images.¹²⁷ The purpose of these restrictions is to prevent the researchers who work with the de-identified medical records from knowing the patient’s identity.

De-identified court records can be linked to no person and so can provide no reputation data. In theory, researchers can still perform statistical analyses with de-identified data. Those analyses, however, suffer from at least three important limitations.

First, the researcher cannot validate the records by checking them against the phenomena that they supposedly represent.¹²⁸ De-identified records are merely an unverifiable representation by the government that such people exist and such events occurred. If the government simply invented an internally consistent, de-identified data set, the researcher receiving the data set would have no way to know it was a fake. Indeed, one method of de-identifying data is to create a “synthetic” data set that supposedly has the same relevant characteristics as the sets for which they

127. 45 C.F.R. § 164.514(e) (2006).

128. The usual method is to check randomly selected entries in the data against the represented events until the researcher is confident of the quality of the data.

are substituted.¹²⁹ None of the records in a synthetic data set, however, actually represents a real person or event. They too are fake.

Researchers and the press verify court-system data using a variety of methods. If the government compiled the data from court records in whole or in part, the researchers may check the data against those court records. This method has turned up startling levels of inaccuracy in the court data furnished to researchers by the Administrative Office of the U.S. Courts¹³⁰ and the U.S. Sentencing Commission.¹³¹ A second method is for researchers or journalists to identify independently people and events that should appear in the government's data and then determine whether they do. This method led to the recent discovery that both the Connecticut and federal courts illegally concealed thousands of court files from the public.¹³² Had

129. Douglas J. Sylvester & Sharon Lohr, *Counting on Confidentiality: Legal and Statistical Approaches to Federal Privacy Law After the USA Patriot Act*, 2005 WIS. L. REV. 1033, 1119–20 (discussing the shortcomings of synthetic data).

130. The history of misleading data from the Administrative Office is lengthy. In the late 1980s, the Administrative Office furnished William Whitford and me with a purported list of bankrupt companies reporting assets of over \$100 million. We checked the list against other sources and were unable to confirm the existence of many of the cases. Later, Bankruptcy Judge Lisa Fenning discovered the source of the problem: large numbers of “phantom” \$100 million cases emanating from a poorly drafted question on the bankruptcy cover sheet. Lisa Hill Fenning & Craig A. Hart, *Measuring Chapter 11: The Real World of 500 Cases*, 4 AM. BANKR. INST. L. REV. 119, 123 (1996) (reporting that in forty-one of 262 chapter 11 cases (16%), debtors erroneously reported assets in the \$100,000 category as being in the \$100 million category). At about that same time, Jennifer Frasier found that “[t]he debtor’s assets, liabilities, and number of creditors as shown on the face sheet are inaccurate 20 to 25 percent of the time.” Jennifer Connors Frasier, *Caught in a Cycle of Neglect: The Accuracy of Bankruptcy Statistics*, 101 COM. L.J. 307, 340 (1996). *Id.* at 340–41 (defining “inaccurate” to mean placed in the wrong one of several categories and describing high levels of errors by the Administrative Office of the U.S. Courts in bankruptcy statistical reporting). The overall error rate rises from fourteen percent to twenty percent as the data are captured from the documents. *Id.*

In early 2007, a group of scholars, myself included, discovered substantial differences between the numbers of bankruptcy filings reported in Federal Judicial Caseload Statistics and the numbers of files on PACER. The discrepancies numbered in the thousands of cases. We ultimately discovered that the Administrative Office classifies the reopening of a bankruptcy case as an additional case, because the debtor pays a fee for reopening. Thus, the Administrative Office counted fees and called them cases. Even after adjustment for that difference, the number of files on PACER still did not reconcile with the number of cases reported by the Administrative Office.

131. Cindy R. Alexander, Jennifer Arlen & Mark A. Cohen, *Evaluating Trends in Corporate Sentencing: How Reliable Are the U.S. Sentencing Commission’s Data?*, 13 FED. SENT’G REP. 108, 108 (2000) (“[A] substantial number of cases seem to be missing from the Commission’s data on organizational sanctions. Indeed, the data do not appear to be representative of the underlying case population.”).

132. *Hartford Courant Co. v. Pellegrino*, 380 F.3d 83, 86 (2d Cir. 2004). The court stated:

Between 2002 and 2003, the newspaper plaintiffs learned that, over the prior 38 years, the Connecticut state court system had adjudicated what appeared to be thousands of cases where sealing procedures prohibited court personnel from allowing the public to access the files in those proceedings and, in certain

the publicly disclosed records been de-identified, neither method could have been used. The concealments might never have been discovered. De-identification removes all information linking the released data set to the underlying reality.

Second, de-identification hampers research by preventing researchers from linking court records with other records regarding the same person.¹³³ Researchers may need to link court records to (1) records of jails and prisons regarding incarceration of criminals;¹³⁴ (2) Megan's Law databases; (3) property-tax records, including those on zillow.com; (4) professional-licensing records for lawyers, doctors, accountants, stock brokers, and other professions; (5) arrest records; (6) records of political contributions; (7) fictitious-name registries; (8) birth and death records; (9) news stories; and (10) databases compiled from any combination of the above.¹³⁵ Some private parties might wish to link them to their own databases of employees, customers, or insureds or to the membership lists of professional organizations. The number of possible studies declines geometrically with reductions in the number of databases available for matching. De-identification limits each research project to a single, isolated database.

Third, de-identification hampers transparency by preventing researchers from matching identities within the data set. Thus researchers would not know whether the filers of two bankruptcy cases were the same

comparatively rare instances, from acknowledging the existence of these cases altogether.

Id. After media organizations reported the disappearance of hundreds of federal criminal cases, see, for example, Press Release, Reporters Comm. for Freedom of the Press, Hidden Docketing System Keeps Hundreds of Cases Shrouded in Nation's Capital (Mar. 4, 2006), available at <http://www.rcfp.org/newsitems/index.php?i=6637>, the Judicial Conference "strongly urged" courts to report "case under seal" not "case does not exist." REPORT OF THE PROCEEDINGS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 12 (Mar. 13, 2007), available at <http://www.uscourts.gov/judconf/07MarchProceedings.pdf>. Illegal secret dockets have also been a problem in other state-court systems. Bruce Rishton, *The Curious Case of 2007MR219: Search Warrant Appears, Disappears, Reappears on Court Docket*, ST. J.-REG. (Springfield, Ill.), May 8, 2007, at 1 (reporting the existence of sealed cases with no public dockets in Illinois).

