

Richmond Journal of Law and Technology

Volume 13 | Issue 3

Article 3

2007

Information Inflation: Can The Legal System Adapt?

George L. Paul

Jason R. Baron

Follow this and additional works at: <http://scholarship.richmond.edu/jolt>



Part of the [Evidence Commons](#), and the [Internet Law Commons](#)

Recommended Citation

George L. Paul & Jason R. Baron, *Information Inflation: Can The Legal System Adapt?*, 13 Rich. J.L. & Tech 10 (2007).

Available at: <http://scholarship.richmond.edu/jolt/vol13/iss3/3>

This Article is brought to you for free and open access by UR Scholarship Repository. It has been accepted for inclusion in Richmond Journal of Law and Technology by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

INFORMATION INFLATION: CAN THE LEGAL SYSTEM ADAPT?

George L. Paul *

*Jason R. Baron***

Cite as: George L. Paul and Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J.L. & TECH. 10 (2007), <http://law.richmond.edu/jolt/v13i3/article10.pdf>.

[1] Information is fundamental to the legal system. Accordingly, lawyers must understand that information, as a cultural and technological edifice, has profoundly and irrevocably changed. There has been a civilization-wide morph, or pulse, or one might say that information has evolved. This article discusses the new inflationary dynamic,¹ which has caused written information to multiply by as much as ten thousand-fold recently.² The resulting landscape has stressed the legal system and

*George L. Paul is a partner in Lewis and Roca, LLP, and is a graduate of Dartmouth College (1979) and Yale Law School (1982).

** Jason R. Baron is Director of Litigation at the National Archives and Records Administration, College Park, Maryland, and is a graduate of Wesleyan University (1977) and Boston University School of Law (1980). The authors wish to thank Jeffrey Albright, John Messing, Doug Oard, Judge Shira Scheindlin, Rachel Spector, Gary Stern, Ted Hirt, Paul Thompson, and Ellen Voorhees for their helpful review, comments and suggestions. Mr. Baron also wishes to gratefully acknowledge the written contributions to Part II.A, *infra*, provided by Andrew O'Malley, J.D. Class of 2008, at The George Washington University Law School, as well as research assistance provided by Sarah Rudgers. The views expressed in this Article are the authors' alone, and should not be attributed, in whole or in part, to either Lewis and Roca, LLP, or to any component of the U.S. government.

¹ The word "inflation" is taken from the hypothesis that the universe in its earliest stages experienced a period of rapid expansion, i.e., "inflation." See ALAN H. GUTH, *THE INFLATIONARY UNIVERSE: THE QUEST FOR A NEW THEORY OF COSMIC ORIGINS* (1997).

² See GEORGE L. PAUL & BRUCE H. NEARON, *THE DISCOVERY REVOLUTION: E-DISCOVERY AMENDMENTS TO THE FEDERAL RULES OF CIVIL PROCEDURE 4-5* (2006)

indeed, it is becoming prohibitively expensive for lawyers even to search through information. This is particularly true in litigation.

[2] As problematic as quantity are the diverse new forms of writing which emerge constantly as a consequence of information inflation. Given that lawyers must retrieve and synthesize information, we must ask how our system should adapt to these new forms of information life. And what tools can be developed to help? It is no exaggeration to say that litigation, as we have known it, is threatened by information's new hyper-flow. The amount of electronically stored information relevant to a case is already a stress point in litigation. What might be the result if, in three or four years, there is ten times as much information in enterprises as there is today—or in ten years, 50 times as much as there is today? To what extent will litigators of the future be able to rely on or reasonably work from a complete evidentiary record? This article suggests and briefly discusses several possible solutions to such challenges.

[3] First, there must be a change in culture among litigation lawyers. The last 30 years have seen truculence, gamesmanship, and a supreme rule of "volunteer nothing." Because of the new complexity and volume of information, however, the game theory underlying much of litigation has changed.³ Litigators must collaborate far more than they have in the past, particularly concerning the discovery of information systems. If they do not, they act against their own self-interest. The new "e-discovery" Federal Rules of Civil Procedure explicitly provide for this collaborative process, but fall short in explaining why such collaboration is essential, the extent to which it must occur, or its necessary iterative fashion. This iterative collaboration signifies a needed revolution in discovery practice.

[4] Next, a family of computer technology employing new types of search methods and techniques beyond the use of mere keywords should now be considered for use in litigation.⁴ In particular, lawyers and judges should

("Organizations now have thousands if not tens of thousands of times as much information within their boundaries as they did 20 years ago.").

³ PAUL & NEARON, *supra* note 2, at 169.

⁴ Generically called "search and retrieval" technology, such methods create a database of all candidate files, including their text and metadata, and then use computer processors to identify documents with a designated word, or combination of words, or probability of appearance of words. *Keyword* searching, *Boolean* searching, *fuzzy logic*, *Bayesian*

be far more tolerant of using statistical techniques, like sampling, as part of a reasonable search process. Litigators can no longer depend on manual review alone.⁵ It is too time-consuming and expensive – with costs often exceeding the amount in dispute. Yet the use of machines to search written records continues to pose a challenge, as language is an ever-evolving, elastic form which has proved notoriously hard to search.⁶ In addition, there is the issue of the necessary skill, or *technique*, to use such computer search tools. How many in the profession have such skill? Have such tools been adequately tested or proven? No. But there is no choice but to shape new tools, and new processes for using them.⁷

[5] Third, there must be innovation in the law, particularly governing inadvertent disclosure of privileged information. New rules on inadvertent disclosure should be given effect, which limit or eliminate the waiver of privilege, or the system will become impossibly expensive through continued need for meticulous review.

[6] Finally, from this point forward lawyers must embrace creative, technological approaches to grappling with knowledge management as information inflation continues apace. Failure to do so will severely hamper the legal profession's ability to meaningfully retrieve and process evidence. All this equates to perhaps the biggest new skill set ever thrust upon the profession – a revolution for the practice. What it means to be a lawyer will change rapidly in the years to come.

belief networks, *vector space models*, and the use of *taxonomies* and *ontologies* are all examples of such search and retrieval technology, as discussed in Part II.B.

⁵ The costs of manual review now lead many litigants to hire armies of “information reviewers” located in India, Pakistan, the Philippines, and elsewhere where costs of educated labor are lower than in the United States. See John Tredennick, *Your Next Office: Bangladore?*, LAW PRACTICE TODAY, July 2005, <http://www.abanet.org/lpm/lpt/articles/fwr07051.html>.

⁶ See, e.g., DAVID BLAIR, WITTGENSTEIN, LANGUAGE AND INFORMATION: ‘BACK TO THE ROUGH GROUND!’ 302 (2006) (summarizing prior 1985 Blair & Maron study in which retrieval effectiveness was measured for 40,000 documents captured in a large corporate litigation, where results of the study showed a large amount of indeterminacy of meaning in natural language in light of the fact that “while [the] lawyers and paralegals were convinced that they were retrieving over seventy-five percent of the desired documents, they were, in actuality retrieving only twenty percent!”). See also *infra* Part II(B).

⁷ See *infra* Part II(B)(2).

I. A RECENT EVOLUTION IN WRITING HAS CAUSED
INFORMATION INFLATION

[7] Writing co-evolved with civilization more than 50 centuries ago.⁸ The original technology combined a system of markings, called graphs or script, with an alteration of physical matter: clay tablets, stone, wax, papyrus, bronze, bark, cloth, parchment, leather, wood, paper, carbon paper, and in short, anything one could write on.⁹ From the inception of the invention, molecules were displaced, leaving an original record readable by human beings. As such, mankind's recorded communications have long been confined to the physical realm – frozen in time as "information artifacts." Information technology was simple, static, material, and lifeless. The technology remained in equilibrium for over 5200 years.¹⁰

[8] There has been only one transformative advance in the original writing technology. Circa 1450 Johannes Gutenberg invented the movable type printing press, which dramatically lowered the cost of producing written records.¹¹ The printing press allowed mass production of information¹² and thus contributed to the Renaissance, the Scientific Revolution, and the

⁸ A mature system of script is estimated to have arisen in approximately 3200 B.C. as a Sumerian invention. *See* ALBERTINE GAUR, A HISTORY OF WRITING (1984); ROY HARRIS, THE ORIGIN OF WRITING (1986).

⁹ Script has its own history and is said to have reached its apex with the invention of the Greek alphabet around 900 B.C. The Greek innovation was to add a sign for vowels to the sign for consonants that already existed in the Semitic pre-alphabet. The alphabet is considered the most efficient of scripts with its few letters able to form any word in a language. It was invented only once but has been borrowed many times, for example by the Romans. We use the Greek alphabet today. *See* 2 DAVID DIRINGER, THE ALPHABET: A KEY TO THE HISTORY OF MANKIND (3d ed. 1968); WALTER DURFEE, ALPHABETICS AS A SCIENCE (1956).

¹⁰ *See* GAUR, *supra* note 8.

¹¹ Before the printing press, a law book cost as much as an average worker's living expenses for over a year. *See, e.g.*, CARLO M. CIPOLLA, MONEY, PRICES AND CIVILIZATION 61 (1956).

¹² The typewriter, which was invented shortly after the American Civil War, was another major writing invention. Originally manufactured by the Remington Arms Company, the typewriter revolutionized business writing and, before the advent of the computer, was probably the number one "business machine." *See* MARTIN CAMPBELL-KELLY & WILLIAM ASPRAY, COMPUTER: A HISTORY OF THE INFORMATION MACHINE 30-34 (1996).

Protestant Reformation.¹³ The invention demonstrates that the more freely information flows, the more dynamic the evolution of culture.

[9] But quite recently there has been an evolutionary burst in writing technology – a jagged punctuation on a 50 century-long sine wave.¹⁴ A quick succession of advances clustered¹⁵ or synced¹⁶ together, to emerge into a radically new and more powerful writing technology. These include digitization;¹⁷ real time computing;¹⁸ the microprocessor;¹⁹ the personal

¹³ Before Gutenberg's invention it took many months to make a book. After the invention the first printers were able to print, or "pull," about 300 pages a day, with the rate up to 1,000 pages a day at the end of the fifteenth century. CARLO M. CIPOLLA, *BEFORE THE INDUSTRIAL REVOLUTION: EUROPEAN SOCIETY AND ECONOMY 1000-1700*, at 106 (1976). Because of the importance of his invention, Gutenberg has been proclaimed the most important person of the last millennium. AGNES HOOPER GOTTLIEB ET AL., *1000 YEARS, 1000 PEOPLE 2* (1998). See also Johannes Gutenberg, Wikipedia, http://en.wikipedia.org/wiki/Johannes_Gutenberg (last visited March 6, 2007) (asserting that Gutenberg's invention stimulated the Renaissance and the Scientific Revolution).

¹⁴ Niles Eldridge and Stephen Jay Gould wrote a seminal paper on the nature of evolution in which they pointed out that forms of life stay stable for long periods, and then quickly morph into new forms in a burst of evolution. See NILES ELDRIDGE & STEPHEN JAY GOULD, *PUNCTUATED EQUILIBRIA: AN ALTERNATIVE TO PHYLETIC GRADUALISM* (1972), reprinted in *MODELS IN PALEOBIOLOGY* 82-115 (Thomas J. M. Schopf ed., 1972). This same "fit and start" pattern can be seen in physico-chemical systems, economic growth, ecosystems, the climate, technology, and the evolution of human culture.

¹⁵ Technologists speak of subsystems "clustering" or "bundling" together to form technologies, which emerge from constituent parts. See ARNULF GRÜBLER, *TECHNOLOGY AND GLOBAL CHANGE* 5, 19 (1998). See also LEWIS MUMFORD, *TECHNICS AND THE NATURE OF MAN* (1966), reprinted in *PHILOSOPHY AND TECHNOLOGY: READINGS IN THE PHILOSOPHICAL PROBLEMS OF TECHNOLOGY* 77-85 (Carl Mitcham & Robert Mackey eds., 1972).

¹⁶ See STEVEN STROGATZ, *SYNC: THE EMERGING SCIENCE OF SPONTANEOUS ORDER* (2003); E.N. Lorenz, *Deterministic Nonperiodic Flow*, 20 *J. ATMOSPHERIC SCI.* 130 (1963); Louis M. Pecora & Thomas L. Carroll, *Synchronization in Chaotic Systems*, 64 *PHYSICAL REV. LETTERS* 821-24 (1990).

¹⁷ Digital computers were invented at the close of World War II, and gradually diffused throughout society as the "mainframe" of the 1950s, 1960s and 1970s. See Computer, Wikipedia, http://en.wikipedia.org/wiki/Digital_computers (last visited March 6, 2007).