133. Matching de-identified records to identified records would defeat the purpose of de-identification by giving the researcher the identities of the de-identified subjects. Matching de-identified records to other de-identified records does not reveal the identities of either set of subjects. It requires, however, a common identifier by which to make the match. Unless a single person de-identified the two sets of records and created the common identifier for the purpose of linking them, such a common identifier would not exist.

134. For example, identified data are available on prisoners held by the Florida Department of Corrections. Fla. Dep't of Corr., Offender Information Search, <http://www.dc.state.fl.us/AppCommon/searchall.asp?Action=Find&SexOffOnly=0> (last visited Sept. 23, 2008).

135. *E.g.*, John Golmant & Tom Ulrich, *Bankruptcy Repeat Filings*, 14 AM. BANKR. INST. L. REV. 169, 174-75 (2006) (reporting a study in which the researchers obtained Social Security numbers from public court records); John Golmant & Tom Ulrich, *Aging and Bankruptcy*, AM. BANKR. J., May 2007, at 27 (reporting a later study in which the researcher matched the Social Security numbers to "public records accessible through Lexis and Westlaw" to obtain birth dates and reports on aging of bankruptcy filers).

person, whether the defendants in two medical-malpractice actions were the same doctor, whether the persons convicted in two criminal cases were the same person, or whether a group of identical lawsuits were multiple filings by the same person or separate filings by different persons.¹³⁶ The government agency that de-identifies the records could make these links for the researcher. But then the researcher could have no way of knowing precisely how, or how well, the agency made the matches.

2. The Difficulty of De-Identifying Court Records

American courts are open to the public. The public and the press have a constitutional right to attend criminal and civil trials and hearings.¹³⁷ No one proposes to eliminate that right. Nor is there any way that the participants in those trials and hearings can be de-identified. To the extent that they are relevant, witnesses will continue to speak the names, addresses, and other identifying information regarding the parties—even when the parties are minor children. The press will continue to report them, and information brokers and busybodies will continue to collect them.

Thus, court-record de-identification can do no more than eliminate one channel—the court file—through which that information might flow. De-identification could reduce the likelihood that any given person could discover any given fact. It could not, however, make what someone said publicly private again. De-identification can protect private information, such as that given in confidence to the census bureau or to a medical provider. But it cannot protect information spoken in a public courtroom.

When data are de-identified, steps are usually taken to prevent re-identification. One such step is to remove non-identifying information that might be combined with other information to re-identify the subject. That is easier to do with medical records than with court records. Medical records report symptoms and bodily conditions likely to be known only to the patient and the physician who made the record. Researchers cannot use their knowledge of particular patients' medical conditions to reattach identities because researchers do not know the medical conditions of significant numbers of patients.

By contrast, court records report events likely to have occurred in public and even likely to have been discussed in newspaper articles. Researchers using de-identified court records could match the stories in

136. The last reference is to a ploy by which some lawyers pick their judges. Most courts assign judges randomly. To beat that system, the lawyer files several copies of a case until the lawyer draws a desired judge. The lawyer then dismisses all of the cases except the lucky one. Several versions of this ploy exist. *E.g.*, *No Judge-Shopping Allowed*, NAT'L L.J., May 5, 1997, at A8 (reporting that an attorney paid sanctions of \$7,500 for filing thirteen lawsuits then withdrawing all but one in a case involving Dr. Jack Kevorkian).

137. *Richmond Newspapers, Inc. v. Virginia*, 448 U.S. 555, 580 (1980) (“[T]he right to attend criminal trials is implicit in the guarantees of the First Amendment . . .”).

those records to the stories in newspaper articles—or gossip—to reattach identities. Stripping all details that would enable researchers to reattach identities would protect “privacy,” but only by stripping away the non-identifying information needed for research.

3. The Privacy Objections

Privacy advocates seek at least four kinds of limitations on court-system transparency. First, they seek to protect subjects from specific kinds of harms that might result from making identified court data public. Second, they seek to preserve the “practical obscurity” of court data even in situations where they can identify no harm that would flow from its release. Third, they seek to de-identify court data so that the data cannot be matched or aggregated by human identities. Finally, they seek to place contractual restrictions on released court data to prevent commercial use or resale.

a. Identity Theft and Other Specific Harms

Identity theft results from the business practice of treating a person’s knowledge of a customer’s Social Security number, credit card number, or birth date (hereinafter “identity-theft data”) as proof that the person is the customer.¹³⁸ So long as businesses continue that practice, revealing a person’s identity-theft data would expose that person to the possibility of identity theft.

To my knowledge, no evidence exists that court records have been a significant source of information used in identity theft. Nevertheless, public-access committees have generally recommended that identity-theft data be omitted from court records or redacted from court records prior to the records’ release. If the court records are not machine readable, finding and removing identity-theft data is impractical. Under the committees’ recommendations, the records may never qualify for release—even if they are in electronic, but not machine-readable, form. If the records are machine readable, court officials can find and remove identity-theft data with virtually one hundred percent accuracy. Thus, requiring the filing of court documents in machine-readable formats is likely to reduce, rather than increase, the risk of identity theft.

Privacy advocates identify a number of other specific harms that can result from the inclusion of personally identifying information in public records. They include stalking, domestic violence, and witness intimidation or killing.¹³⁹ Authority already exists for courts to seal court records when

138. Lynn M. LoPucki, *Human Identification Theory and the Identity Theft Problem*, 80 TEX. L. REV. 89, 99–100 (2001).

139. *E.g.*, RECOMMENDATIONS, *supra* note 123, at 8–12, 9 n.10 (discussing witness intimidation).

they create a significant risk of such problems.¹⁴⁰ Courts, however, seal only a small proportion of records, and no evidence exists showing that the current level of record sealing is inadequate to address the problems.

Sealing court records inhibits research by creating two kinds of problems for researchers. The obvious one is to reduce the number of records remaining available for research. For example, if the courts routinely seal the bankruptcy records of domestic-abuse victims, insufficient numbers of such records might remain for research on domestic-abuse victims' bankruptcies.