¹⁸ Although computers were used by institutions, governments, and large businesses in the 1950s, 1960s, and 1970s, these were primarily expensive mainframes that performed "batch processing." Often, the results of the computation were not available for hours, or until the next day, because the process of computation was separate from the human user interface. The first commercial real-time computing system was American Airlines's SABRE reservations system, which was deployed in 1964. It was not until the 1980s that

computer,²⁰ e-mail;²¹ local and wide-area networks leading to the Internet;²² the evolution of software, which has “locked in” seamless editing as an almost universal function;²³ the World Wide Web;²⁴ and of

“real-time” computing became widely available. CAMPBELL-KELLY & ASPRAY, *supra* note 12, at 170-75.

¹⁹ Three inventions claim to have been the first computer in a chip or microprocessor: The Central Air Data Computer (CADC), the Intel 4004, and the Texas Instruments (TI) TMS 1000. The CADC system was completed for the Navy's "TomCat" fighter jets in 1970; the TI TMS 1000 was first to market in calculator form; the first stand-alone microprocessor was the Intel 4004, which was introduced in November 1971. See W. Warner, *Great Moments in Microprocessor History*, <http://www-128.ibm.com/developerworks/library/pa-microhist.html> (last visited March 6, 2006).

²⁰ These machines used the newly-invented microprocessor and spread after 1977, with leading products being the Apple II and Commodore PET. Such “personal computers” gained ascendancy in business in the mid-1980s. See CAMPBELL-KELLY & ASPRAY, *supra* note 12, at 247-53.

²¹ See KATIE HAFNER & MATHEW LYON, *WHERE WIZARDS STAY UP LATE: THE ORIGINS OF THE INTERNET* 187-218 (1996). See also PAUL & NEARON, *supra* note 2, at 4; Ray Tomlinson, Wikipedia, http://en.wikipedia.org/wiki/Ray_Tomlinson (last visited March 6, 2007) (describing the first e-mail, which was sent in 1971 between users on different hosts connected on a network).

²² The networking dynamic is fundamental to the new writing paradigm. Corporate networks using communication protocols gained ascendancy in the 1980s, and were fully deployed by the mid-1990s. See CAMPBELL-KELLY & ASPRAY, *supra* note 12. See generally HAFNER & LYON, *supra* note 21 (describing the growth of ARPANET).

²³ One of the designs “locked into” most business productivity software is seamless editing. See also THOMAS L. FRIEDMAN, *THE WORLD IS FLAT: A BRIEF HISTORY OF THE TWENTY-FIRST CENTURY, VERSION 2.0*, at 76-92 (2006) (discussing the importance of workflow applications).

²⁴ The World Wide Web (the “Web”), a communications network that operates on the Internet, dates to circa 1990, when Tim Berners-Lee of CERN invented hypertext markup language (HTML). See generally TIM BERNERS-LEE, *WEAVING THE WEB: THE ORIGINAL DESIGN AND ULTIMATE DESTINY OF THE WORLD WIDE WEB BY ITS INVENTOR* (1999). Mosaic Communications Corporation was the first company to attempt to capitalize on the Web when it released a “web browser” known as Mosaic Netscape 0.9, later renamed Netscape Navigator. On August 9, 1995, newly renamed “Netscape” went public and the Web caused the Internet to grow with inflationary force. See Netscape, Wikipedia, http://en.wikipedia.org/wiki/Netscape_Communications_Corporation (last visited March 6, 2007). The Web went from 20 host servers in 1992; to 50 as of January 1993; 200 as of August 1993; to 100,000 (post-Netscape) in January 1996; 650,000 in January 1997; and 4.3 million in 1999. See BERNERS-LEE, at 67; History of the Virtual Library, <http://vlib.org/admin/history> (last visited March 6, 2007). Now there are *tens of millions* of hosts and complete networks are added to the Internet daily. ROMUALDO PASTOR-SATORRAS & ALESSANDRO VESPIGNANI, *EVOLUTION AND STRUCTURE OF THE INTERNET: A STATISTICAL PHYSICS APPROACH* 7 (2004).

course people²⁵ and their *technique*.²⁶ These constituents have swirled into an information complex, now known as the "Information Ecosystem."²⁷ In such a system, the whole exhibits an emergent behavior that is much more than the sum of the parts.²⁸ Critically for law, such systems cannot be understood or explained by any one person.²⁹ As a result, writing has now grown into something akin to a new "form of life."³⁰ Because of its long-standing stasis and the importance of writing as a global technology, such a development may legitimately be said to herald a new phase of civilization.³¹

²⁵ Philosophers and technologists have described the synthesis of human beings with their technological hardware as a sort of "archetypal machine composed of human parts" – a sort of "mega-machine," where technology and human beings are linked to emerge into a larger whole, such as a civilization or an economy's "invisible hand." See, e.g., MUMFORD, *supra* note 15. See also 1-2 ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS (R.H. Campbell & A.S. Skinner, eds., photo. reprint 1981) (1776).

²⁶ Technologists use the French word *technique* to refer to the disembodied nature of technology – the knowledge base of "how" to use technology "hardware." See GRÜBLER, *supra* note 15, at 20.

²⁷ George L. Paul & Robert Copple, *No, You Can't Call Them Documents Anymore*, BUS. LAW TODAY, Mar.-Apr. 2005, at 39; PAUL & NEARON, *supra* note 2, at 13. Such an information ecosystem can be referred to as an *oikos*, the ancient Greek word for "home, house or habitation," which was used by German zoologist Ernst Haeckel (1843-1919) in 1873 to coin the word "ecology" or *Okologie*, from Gk. *oikos* + *-logia* "study of."

²⁸ See PAUL & NEARON, *supra* note 2, at 4. See also GREGOIRE NICOLIS & ILYA PRIGOGINE, EXPLORING COMPLEXITY: AN INTRODUCTION (1989); ILYA PRIGOGINE & ISABELLE STENGERS, ORDER OUT OF CHAOS: MAN'S NEW DIALOGUE WITH NATURE (1984) (explaining the concepts of "emergence," "complexity," "behavior," and providing a general description of far-from-equilibrium thermodynamic systems).

²⁹ ROBERT W. RYCROFT & DON E. KASH, THE COMPLEXITY CHALLENGE: TECHNOLOGICAL INNOVATION FOR THE 21ST CENTURY 3 (John de la Mothe, ed., 1999).

³⁰ LUDWIG WITTGENSTEIN, THE PHILOSOPHICAL INVESTIGATIONS § 19 (G.E.M. Anscombe, trans., The Macmillan Co., 1953) ("[T]o imagine a language is to imagine a form of life.").