The less obvious problem is to call into question the representativeness of the unsealed records. To continue with the same example, if a matrimonial court were to seal twenty percent of its files, researchers would know nothing about the contents of those files. All might show abuse, or none might show abuse. If examination of the court's remaining files were to reveal that two percent show domestic abuse, the researcher could conclude only that domestic abuse is present in two percent to twenty-two percent of cases. That range is so broad that such a study would have little value. Current levels of sealing are, however, low enough not to interfere with most kinds of court-record research.

b. Practical Obscurity

Practical obscurity is perhaps the most commonly raised objection to the electronic release of public court records.¹⁴¹ This objection requires no threat of harm to anyone from record release.¹⁴² Rather, the objection

140. See authorities cited *supra* note 112.

141. Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1176 (2002) (arguing that the public-private distinction is outmoded and should be replaced by a new paradigm in which individuals have privacy rights in information on the public record). Solove argued:

I contend that information privacy must be reconceptualized in the context of public records to abandon the longstanding notion that there is no claim to privacy when information appears in a public record I suggest that privacy must be understood as an expectation of a limit on the degree of accessibility of information.

Id. at 1140; Peter A. Winn, *Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information*, 79 WASH. L. REV. 307, 325 (2004) ("Courts . . . recreate in cyberspace a system that protects the same value that [was] protected by the practical obscurity of paper [court] records.").

142. For example, Solove concludes:

Privacy involves an expectation of a certain degree of accessibility of information. Under this alternative view, privacy entails control over and limitations on certain uses of information, even if the information is not concealed. Privacy can be violated by altering levels of accessibility, by taking obscure facts and making them widely accessible.

Solove, *supra* note 141, at 1178.

asserts that public court records should be private because they were difficult to access in the past, and people have come to expect difficulty accessing them.¹⁴³

The objection is based on those scholars' misreading of the Supreme Court's opinion in *United States Department of Justice v. Reporters Committee for Freedom of the Press*.¹⁴⁴ In that case, the Court held that the government could not disclose the contents of an FBI rap sheet under the Freedom of Information Act ("FOIA") because the subject of the rap sheet had a privacy interest in the rap sheet, and its disclosure would constitute an "'unwarranted' invasion of privacy."¹⁴⁵

The Court did not hold, however, that the subject had a privacy interest in the court records from which the compilation was made. To the contrary, the Court said that "[p]lainly there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information."¹⁴⁶ The Court recognized that "the power of compilations to affect personal privacy that outstrips the combined power of the bits of information contained within."¹⁴⁷ The Court recognized a privacy right in the FBI's compilation. Nothing in the Court's opinion, however, suggests that the bits of information in the courts' records should no longer be public or that members of the public should be barred from compiling their own rap sheets from them.

To the extent that *Reporters Committee* relies on the practical obscurity of court records, it was a decision grounded in the technology of the time. As the Court put it:

The very fact that federal funds have been spent to prepare, index, and maintain these criminal-history files demonstrates that the individual items of information in the summaries would not otherwise be "freely available" either to the officials who have access to the underlying files or to the general public. Indeed, if the summaries were "freely available," there would be no reason to invoke the FOIA to obtain access to the information they contain.¹⁴⁸

143. *Id.* at 1141 ("I suggest that privacy must be understood as an expectation of a limit on the degree of accessibility of information."); *id.* at 1178 ("Our expectation of limits on the degree of accessibility emerges from the fact that information in public records has remained relatively inaccessible for much of our history."); Winn, *supra* note 141, at 325.

144. *U.S. Dep't of Justice v. Reporters Comm. for Freedom of the Press*, 489 U.S. 749 (1989).

145. *Id.* at 776 (quoting 5 U.S.C. § 552(b)(7)(C) (2000)).

146. *Id.* at 764.

147. *Id.* at 765.

148. *Id.* at 764.

The technology has since changed. Specifically, the Internet developed and both the state and federal courts computerized their case records. Private sources now compile rap sheets and make them available.¹⁴⁹ A job applicant can no longer assume that a prospective employer does not have access to the applicant's criminal record.¹⁵⁰ Thus, criminal records are no longer practically obscure. Efforts to *make* court records practically obscure by erecting artificial barriers to obtaining them are not mandated, and arguably would violate the Americans with Disabilities Act.¹⁵¹

Consistent with this view, public-access committees generally recommend that court records be available online to the same extent that they are available at the courthouse. For example, California court rules require generally that courts keeping electronic records must provide the public access electronically, both remotely and at the courthouse.¹⁵² But those rules also provide an exception for six kinds of records that need only be made available at the courthouse.¹⁵³

Critics of such exceptions have noted the potential for data arbitrage. If data that were not available online were available at the courthouse, service companies would go to the courthouse, get the data, and sell it online.¹⁵⁴ The sale price would reflect the difficulty of the courthouse data transfer. The court might permit flash-drive downloads, thus making data cheap. The court might limit data transfer to the printing of hard copy. The court might display the information on the computer screen and provide no means of printing it. The court might even prohibit users from photographing the screens, effectively requiring them to retype the information. The more onerous the data-transfer requirements, the higher the resale price of the data is likely to be. Such arbitrage would convert privacy into a mere price

149. *E.g.*, SentryLink, National Criminal Background Check & Sex Offender Check, <http://www.sentrylink.com/web/loadCriminalReport.do> (last visited Sept. 23, 2008) (offering a criminal-records search that covers the bulk of criminal courts in the United States).

150. DEREK HINTON, CRIMINAL RECORDS BOOK 79–86 (2002) (describing the complexities of employer acquisition and use of criminal-record information). Complete or partial criminal records from fifteen states are available on LEXIS, Library, DOCKRT, File, and STCRIM.

151. 42 U.S.C. § 12132 (2000) (“No qualified individual with a disability shall, by reason of such a disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of a public entity, or be subjected to discrimination by any such entity.”); *In re Estate of Engelhardt*, 804 N.E.2d 1052, 1058 (Ohio Prob. Ct. 2004) (“Removing case files from the Internet may implicate the ADA because such removal may preclude access to public records for those individuals whose disabilities prevent them from traveling to the court.”).

152. CAL. R. CT. 2.503(b).

153. *Id.* 2.503(c); *see supra* note 124 (listing the exceptions).