³¹ Suddenly, the amount of information that an ordinary human being can communicate through writing has increased exponentially. Because of the complex, emergent aspects of the various subsystems syncing to form the new writing technology, this new phase of civilization might be called the Age of Information Complexity. See PAUL & NEARON, *supra* note 2, at 4-8.

A. THE LEGAL PROFESSION CONFRONTS AN INFLATIONARY EPOCH

[10] In the original writing technology, the rate of flow of information was limited because it depended on distribution of information artifacts. With the plastic and networked nature of new age writing, we are no longer wedded to original records. We can distribute thousands or even millions of identical records in an instant. These can be read in real time, affecting other people, who are also imbedded in the system.³² Recipients modify the system further by sending out their own writings, often editing messages they receive. All this leaves multitudinous records of thought, word, and action as evidence.³³ We can edit; change formats; respond; converse with twenty people at once; and even move, speak, and write in virtual worlds as an avatar.³⁴

[11] A scientist would say that the *flux* enabled by our new writing is qualitatively different than in civilization's original phase.³⁵ Information

³² More than television, interactive participation in the new writing stimulates hormonal systems and creates changes in behavior. *Resolution on Violence in Video Games and Interactive Media*, 61 AMER. PSYCHOLOGIST 490, July 2006 (citing seventy-one published psychological studies).

³³ The explosion of instant messaging is just one recent application of this effect. *See* Growth of Instant Messaging, <http://www.stanford.edu/~sdouglas/Instant%20Messenger/node%201.htm> (last visited March 6, 2007) (projecting 1.4 billion instant message accounts in 2007).

³⁴ The modern mega-multiplayer interactive society is a good example of the richness of the new forms of writing. Tens of millions of people spend hours a day playing such games as *Everquest 2*, *World of Warcraft*, and *Second Life*. They live in a virtual world, interacting sometimes to attack each other, but often merely to communicate or to join forces in guilds to fight against creatures called "The Boss." They sell land, sell their bodies, run gambling parlors, design and construct buildings, buy and spend virtual money, hack into each others' accounts to steal virtual property, and now even sue one another in "reality" for being defrauded in virtual transactions. Players often simultaneously move their avatar, chat, and talk over a VOIP communications channel. Voluminous logs of chat are kept by administrators of the game to resolve disputes and to police the virtual world for adherence to rules. This new, highly evolved "writing of the avatars" is an excellent example of the degree to which new forms of writing will rapidly evolve.

³⁵ *See* NICOLIS & PRIGOGINE, *supra* note 28, at 54. There are many fluxes used in the study of "transport phenomena." Some of the most common are: (1) *heat flux*, the rate of heat flow across a unit area (*Fourier's Law*); (2) *chemical flux*, the rate of movement of molecules across a unit area (*Fick's law of diffusion*); (3) *volumetric flux*, the rate of volume flow across a unit area (*Darcy's law*); (4) *mass flux*, the rate of mass flow across

is now zipping, pulsing, and being channeled by catalysts into new social order much more rapidly than before.³⁶ Inflation causes both an outward expansion of the quantity of information, and reveals a dynamic³⁷ of innovation – of relentless diversity of new form. Information inflation has created a new age explosion of conceptual life.³⁸

[12] Probably close to 100 billion e-mails are sent daily,³⁹ with approximately 30 billion e-mails created or received by federal government agencies each year.⁴⁰ The amount of stored information continues to grow exponentially.⁴¹

a unit area (either an alternate form of *Fick's law* that includes the molecular mass, or an alternate form of *Darcy's law* that includes the density); and (5) *radiative flux*, the amount of energy moving in the form of photons at a certain distance from the source per steradian per second (used in astronomy to determine the magnitude and spectral class of a star). See Flux, Wikipedia, <http://en.wikipedia.org/wiki/Flux> (last visited March 6, 2007). This article proposes a new “writing flux” of information flowing from human beings.

³⁶ Human beings’ thoughts and existence are entangled in their communications, both verbal and written/read. Humans have a “scaffolding” that integrates their being into reality, going beyond the “mind.” See BLAIR, *supra* note 6, at 270-85; ANDY CLARK, BEING THERE: PUTTING BRAIN, BODY, AND WORLD TOGETHER AGAIN (1997). The “scaffolding” for human beings has radically changed recently. One can see this with young people, expressing their new forms of social life; with adults sending text messages with omnipresent cell phones; and in the other ways in which business communications have changed. Another example of how “scaffolding” has changed is the degree to which we are now addicted to our computer networks, as compared to the workplace of 20 years ago.

³⁷ Living systems tend to maximize the construction of the diversity of their components. They maximize “what can happen next.” They maximize the average sustained growth of their own dimensionality, unfolding in “phase space” or “state space.” See STUART KAUFFMAN, INVESTIGATIONS 3-4 (Oxford 2000); NICOLIS & PRIGOGINE, *supra* note 28, at 79-89. An oikos also unfolds in phase space.

³⁸ Just as life *exploded* with diversity after the innovation of complex, multi-cellularity (the “Cambrian Explosion”), so too has information diversified after writing became “complex.” See STEPHEN JAY GOULD, WONDERFUL LIFE: THE BURGESS SHALE AND THE NATURE OF HISTORY (1989) (discussing the Cambrian Explosion).

³⁹ Peter Lyman & Hal R. Varian, *How Much Information?* (2003), <http://www.sims.berkeley.edu/how-much-info-2003>.

⁴⁰ Jason R. Baron, *E-mail Metadata in a Post-Armstrong World*, Presented at the 3rd IEEE Computer Soc’y Metadata Conf. (1999), *available at* <http://www.archives.gov/era/pdf/baron-email-metadata.pdf>.

⁴¹ See Lyman & Varian, *supra* note 39.

[13] Perhaps more easily grasped, the amount of information in business has increased by thousands, if not tens of thousands of times in the last few years. In a small business, whereas formerly there was usually one four-drawer file cabinet full of paper records, now there is the equivalent of two thousand four-drawer file cabinets full of such records, all contained in a cubic foot or so in the form of electronically stored information.⁴² This is a sea change.

B. WHITE HOUSE CASE STUDY: THE PROBLEM OF A BILLION E-MAILS

[14] The explosive growth of information is perhaps best illustrated by litigation over White House e-mail. This trove of government e-records has fed Washington scandals and spawned a number of related lawsuits.⁴³

[15] In 1989 on the last day of the Reagan Administration, public interest groups filed suit to ensure that backup tapes containing Iran-Contra e-mails were preserved during the transition to the Bush Administration.⁴⁴ At the outset of the case, some 392 backup tapes were subject to a

⁴² See PAUL & NEARON, *supra* note 2, at 4-5.