154. *See* JUDICIAL CONFERENCE PRIVACY REPORT, *supra* note 8, at A-7 (referring to “data resellers who, if remote electronic access were restricted, could go to the courthouse, copy the files, download the information to a private website, and charge for access to that website, thus profiting from the sale of public information and undermining restrictions intended to protect privacy”).

differential. Whether the government could eliminate arbitrage by prohibiting resale of the data is considered below.

The need for such arbitrage would cripple court-system transparency. Even small per-datum transfer costs become prohibitively large when multiplied by the quantities of data necessary to achieve court-system transparency.

c. Personal Data Aggregation

One of the principal fears of privacy advocates is that, unless court records are de-identified, users will match them to other records about the same person and aggregate the information.¹⁵⁵ (Privacy advocates provocatively describe this as the compilation of “dossiers”—an apparent attempt to link data aggregation in the U.S. to records kept on individuals by the secret police in totalitarian regimes.) As previously noted, such matching and aggregation are necessary to achieve court-system transparency.¹⁵⁶ To the extent that privacy rules prevent aggregation, they impair the courts’ ability to provide reputation data as well as researchers’ ability to analyze court operations.

Public-access committees universally recommend redacting records in ways calculated to prevent identity theft.¹⁵⁷ They split, however, on whether to leave sufficient information on the public record to support data aggregation.¹⁵⁸

The relatively pro-transparency federal redaction rules require redaction of Social Security numbers and financial-account numbers to the last four digits,¹⁵⁹ redaction of birth dates to the birth year,¹⁶⁰ redaction of the names of minors to initials, and in criminal cases, redaction of addresses to city and state.¹⁶¹ The federal rules require no other redactions. The effect is to leave enough identification information to support aggregation. Nearly

155. See DANIEL J. SOLOVE, *THE DIGITAL PERSON* 44–55 (2004) (describing harms that supposedly flow from “the aggregation effect”); Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 506–11 (2006) (identifying the aggregation of data about individuals as one of the principal privacy problems).

156. See *supra* Part IV.B.1 (discussing the need for matching and aggregation).

157. See Lynn M. LoPucki & Shahab Asghar, *Summary of State Reports on Public Access to Court Records passim* (2007) (unpublished manuscript, on file with the author) (showing recommendations that Social Security numbers and financial-account numbers be partially or completely redacted).

158. See *id.* (showing diversity in recommendations regarding redaction of birth dates and addresses and regarding the extent of redaction of Social Security numbers).

159. FED. R. BANKR. P. 9037 (proposed), available at <http://www.supremecourtus.gov/orders/courtorders/frbk07p.pdf> (last visited Sept. 23, 2008).

160. *Id.*

161. FED. R. CRIM. P. 49.1(a).

half the states have rules adopting, or reports recommending that they adopt, the federal redaction scheme.¹⁶²

On the other hand, slightly over half the states have adopted or recommended pro-privacy redaction rules designed to block aggregation.¹⁶³ That view would redact the entire Social Security number and the entire date of birth, but leave the individual's name on the public record. Preservation of names would enable the public record to continue to provide some reputation data, though with considerable risk that the reputations would attach to the wrong people. Complete redaction of Social Security numbers and birth dates, however, would make both aggregation and statistical analyses based on aggregation impossible.

The split regarding redaction rules is about aggregation, not privacy in the traditional sense. In the traditional sense, leaving the individual's name and story on the public record invades privacy. Private detectives and busy-body neighbors interested in a specific individual usually would have no trouble identifying that individual from the court file.¹⁶⁴ Names and stories, however, are not sufficient for research. Research requires matching human identities in national data sets.¹⁶⁵

The pro-privacy redaction rules provide no additional protection against identity theft. Some creditors and credit-reporting agencies treat knowledge of a person's Social Security number and birth dates as evidence that the knower is the person.¹⁶⁶ Both sides in the privacy–transparency debate condemn that practice and recommend legislation to end it.¹⁶⁷ The practice, however, continues. As a result, one who knows another's entire Social Security number and birth date might be able to impersonate the other. Four digits of the Social Security number and the year of birth, however, do not create that risk.

A person's name, together with the last four digits of his or her Social Security number and date of birth, are sufficient for matching in national databases. The only apparent reason for the federal policy that preserves the last four digits of Social Security numbers is to make such matching possible and thus facilitate aggregation. Thus, it appears that the federal rules are deliberately pro-aggregation.

162. LoPucki & Asghar, *supra* note 157 (showing that a substantial minority of states recommend not redacting the last four digits of Social Security numbers).

163. *Id.*

164. Jennifer 8. Lee, *Dirty Laundry, Online for All to See*, N.Y. TIMES, Sept. 5, 2002, at G1 (reporting complaints about neighbors snooping when court records were placed online).

165. Many people have the same names, and the form of a person's name often varies through the use of nicknames, initials, aliases, and suffixes.

166. See Lynn M. LoPucki, *Did Privacy Cause Identity Theft?*, 54 HASTINGS L.J. 1277, 1281–82 (2003) (describing the process).

167. E.g., Daniel J. Solove, *Identity Theft, Privacy, and the Architecture of Vulnerability*, 54 HASTINGS L.J. 1227, 1270 (2003) (“An SSN, mother's maiden name, and birth date should be prohibited as the method by which access can be obtained to accounts.”).

That facilitation of aggregation is imperfect. While no two people have the same name and Social Security number, about one person in twenty-three has the same name and last four digits as someone else.¹⁶⁸ As a result, matching large databases by name and last four digits would produce a significant number of false positives. Depending on the specific type of study, those false positives may be a minor irritant to the researcher or may make the research impossible.

Bankruptcy rules require bankruptcy filers to identify themselves by full name, aliases, and Social Security numbers.¹⁶⁹ The federal redaction rule thus assures that sufficient information remains for matching the identities of bankruptcy filers with identities in other databases. In civil and criminal cases, however, the courts typically do not require parties to identify themselves by Social Security numbers or dates of birth. As a result, the court file may not contain an identity sufficient for automated matching across databases.

Courts adopting the federal view should consider requiring litigants to identify themselves by last four digits of the Social Security number and year of birth. That information would facilitate court-system accountability and make reputation data more accurate without invading any protected privacy interest of the litigants.