⁴³ *Armstrong v. Bush*, 721 F. Supp. 343 (D.D.C. 1989), *aff'd in part, rev'd in part*, 924 F.2d 282 (D.C. Cir. 1991). See also *Armstrong v. Executive Office of the President*, 1 F.3d 1274 (D.C. Cir. 1993) (ordering continued preservation of PROFS backup tapes holding Iran-Contra e-mail, and ruling that various EOP components must preserve the electronic versions of e-mail in light of metadata); *Armstrong v. Executive Office of the President*, 90 F.3d 553 (D.C. Cir. 1996), *cert. denied*, 520 U.S. 1239 (1997) (holding e-mail created by the National Security Council to be within scope of Presidential Records Act, 44 U.S.C. § 2201, and thus not subject to immediate FOIA access); *Alexander v. FBI*, 198 F.R.D. 306 (D.D.C. 2000) (missing White House e-mail gives rise to lengthy evidentiary proceedings); *Public Citizen v. Carlin*, 2 F.Supp.2d 1 (D.D.C. 1996), *rev'd*, 184 F.3d 900 (D.C. Cir. 1999), *cert. denied*, 529 U.S. 1003 (2000) (upholding Archivist's authority to issue a general records schedule allowing for deletion of e-mail subject to conditions); *American Historical Ass'n v. Peterson*, 876 F. Supp. 1300 (D.D.C. 1995) (invalidating controversial Bush-Wilson agreement); U.S. GEN. ACCOUNTING OFFICE, ELECTRONIC RECORDS: CLINTON ADMINISTRATION'S MANAGEMENT OF EXECUTIVE OFFICE OF THE PRESIDENT'S E-MAIL SYSTEM, (April 2001), available at www.gao.gov/new.items/d01446.pdf. For a detailed discussion of *Armstrong* and related cases, see Jason R. Baron, *The PROFS Decade: NARA, E-mail, and the Courts*, in THIRTY YEARS OF ELECTRONIC RECORDS 105-37 (Bruce Ambacher ed., 2003); Baron, *supra* note 40.

⁴⁴ See *Armstrong v. Bush*, 721 F. Supp. at 347.

preservation order.⁴⁵ After a decade of litigation, the universe of e-mail and backup tapes swept into the case had ballooned to encompass nearly 6,000 original backup tapes.⁴⁶ Additionally, the White House decided to settle the litigation, at least in part, by agreeing in July 1994 to implement a form of electronic recordkeeping system known as “ARMS” (the Automated Records Management System).⁴⁷ As a result, by the end of the Clinton Administration over 32 million e-mails created or received within the Executive Office of the President had been electronically preserved for accessioning with the National Archives and Records Administration (NARA).⁴⁸

[16] The 32 million e-mails were a prime subject of discovery in *U.S. v. Philip Morris*, the RICO lawsuit filed by the government in 1999 against numerous tobacco companies.⁴⁹ In that action, defendants filed 1,726 requests to produce against 30 agencies, requesting U.S. records going back to the early 1950s, as well as all relevant e-mails. The government responded to discovery as best it could using its available resources – but decidedly in what will be deemed here as a “pre-FRCP rules change mode,” namely: with respect to e-records, government lawyers oversaw searches by unilaterally choosing the set of search terms using simple keywords.⁵⁰ Later, a secondary search was conducted based on limited negotiations held with and input from defendants with respect to additional keywords.⁵¹

⁴⁵ See *id.* at 347-48.

⁴⁶ Thomas E. Brown, *History of NARA’s Custodial Program for Electronic Records: From the Data Archives Staff to the Center for Electronic Records, 1968-1998*, in THIRTY YEARS OF ELECTRONIC RECORDS 16 (Bruce Ambacher ed., 2003).

⁴⁷ See *Armstrong v. Executive Office of the President*, 877 F. Supp. at 715 (Exhibit C, setting out guidance issued to White House staff at what was the onset of the “paperless” era for retention of e-mail within the EOP).

⁴⁸ See Allen Weinstein, Archivist of the United States, Ask the White House (Jan. 17, 2006), <http://www.whitehouse.gov/ask/20060117.html>.

⁴⁹ See *United States v. Philip Morris USA*, 449 F.Supp.2d 1 (D.D.C. 2006) (1,600 page opinion by Judge Gladys Kessler).

⁵⁰ Keywords included such common terms as “tobacco,” “smoking,” “cigarette,” “tar,” “nicotine,” and “Philip Morris.” See Jason R. Baron, *Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery*, 6 SEDONA CONF. J. 237, 239 (2005).

⁵¹ *Id.*

[17] For the 18 million presidential record e-mails within NARA's legal custody, the search process found a universe of some 200,000 "hits," of which over 100,000 were later determined to be responsive in evidencing tobacco policies and practices.⁵² In undertaking a second-stage manual search to determine responsiveness, it was necessary to put into place a team of twenty-five lawyers, law clerks, and archivists working at NARA more or less full time over a period of six months, reviewing 200,000 e-mails on CDs, e-mail by e-mail, attachment by attachment, for the purpose of printing out hard copies to be used.⁵³ Additional time and resources were needed to make a further review for privilege.⁵⁴

[18] The approach taken by the government in *Philip Morris* most charitably represented a stopgap set of measures put in place to deal with a burdensome search request which placed strains on administrative resources. But the essentially unilateralist, pre-December 1, 2006 FRCP rules-changes approach to confronting a difficult search task clearly fails to *scale up*; for were the e-mail universe ten times as large, it would have been impossible to assemble a large enough team of reviewers to devote the time for them to manually review each e-mail "hit," and its attachments, for responsiveness and privilege.

[19] Critically, as of January 20, 2009, NARA expects to receive substantially over a hundred million e-mails from the current incumbent White House.⁵⁵ At the present rate of e-mail creation, NARA expects to receive over *one billion* e-mails over the course of the next decade as permanently accessioned records of the government. Some parties in litigation have apparently *already* crossed the billion electronic document threshold.⁵⁶

⁵² *Id.*

⁵³ *Id.*

⁵⁴ *Id.* at 239-40.

⁵⁵ See Allen Weinstein, *supra* note 48.

⁵⁶ See John H. Jessen, *Special Issues Involving Electronic Discovery*, 9 KAN. J.L. & PUB. POL'Y 425 (2000). "We have had about half a dozen cases now where the total number of electronic things brought into play – not that were available in a global set, but which were available after a reasonable initial review of the set – went over one billion. A billion pieces of discovery material," reports Jessen, the chief officer of his electronic discovery firm. *Id.* at 428. He further points out, "Those kind of numbers introduce a whole host of issues about scope and management. How do you manage a billion things . . . ?" *Id.*

[20] Take then, for example, litigation in which the universe subject to search stands at one billion e-mail records, at least 25% of which have one or more attachments of varying length (1 to 300 pages). Generously assume further that a model “reviewer” (junior lawyer, legal assistant, or contract professional) is able to review an average of fifty e-mails, including attachments, per hour.⁵⁷ Without employing *any* automated computer process to generate potentially responsive documents, the review effort for this litigation would take 100 people, working ten hours a day, seven days a week, fifty-two weeks a year, over fifty-four *years* to complete.⁵⁸ And the cost of such a review, at an assumed average billing of \$100/hour, would be \$ 2 billion.⁵⁹ Even, however, if present-day search methods (such as in the tobacco litigation example) are used to initially reduce the e-mail universe to 1% of its size (i.e., 10 million documents out of 1 billion), the case would still cost \$20 million for a first pass review conducted by 100 people over 28 weeks, without accounting for any additional privilege review. Given the exponential growth of e-mail, as well as its continued accumulation on a legacy basis in many corporate databases, it is not far-fetched to project that in the future the need to manually review even as much as “1%” of what will be ever-growing corporate data sets will prove unmanageable if the quality of the initial search is not improved.⁶⁰ The numbers add up to more of a burden than any party should assume, no matter how rich in resources, without changes being made to the way cases are litigated and to techniques used in discovery.

⁵⁷ Based on the authors’ experiences with the TREC legal track research project, see *infra* note 122, this estimate generously assumes a *greater* industry on the part of our hypothetical reviewer than was evidenced by forty or more actual (human) assessors participating in the first year of this project.

⁵⁸ One billion e-mails at the rate of fifty e-mails per person/hour equals a total of twenty million “person hours” to review. One hundred people working seventy hours per week, fifty-two weeks a year equals 364,000 person hours/year. Dividing twenty million person hours by 364,000 person hours per year = 54.95 years.

⁵⁹ Lowering the cost of labor to \$10/hour by going “off-shore,” the cost would still be over \$200 million.

⁶⁰ “Searchers are faced with the classic problem of finding a needle in a haystack, but the haystack is growing so rapidly that there is a continual demand for improved search technology.” Douglas W. Oard & Jinmook Kim, *Modeling Information Content Using Observable Behavior*, Presented at the Nov. 2001 Conf. of the Am. Soc’y for Info. Sci. & Tech. (2001), available at <http://www.glue.umd.edu/~oard/research.html#recommender>.

C. BRAVE NEW WORLDS OF INFORMATION

[21] Of course, e-mail is not the only e-record on the corporate and government desktop: the diverse universe of information objects in 2006 currently consists of instant messaging, word processing with hyperlinks, integrated voice mail in “.wav” file format, structured databases of all kinds, Web pages, blogs, and e-data in all conceivable forms. Instant messaging alone is a booming phenomenon, which in some institutions, may yet even eclipse total e-mail traffic.⁶¹ What might be perceived as the e-discovery of the “near-term” future is tantalizing, and always seems to arrive more rapidly than predicted. For example, Professor Cass R. Sunstein posits in his 2006 book, *Infotopia*, a “possible future” where:

[T]he day-to-day operations of the department are strikingly different from what they were in the earliest years of the twenty-first century. Many of the department’s internal documents are ‘wikis’ – Web pages that are highly secure but can be freely edited by anyone who has access to them Department of Defense lawyers have a wiki for critical legal issues, informally named Wikilaw and containing an extraordinary amount of material about legal problems of particular concern to the department. Some important files involving national security operate as wikis, too. These files are edited several times each day, as new information emerges.⁶²

[22] Yet, as of late 2006, some government agencies already had begun to utilize collaborative software in the form of wikis, including in the area of national security.⁶³ Accordingly, these agencies should expect future

⁶¹ A reported twelve billion instant messages are sent daily worldwide. Gene J. Koprowski, *Instant Messaging Grew by Nearly 20 Percent in 2005*, TECHNEWSWORLD, Nov. 10, 2005, <http://www.technewsworld.com/story/47270.html>. See also *Don’t Shoot the Instant Messengers*, SIEMENS BUS. SURV., Nov. 24, 2004 (“[I]nstant messaging is eclipsing e-mail and phone use at the office”), <http://www.sbs-sa.siemens.com/Press/11.16.04.asp>.

⁶² CASS R. SUNSTEIN, *INFOTOPIA* 3-4 (2006).

⁶³ See Clive Thompson, *Open-Source Spying*, N.Y. TIMES SUNDAY MAG., Dec. 3, 2006, at 54. As of late 2006, NARA issued records management guidance recognizing the growing use of wikis as part of the daily business of government. See *The National Archives: Implications of Recent Web Technologies For NARA Web Guidance*,

access demands of all kinds encompassing these types of records, including under FOIA, pursuant to subpoenas and investigations, and, of course, in civil discovery.

[23] As there is no end in sight for the continued exponential expansion of quantity and form across all types of e-discovery boundaries, private corporations and public institutions of all kinds are in the same inflationary boat, i.e., the volume and forms implicated in e-discovery require new strategies if there is any hope of accomplishing the task of finding responsive information in finite time periods. Without employing new strategies and techniques, it is foreseeable that some parties in complex litigation will operate at an increased risk of sanction – merely because of their inability to process the requisite volume of information. Notwithstanding the December 1, 2006 modifications to the federal discovery rules, this litigation threat can only be expected to grow unless there are changes in legal practice and behavior.⁶⁴ But more, can the system as we have known it even function without an ability to meaningfully examine information?

<http://www.archives.gov/records-mgmt/initiatives/web-tech.html> (discussing federal records management issues concerning wikis, blogs, Web portals, and Really Simple Syndication (RSS) feeds).