168. Ten thousand combinations of four digits exist. Thus the odds that two people will have the same last four digits are about one in 10,000. About 420 million Social Security numbers have been issued since the program began. Soc. Sec. Admin., *New Social Security Numbers*, <http://www.socialsecurity.gov/history/ssn/ssnvolume.html> (last visited Sept. 23, 2008). If 300 million of those numbers are currently outstanding, about 30,000 people have any particular combination of last four digits (300 million / 10,000). I estimate that in a population of 19,000 people, about 858—roughly one in twenty-three—have the same name. That estimate is based on the examination of the names on three pages of the UCLA Telephone Directory 2002–2003. That directory contains about 19,000 names on 198 pages. On the three pages examined, I found eleven duplicate names and one triplicate name. The standard for determining names to be duplicates was that the last name be identical, the first name be identical or a corresponding nickname (e.g., Cathy and Catherine), and that the middle names or initial not be inconsistent (e.g., John Williams is considered a duplicate of John K. Williams). Projecting that total to the entire directory yields an estimate of 858 duplicates among 19,000 people (one in twenty-three). Because the number of possible names in a population is finite, adding more names from that same population should result in a higher proportion of duplicates. Thus the ratio of duplicates to population for a population of 30,000 would be higher than the ratio for a population of 19,000.

169. FED. R. BANKR. P. 9009, Official Form 1 (requiring “name of debtor,” “all other names used in the last 8 years,” and “last four digits of Soc. Sec. . . . No.”); FED. R. BANKR. P. 1007(f) (“An individual debtor shall submit a verified statement that sets out the debtor’s social security number, or states that the debtor does not have a social security number.”).

d. Contract Restrictions to Prevent Commercial Use

Current law permits courts to deny access to court records on the basis of the requestor's intended use.¹⁷⁰ Once the requestor receives access, however, the requestor can use the records for any lawful purpose. Professor Daniel Solove argues that government can, and should, restrict use further. Specifically, he argues that the government can make a public record available on the condition that certain information is not disclosed or used in a certain manner,¹⁷¹ but cannot establish post-access restrictions on the disclosure or use of information that is publicly available.¹⁷² Once the information is made available to the public, the *Florida Star v. B.J.F.* case prohibits a state from restricting use "by making access conditional on accepting certain responsibilities when using data—such as using it for specific purposes, not disclosing it to others, and so on, certain functions of transparency can be preserved at the same time privacy is protected."¹⁷³ What Solove apparently has in mind is to allow the kind of research proposed in this Article,¹⁷⁴ while disallowing the commercial use of court-system data.¹⁷⁵

The Administrative Office of the U.S. Courts uses this approach with respect to an abbreviated case-record database it has released annually since 1970.¹⁷⁶ But the use restrictions the Administrative Office imposes are calculated to prevent research rather than to facilitate it. Specifically, the government strips the names of the judges and the parties and conditions the download of the information on the execution of a clickwrap agreement that prohibits reattaching the names.¹⁷⁷ Such restrictions would be fatal to court-system transparency.

Restrictions that would prohibit the use of court data—however obtained—to prepare marketing lists would not significantly hinder court-system transparency. But Solove's proposal to bar researchers from disclosing data to others would. The scientific method relies on the ability of

170. *Nixon v. Warner Commc'ns, Inc.*, 435 U.S. 589, 598 (1978) ("It is uncontested, however, that the right to inspect and copy judicial records is not absolute. Every court has supervisory power over its own records and files, and access has been denied where court files might have become a vehicle for improper purposes.").

171. Solove, *supra* note 141, at 1213.

172. *Id.*

173. *Id.* at 1213–14 (discussing *Fla. Star v. B.J.F.*, 491 U.S. 524 (1989)).

174. *Id.*

175. *Id.* at 1216 ("[A] more appropriate approach is to curtail broad categories of uses (i.e., commercial, information brokering, further disclosure, and so on) . . .").

176. ICPSR Inter-Univ. Consortium for Political & Soc. Research, Federal Court Cases: Integrated Data Base, 2006 (2007), available at <http://www.icpsr.umich.edu/cocoon/ICPSR/STUDY/04685.xml>.

177. Lynn M. LoPucki, *The Politics of Access to Federal Court Data*, 80 TEX. L. REV. 2161, 2169–70 (2002) (describing the arrangement).

other researchers to reproduce research results.¹⁷⁸ If the second researcher proposes to reproduce the research from scratch, the second researcher would obtain new data from the original source. Restrictions on the first researcher's data are then irrelevant. Many data sets, however, are too expensive to reproduce. Researchers today address the problem by making their data sets publicly available so that later researchers can check their work.¹⁷⁹ In addition, organizations such as the Inter-University Consortium for Political and Social Research acquire and preserve data sets for use by other researchers.¹⁸⁰ My research results are credible, because, if they were wrong, anyone could use my posted data to prove them wrong. Because my data sometimes include identified information, businesses could download it and use it to make marketing lists. Thus, my practice would violate Solove's restrictions.

Second, the mechanisms of transparency must to some degree be funded by the users of the research or their patrons. To verify the research results, both the users and their patrons will sometimes insist upon seeing the underlying data.

Third, the same data processing would provide both transparency data and reputation data. The processing costs will be shared by the two kinds of users. Solove's restrictions would bar the resale of reputation data, thus depriving court-system transparency of a principal source of funding.¹⁸¹

Some access restrictions might be tolerable. The courts could prohibit the use of court-file data to compile lists of people for marketing purposes. The courts could still permit the resale of court-file data, but require that in the process of resale, the reseller impose the same restrictions. That is, however, a dangerous path to take. As occurred with credit information, the result may be the de facto denial of access to individuals.¹⁸²

178. *Ars Technica*, Scientists on Science: Reproducibility, <http://arstechnica.com/journals/science.ars/2006/10/25/5744> (last visited Sept. 23, 2008) ("If experimental or descriptive data cannot be reproduced, then they are generally discarded. Reproducibility was mentioned by several [scientists] as a mechanism by which scientific data becomes viewed as less tentative.").

179. For example, my coauthor and I have posted the data from our studies of large, public-company bankruptcies at <http://www.law.ucla.edu/erg/pubs>.

180. Inter-Univ. Consortium for Political & Soc. Research, Mission Statement, <http://www.icpsr.umich.edu/ICPSR/org/mission.html> (last visited Sept. 23, 2008).

181. PACER, PACER On-Line Registration, <https://pacer.psc.uscourts.gov/psco/cgi-bin/regform.pl> (last visited Jan. 9, 2009) (setting forth the acknowledgment of policies and procedures that places no restrictions on data use).