⁶⁴ On the one hand, the Advisory Committee to the 2006 rules amendments appears to assume that technological progress in the area of “search” has made litigation easier, with respect to large volumes of information; on the other hand, the Committee also recognizes that considerations of “volume” may yet have deleterious effects on litigation across a number of e-discovery areas. *Compare* FED. R. CIV. P. 26(b)(2) (COMM. NOTE) (“The *volume* of – and the ability to search – much electronically stored information [ESI] means that in many cases the responding party will be able to produce information from reasonably accessible sources that will fully satisfy the parties’ discovery needs.”), *with* FED. R. CIV. P. 26(f) (COMM. NOTE) (“The *volume* and dynamic nature of [ESI] may complicate preservation obligations . . . The *volume* of such data, and the informality that attends use of email and some other types of [ESI] may make privilege determinations more difficult, and privilege review correspondingly more expensive and time consuming.”) (emphasis added).

II. STRATEGIES FOR COPING WITH INFORMATION INFLATION IN LITIGATION

[24] Information inflation calls for new and different approaches to the problems of e-discovery in all areas, from collection, to format preservation, to search and retrieval, to production, to ultimate access for use as evidence.

[25] At least two present-day “meta-strategies” are apparent. First, it is essential that the litigation system adopt new strategies for cooperation in the context of the adversary process, for promotion of efficiency and transparency. Next, in many cases litigation will require new ways of thinking about the computer-assisted search and retrieval of information. A more complete answer may yet arrive – the use of various forms of artificial intelligence including sophisticated data mining, links analysis, and other more or less science-fiction sounding measures yet to be harnessed in litigation contexts.

A. INVOKING A NEW STRATEGIC COOPERATION PARADIGM

[26] Lawyers need to re-engineer the process of interacting with opposing counsel to promote efficiency, transparency, and the “just and speedy” resolution of disputes consistent with Rule 1 of the Federal Rules of Civil Procedure. A new, collaborative paradigm for the 21st century is in order. There is a need for the development of case law that makes explicit what, for the past 70 years or so, has been left as a largely unstated goal of “cooperation” within the adversary system. If early meet and confer obligations result in a body of case law better defining cooperative behavior, it will likely be the most profound, “emergent” aspect of legal practice under the December 2006 changes to the Federal Rules of Civil Procedure.

[27] Quite simply, as courts and commentators have increasingly come to expressly recognize, the volume and complexity of electronically stored information demand new forms of collaboration.⁶⁵ In turn, in many such

⁶⁵ See *BG Real Estate Serv., Inc. v. Am. Equity Ins. Co.*, 2005 WL 1309048 (E.D. La. May 18, 2005) (noting, in the context of an e-discovery dispute involving preservation issues, that “it is hoped that reasonable lawyers can cooperate to manage discovery without the need for judicial intervention.”); Robert D. Brownstone, *Collaborative*

instances, a tipping point can be said to have been reached where the game theoretical aspects of litigation practice, dictating what is in one's self-interest, have necessarily changed. Without greater cooperation among adversaries,⁶⁶ parties are doomed to any number of defeating consequences, not the least of which will be a real or perceived information "gap" in ferreting out evidence.

[28] Some history is in order – the Federal Rules of Civil Procedure have never expressly imposed an affirmative duty to cooperate within what is otherwise an adversarial process.⁶⁷ Nevertheless, the discovery rules crafted in 1938 and still basically in force, were a reaction to the pre-1938 system in which an attorney relied primarily on her opponent's pleading for discovery, without much information otherwise being disclosed.⁶⁸

Navigation of the Stormy e-Discovery Seas, 10 RICH. J.L. & TECH. 53 (2004), <http://law.richmond.edu/jolt/v10i5/article53.pdf>; Kenneth J. Withers, *Computer-Based Discovery in Federal Court Litigation*, 2000 FED. CTS. L. REV. 2. See also Richard H. Agins, *An Argument for Expanding the Application of Rule 53(b) to Facilitate Reference of the Special Master in Electronic Data Discovery*, 23 PACE L. REV. 689, 693 (2003) ("Although the Rules of Procedure characterize discovery as a cooperative process, it is in fact highly adversarial and presents substantial opportunity for overreaching and abuse. As the volume of data increases, and as retrieval of that data becomes easier, the potential for abuse increases correspondingly.").

⁶⁶ Of course, greater cooperation is also necessary among all individuals aligned on the same side of the "v.," including counsel of record, in-house counsel, IT staff, and all key players and material witnesses within corporations and other institutions subject to suit. See Gregory D. Shelton, *Don't Let the Terabyte You: New E-Discovery Amendments to the Federal Rules of Civil Procedure*, 73 DEF. COUNSEL J. 324, 325 (2006) ("Staying ahead of the terabytes of information that are accumulating in companies' computer systems and managing that information for discovery will require a great deal [of] cooperation and communication between companies and their counsel."). This is true, in part, because information systems are so complex and emergent, that no one person at an enterprise can usually explain them. See PAUL & NEARON, *supra* note 2, at 6.

⁶⁷ John S. Beckerman, *Confronting Civil Discovery's Fatal Flaws*, 84 MINN. L. REV. 505, 530 (2000) (suggesting that the drafters "took for granted, or at least did not explicitly address, a lawyer's duty to cooperate in the discovery process.").

⁶⁸ *Hickman v. Taylor*, 329 U.S. 495, 500 (1947) (describing the pre-1938 discovery system and noting that the new system is one of "significant innovations"). See also Griffin B. Bell et al., *Automatic Disclosure in Discovery – The Rush to Reform*, 27 GA. L. REV. 1, 6 (1992). But cf. Jeffrey J. Mayer, *Prescribing Cooperation: The Mandatory Pretrial Disclosure Requirement of Proposed Rules 26 and 37 of the Federal Rules of Civil Procedure*, 12 REV. LITIG. 77, 86 (1992) (citing *Shaw v. Ohio Edison Installation Co.*, 9 Ohio Reprint 809, 812 (1887) ("There is no objection that I know, why each party should not know the other's case.") (Taft, J.)).