182. Because resellers of credit information have legal responsibility for the ultimate use of the data, credit-reporting agencies are membership organizations. Individuals rarely need credit information frequently enough to warrant the cost of membership. As a result, only businesses have de facto access to credit information.

C. JUDICIAL INDEPENDENCE

A variety of external pressures and threats may make it difficult for judges to decide cases impartially. Legislators may threaten judges with budget cuts, job cuts, or impeachment. The press may incite the public against them. Interest groups may threaten them with political campaigns that could remove them from office. Disgruntled litigants or political extremists may threaten them with violence. Describing “the threat to judicial independence,” Justice Sandra Day O’Connor recently asserted that “the breadth and intensity of rage currently being leveled at the judiciary may be unmatched in American history.”¹⁸³

If the courts operated in secrecy, none of those groups would know what the courts were doing, and the level of rage about court decisions undoubtedly would decline. That solution would fail, however, to impose necessary limits on judicial independence. Scholars from across the political spectrum agree that judicial independence is not an end in itself. It is merely the means of assuring that judges are free to decide cases impartially.¹⁸⁴ As one commentator put it:

Most thoughtful scholars recognize that judicial independence is an instrumental value—a means to achieve other ends. As an instrumental value, judicial independence has limits, defined by the purposes it serves. Disagreement persists as to what those purposes are, but most would accept some variation on the theme that judicial independence enables judges to follow the facts and law without fear or favor, so as to uphold the rule of law, preserve the separation of governmental powers, and promote due process.¹⁸⁵

Thus, judicial independence is not freedom to do as a judge pleases, but rather, freedom to do what a judge should. Some kind of oversight remains necessary.

In a democracy, that oversight must necessarily come from the public.¹⁸⁶ The public acts directly in electing judges¹⁸⁷ and indirectly in

183. Sandra Day O’Connor, Op-Ed., *The Threat to Judicial Independence*, WALL ST. J., Sept. 27, 2006, at A18.

184. *E.g.*, Pamela S. Karlan, *Judicial Independences*, 95 GEO. L.J. 1041, 1059 (2007) (“[J]udicial independence is not an end in itself; rather, it is a means of ensuring freedom and the rule of law.”); Frances Kahn Zemans, *The Accountable Judge: Guardian of Judicial Independence*, 72 S. CAL. L. REV. 625, 632 (1999) (“But judicial independence is only a means to an end; it is the mechanism chosen by the Founders to ensure the rule of law.”).

185. Charles Gardner Geyh, *Rescuing Judicial Accountability from the Realm of Political Rhetoric*, 56 CASE W. RES. L. REV. 911, 915–16 (2006).

186. *E.g.*, Cynthia Gray, *The Line Between Legal Error and Judicial Misconduct: Balancing Judicial Independence and Accountability*, 32 HOFSTRA L. REV. 1245, 1246 (2004) (noting that the code of judicial conduct requires judges to “respect and comply with the law,” to “be faithful to the law,” and to accord to every person “the right to be heard according to the law”).

electing those who appoint, impose discipline on, and promote judges. To enable the public to make those decisions, the court system must require judges to work under some amount of public scrutiny.

Courts institutionalize democracy's claim that it imposes constraints on state power. In open courts, government judges have to account for their own authority by letting others know how and why power is used. Bentham's widely-quoted phrase captures this activity: "Publicity is the very soul of justice [. . .] It keeps the judge himself, while trying, under trial."¹⁸⁸

The need for public scrutiny has long been the basis for the current policy that court records are available to the public. Secret courts are not a realistic option. The choice is between high and low levels of public information about judges' actions.

Thus, the systems to be compared are the current one in which court records are merely public and the proposed one in which court records would be fully transparent. The standard for comparison is how well each of these systems could serve the combined goals of judicial independence and judicial accountability.

In the current system, voters have low levels of information about judicial candidates.¹⁸⁹ That creates a volatile situation in which particular decisions can be considered out of context and so provide the basis for superficial analysis or political attack.¹⁹⁰ Commentators generally recommend that judges protect themselves by making more information about themselves and their decisions available to the public.¹⁹¹ For example, Frances Kahn Zemans argues:

[W]hen it comes to retention in office, it is the individual judge who must face evaluation. Reporting possible suspect behavior becomes magnified because typically it is the only information the

187. "The great majority of judges in the United States must periodically win elections in order to retain their positions." Lawrence Baum, *Judicial Elections and Judicial Independence: The Voter's Perspective*, 64 OHIO ST. L.J. 13, 16 (2003).

188. Judith Resnik & Dennis E. Curtis, *From 'Rites' to 'Rights' of Audience: The Utilities and Contingencies of the Public's Role in Court-Based Processes*, in REPRESENTATIONS OF JUSTICE 195, 229 (Antoine Masson & Kevin O'Connor eds., 2007).

189. Baum, *supra* note 187, at 19 (noting that "contests for judgeships fall firmly in this 'low-information' category").

190. For example, an Iowa state district judge's grant of a divorce to a lesbian couple resulted in a national controversy. Kathleen Burge, *Iowa Judge Causes Stir in Granting Gay Divorce*, BOSTON GLOBE, Dec. 13, 2003, at B1.

191. Zemans, *supra* note 184, at 654 (recommending that to sustain support for judicial independence, judges should "communicat[e] with the public in ways that will enhance their legitimacy and justify their independence"); see also Robert E. Drechsel, *Dealing with Bad News: How Trial Judges Respond to Inaccurate and Critical Publicity*, 13 JUST. SYS. J. 308, 309 (1988) (arguing that if the "judiciary is to be meaningfully accountable and understood, the public must have accurate informative news" from the media).

public has received. There is not a base of knowledge about individual judges or the judiciary as a whole against which the public can evaluate reports of judicial misbehavior. Thus the invisibility of most judges, while providing the appearance of some protection against attack, may in fact have the opposite effect. That is, with no other information about a judge against which to evaluate a story, whatever is printed or broadcast becomes accepted “truth.”¹⁹²

One solution, Zemans proposes, is to “provide a template of information against which a reporter can evaluate judicial behavior.”¹⁹³ When the judge is right and critics are wrong, the effect of making raw court data available is to vindicate the judge and check the power of critics.

Even in a transparent court system, judicial-election voters would not know firsthand the details of a judge’s decisional record. They would still rely on intermediaries to process and evaluate those records.¹⁹⁴ The difference would be in the levels of information available to the intermediaries. Instead of being dependent on anecdote, the intermediaries would have access to full records. They could assess in depth the pattern of all of a judge’s decisions, instead of skimming the surface of a few.

The function of public oversight is not merely to provide assurance that all is well in the court system. In fact, all is not well. Even at current levels of transparency, researchers have been able to document the systematic influence of judicial self interest. Numerous factors that should be irrelevant influence judicial decisions, including the desire of some judges to attract cases,¹⁹⁵ the threat of electoral challenge, and the proximity of the next election.¹⁹⁶ Transparency is needed to prevent judges from acting improperly.

D. COPYING LAWYERS’ WORK PRODUCT

A substantial portion of all lawyers’ work product is in the form of documents filed with the courts. Those documents are entitled to little, if

192. Zemans, *supra* note 184, at 640.

193. *Id.*

194. In many parts of the United States, for example, lawyers recommend for or against the retention of particular judges by publicly voting on them prior to the election. *See, e.g., Lawyers Strongly Support Judge Who Approved Lesbian Divorce*, ADVOCATE.COM, Sept. 23, 2004, http://www.advocate.com/news_detail.asp?id=06629 (noting that the judge who granted an unpopular divorce to two lesbians received a ninety-three percent approval rating in the Iowa State Bar Association’s election-year poll).

195. LOPUCKI, *supra* note 46, at 137–80 (documenting the effect of court competition on substantive decisionmaking).

196. Nancy J. King, *How Different Is Death? Jury Sentencing in Capital and Non-Capital Cases Compared*, 2 OHIO ST. J. CRIM. L. 195, 204–06 (2004) (reviewing the literature regarding the effects that upcoming elections have on judicial modifications of jury sentences).

any, copyright protection.¹⁹⁷ Because the same problems and legal issues recur, lawyers often copy the work of other lawyers. The principal impediment to such copying is lawyers' difficulty in finding and evaluating the most relevant documents. Today, court-document copying is generally confined to the documents produced by other members of a lawyer's firm. That is because the lawyer can find those documents by searching the firm's archives, and because the firm's standard of quality generally assures the particular document's quality.

In a transparent court system, a lawyer could search among all documents filed with the courts. To the extent that the lawyer was familiar with the standards of quality of the firms that produced the documents, those standards would assure the documents' quality. In essence, the court system would become the world's largest legal form book. Lawyers could copy pleadings, motions, memoranda, briefs, and even the contracts that frequently appear in the court files as attachments or evidence. The fact that this "form book" would be free would tend to reduce lawyers' costs.

In addition, lawyers could use filed transcripts of depositions or hearings in similar cases to prepare for depositions or hearings in their own cases. For example, they could see the problems encountered by an earlier lawyer in questioning a particular expert witness or a type of expert witness and perhaps improve on the earlier lawyer's approach. They could see what arguments were made when an earlier lawyer argued a particular issue before a court and see from the court's decision how successful that argument was. Thus, the principal effects of such copying would be to improve the quality of legal services and reduce prices.

The principal problem with copying in other contexts—that it discourages the production of originals—is not present in the legal context. Clients pay for the production of originals, and lawyers are ethically restrained from charging later clients for work already done. Thus, transparency and copying would not significantly reduce the production of originals.

The effect may, in fact, be to increase the production of originals. When a single client cannot afford the entire cost of producing the original, the client's lawyer could search the court system for other parties who need the same documents and might be willing to share the cost. For example, the client who loses on an issue and seeks to appeal could share the appeal's cost with other clients whose cases would be affected by the appeal's outcome.

197. Davida H. Isaacs, *The Highest Form of Flattery? Application of the Fair Use Defense Against Copyright Claims for Unauthorized Appropriation of Litigation Documents*, 71 MO. L. REV. 391, 402–11 (2006) (observing that many complaints and memoranda may lack the requisite originality and creativity necessary for copyright protection though some legal documents may contain portions that deserve copyright protection).

E. FLIGHT TO PRIVATE ADJUDICATION

While litigation is generally public, arbitration is generally private. As one commentator put it:

Most arbitrations are private. Privacy permeates the atmosphere of arbitration and is often perceived by contracting parties as an advantage over public litigation. In addition to the private location of hearings, the results of arbitration are also private; published opinions are rare. The typical arbitration concludes with a terse, non-explanatory written award that is not disclosed to the public. In short, most arbitration results are essentially secret.¹⁹⁸

Under current law, parties have, in general, the right to choose between arbitration and litigation by contract.¹⁹⁹

That competition between litigation and arbitration is not a healthy one from the perspective of society as a whole. Public adjudication produces public benefits; private arbitration does not. Because the parties that choose between the two have no incentive to take the public benefits of public adjudication into account, the result is a more-than-economically-optimal amount of arbitration.

This has led some commentators to speculate that “increased transparency in civil litigation may have wrought an unintended and unwelcome consequence—the diversion of more civil disputes into alternative dispute resolution proceedings like arbitration, where the public is ‘shut out of information almost completely.’”²⁰⁰ If that is true, the increased transparency proposed in this Article would tend to divert even more disputes to arbitration.

Numerous commentators have lamented arbitration’s opacity²⁰¹ and proposed that the law require more transparency.²⁰² Some states have

198. Edward Brunet, *Arbitration and Constitutional Rights*, 71 N.C. L. REV. 81, 84–85 (1992) (citations omitted).

199. The Federal Arbitration Act provides that arbitration agreements are enforceable except “upon such grounds as exist at law or in equity for the revocation of any contract.” 9 U.S.C. § 2 (2000).

200. Laurie Kratky Doré, *Public Courts Versus Private Justice: It’s Time to Let Some Sun Shine in on Alternative Dispute Resolution*, 81 CHI-KENT L. REV. 463, 465 (2006) (quoting Jack B. Weinstein & Catherine Wimberly, *Secrecy in Law and Science*, 23 CARDOZO L. REV. 20 (2001)). Evidence exists to the contrary. Despite the fact that “[p]ublicity is the price a decedent pays for using ‘court regulated devices’ such as wills or testamentary trusts as opposed to a ‘private arrangement’ such as a revocable trust,” the tabloids suffer no shortage of wills of the rich and famous. Foster, *supra* note 102, at 559. Despite the fact that substantially every bankruptcy file is publicly available on PACER, more than 1.7 million Americans filed bankruptcy annually before the recent “reforms.” U.S. Bankr. Courts, Table F-2, http://www.uscourts.gov/bnkrpctstats/bankrupt_f2table_jun2006.xls (last visited Sept. 23, 2008) (reporting on the twelve month period that ended June 30, 2006).

201. Jean R. Sternlight, *Creeping Mandatory Arbitration: Is It Just?*, 57 STAN. L. REV. 1631, 1658 (2005) (“Unfortunately, researchers have found it very difficult to evaluate mandatory

already begun to address the problem.²⁰³ The solution might be for the courts to enforce only arbitration awards obtained in transparent proceedings—on the principle that the government should not enforce the decisions of secret tribunals. That would cause many more disputants to opt for public adjudication.

Some people may choose to compromise their rights rather than suffer the embarrassment of public adjudication. Undoubtedly, many already do. In establishing public courts, the government already has determined that those courts should hear disputes only if they rise to such a level of importance that the litigants are willing to incur costs, go public, and suffer some amount of discomfort.

Some types of disputes are highly embarrassing, but nevertheless require government resolution. Public adjudication of those disputes might not be appropriate. But the law already provides for the sealing of records and the closing of trials and hearings in such cases.²⁰⁴ In the transition from the current system to a transparent system, some increase in the proportion of sealed records and closed hearings and trials might be needed.²⁰⁵ But the existence of a small number of matters requiring privacy should not prevent society from capturing the huge benefits of transparency in other matters.

arbitration, for a number of reasons. First, to a large extent, researchers cannot obtain access to the data they need to perform good studies. . . . [O]ne of the fundamental traits of arbitration is that it is typically private.”); *see also* TOWARDS A SCIENCE OF INTERNATIONAL ARBITRATION: COLLECTED EMPIRICAL RESEARCH 81 (Christopher R. Drahozal & Richard W. Naimark eds., 2005) (“Arbitration proceedings themselves are private and thus difficult to study. The nature of the process adds to that difficulty. Much of what happens (such as with respect to discovery) may not be documented in any central case file.”).

202. *See, e.g.,* Doré, *supra* note 200, at 520 (advocating “[i]ncreased transparency and accessibility to at least some aspects of ADR in at least some cases”); Judith Resnik, *Uncovering, Disclosing, and Discovering How the Public Dimensions of Court-Based Processes Are at Risk*, 81 CHI-KENT. L. REV. 521, 569 (2006) (applauding legislation and proposed legislation that requires arbitrators of some kinds of disputes to provide reasons for their decisions and make them available to the public).

203. *See, e.g.,* CAL. CIV. PROC. CODE § 1281.96 (West 2007) (requiring arbitration providers to publish data on arbitration outcomes).

204. *Richmond Newspapers, Inc. v. Virginia*, 448 U.S. 555, 580–81 (1980) (specifying conditions under which courts may close trials and hearings).

205. Evidence that no adjustment is needed also exists.

[I]t is the experience of the ECF prototype courts and courts which have been imaging documents and making them electronically available that reliance on judicial discretion has not been problematic and has not dramatically increased or altered the amount and nature of motions to seal. It is also the experience of those courts that have been making their case file information available through PACERNet that there have been virtually no reported privacy problems as a result.

JUDICIAL CONFERENCE PRIVACY REPORT, *supra* note 8, at 6.

V. CONCLUSIONS

The federal courts may soon begin requiring the use of data-enabled PDF forms for some kinds of petitions, motions, and orders. Those forms constitute a landmark advance in the potential for court-system transparency because researchers can automatically extract data from them. They would remove the last technological barrier to the complete automation of the process of gathering, analyzing, and reporting basic court-system research.

Once data-enabled forms were in widespread use, government-imposed fees would be the sole remaining barrier to automated transparency. Eliminating those fees would open a real-time window on court-system operation superior to any previously possible. Not only researchers, but parties, lawyers, the government, and the public could see every important aspect of how the courts operate.

The benefits would be tremendous. Policymakers, litigants, and the public could see the amounts of damages granted in personal-injury cases, the lengths of criminal sentences, the likelihood of success on various kinds of motions, the differences in outcomes among courts, the relative effectiveness of lawyers and expert witnesses, and the answers to a myriad of other questions. Policymakers would have the feedback they need to fundamentally improve or fine tune the system. Lawyers could predict the outcomes of their cases. Legal planners could see what works and what does not. The public would, for the first time, be able to see what courts actually do and whether the precautions that members of the public are taking “for legal reasons” are the right ones.

Initially, the wash of new knowledge would embarrass the legal system and the people who run it. Judges and other court officials would find themselves working in a goldfish bowl. Transparency would reveal huge differences in outcomes based on factors that are not supposed to produce any differences at all. Principal among those factors would be the identities of the judges and lawyers. Although the U.S. legal system seems relatively free of the crude kinds of corruption in which litigants bribe judges to change outcomes, transparency might reveal that the system suffers from other kinds of corruption: judges seeking to attract particular kinds of cases, manifesting their ideologies, or favoring particular litigants for personal gain.

Because transparency would shift power, entrenched interests can be expected to oppose it. They will argue that transparency would be costly, but in fact, private volunteers would bear nearly all of the cost. They will argue that transparency would expose parties and witnesses to the threat of identify theft and other harms. In fact, only a few pieces of information raise that threat and transparency is the approach best calculated to remove those pieces from the public record. Other threats to individuals can be addressed through the courts’ power to seal particular documents or portions of documents, without sacrificing transparency’s benefits. Transparency’s

opponents will argue that transparency would make previously difficult-to-obtain public-record information about individuals more easily available. That, however, is an advantage of transparency, not a drawback.

Lastly, they will argue that transparency would result in public pressures on judges to decide cases differently. Some of those pressures are, however, entirely appropriate. Transparency would aid in sorting the pressures that are appropriate in a democracy from those that are not. It would provide the public with the information the public needs to fulfill its oversight function with respect to the courts.

Once the public could see the courts as they actually are, the political pressure to reform them would be intense. Those advantaged under the status quo would fight back with arguments for privacy and confidentiality. But if American institutions are capable of reforming the courts to comply with American ideals of transparency and justice, the human benefits may be at a level unprecedented in history.