The 1938 switch to notice pleading and liberal discovery was intended to ensure cases would be decided on the merits – by allowing full disclosure of the pertinent facts prior to trial, thereby avoiding unfair surprise.⁶⁹ Thus, as created by the 1938 rules, modern discovery was envisioned as “an essentially cooperative, self-regulating process,” requiring minimal judicial involvement, and allowing for disclosure of all relevant facts pertinent to the case.⁷⁰

[29] But the drafters did not specifically require a duty to cooperate, although a “failure to cooperate” could result in sanctions under Rule 37.⁷¹ Instead, the drafters believed that a cooperative, largely self-run and self-monitored model of fact discovery would be sustained through “self-interested reciprocity.”⁷² Notwithstanding the inherent tension in grafting notions of cooperation onto the discovery process – within an adversarial paradigm in which lawyers have an obligation to zealously represent their clients,⁷³ in general, the discovery process worked adequately for at least the first 30 years.⁷⁴

⁶⁹ *United States v. Procter & Gamble*, 356 U.S. 677, 682 (1958) (stating that modern discovery practices “make a trial less a game of blind man’s buff and more a fair contest with the basic rules and facts disclosed to the fullest practicable extent”); *Hickman*, 239 U.S. at 500, 507 (discussing the trouble with the pleading system and noting that “[m]utual knowledge of all the relevant facts gathered by both parties is essential to proper litigation.”). *See also* *Cine Forty-Second St. Theatre Corp. v. Allied Artists Pictures Corp.*, 602 F.2d 1062, 1063 (2d Cir. 1979) (stating that modern discovery rules brought principles of equity to discovery “in the hope of expediting the litigation process and of transforming the sporting trial-by-surprise into a more reasoned search for truth.”); Bell, *supra* note 68, at 7 (suggesting that one premise on which modern discovery is based is that more information is better, and that the pre-1938 system was inadequate); Mayer, *supra* note 68, at 82 (stating that disclosure of facts makes trials fair and promotes efficiency).

⁷⁰ Beckerman, *supra* note 67, at 513.

⁷¹ *Id.* at 554.

⁷² *Id.* at 515-16 (discussing the similarity between the drafter’s ideal and the predictive model of discovery posited by game theorists); Bell, *supra* note 68, at 7-8 (noting that one premise on which modern discovery is based is that mutual self-interest will guide lawyers and discovery will require little regulation).

⁷³ Beckerman, *supra* note 67, at 517.

⁷⁴ William W. Schwartz, *The Federal Rules, the Adversarial Process, and Discovery Reform*, 50 U. PITT. L. REV. 703, 704 (1989) (explaining the results of a 1968 study showing that the discovery rules did not result in abuse or harassment, and noting further steps taken to reduce court intervention in the discovery process).

[30] Changes in the litigation environment in the 1970s and 1980s created new problems in the discovery process that the drafters of the 1938 rules could not have foreseen.⁷⁵ Of several contributing factors, certain new technologies expanded the scope and volume of discovery.⁷⁶ Also significant was the increasing complexity of litigation, burgeoning bar membership, the establishment of mega-firms, and economic forces such as the billable hour. One result of these developments was the use of liberal discovery rules and procedures as a litigation strategy and a “means to an end,” with a corresponding increase in incidents of overuse and abuse.⁷⁷

[31] In response to the problems of abuse and overuse of discovery, the rules were later amended and revised several times in the 1980s, mostly by adding “limits, penalties, sanctions, and admonitions seeking to bully attorneys into cooperative discovery.”⁷⁸ For example, Rule 26(f) was adopted in 1980 to create discovery conferences supervised by judges.⁷⁹ The rules were amended in 1983 “to provide for greater judicial case management and control of discovery, as well as mandatory sanctions against frivolous litigation and discovery abuse.”⁸⁰ This included adding Rule 26(g), which instituted a “signing obligation for discovery requests, responses and objections.”⁸¹ The Advisory Committee Comment to Rule 26(g) suggests that the Rule “imposes ‘an affirmative duty to engage in pretrial discovery in a responsible manner that is consistent with the spirit and purposes’ of the discovery rules.”⁸²

[32] On issues relating to cooperation, perhaps the most significant and controversial rule change prior to 2006 occurred in 1993, when Rule 26 was amended to impose an affirmative duty to disclose, without awaiting

⁷⁵ *Id.* at 704-05 (discussing problems with rules in the 1970s and 1980s and an increase in judicial involvement).

⁷⁶ Beckerman, *supra* note 67, at 518-20 (describing changes in the litigation environment during the period).

⁷⁷ Bell, *supra* note 68, at 11 (discussing results of the transformation of the litigation environment).

⁷⁸ Mayer, *supra* note 68, at 107.

⁷⁹ Beckerman, *supra* note 67, at 531.

⁸⁰ Bell, *supra* note 68, at 14.

⁸¹ Beckerman, *supra* note 67, at 530.

⁸² *Id.* at 531.

formal discovery, certain basic facts needed to prepare for trial or make an informed settlement decision.⁸³ The hope of the Advisory Committee in providing for automatic disclosure was that “the requirement would change the atmosphere in litigation to create a more cooperative attitude and indeed ‘change the culture of adversariness.’”⁸⁴

[33] This background provides context for the historic 2006 changes to the discovery rules, which carve out the newly defined “electronically stored information” as a category of discoverable information, and go on to emphasize the need for greater dialogue and cooperation in connection with e-discovery obligations, including relating to search tasks. Under Rule 26(f), parties must sit down together at an early “meet and confer” conference to discuss a range of issues involving electronically stored information.⁸⁵ Such a conference is intended to be broad in scope and to cover the gamut of preservation, scope, formatting, and accessibility issues.⁸⁶

[34] If the parties cannot reach consensus, courts will not be reticent in imposing their own solutions, either on motion or *sua sponte*. For example, in *Treppel v. Biovail*, plaintiff refused to cooperate with defendant’s suggestion that the parties enter into a stipulation defining the scope of their e-discovery obligations, including identifying sources of information and agreeing on a set of search terms.⁸⁷ Stating that plaintiff’s refusal was a “missed opportunity,” but that defendant nevertheless had the burden of responding, the court went on to require the use of certain search terms as an “interim step” in discovery.⁸⁸ Other recent cases of court-ordered “search protocols” are emerging.⁸⁹

⁸³ *Id.* at 515. See Richard Marcus, *Discovery Containment Redux*, 39 B.C. L. REV. 747, 764-69 (1998).

⁸⁴ Bell, *supra* note 68, at 40 (citing minutes from Advisory Committee meeting). See also Mayer, *supra* note 68; John J. Carroll, *Developments in the Law of Electronic Discovery*, 27 AM. J. TRIAL ADVOC. 357, 360 (2003) (noting “the cooperation theme of the 1993 amendments”).

⁸⁵ See FED. R. CIV. P. 26(f) (COMM. NOTE).

⁸⁶ *Id.*

⁸⁷ *Treppel v. Biovail Corp.*, 233 F.R.D. 363, 368-69 (S.D.N.Y. 2006).

⁸⁸ *Id.* at 374-75. One commentator, writing in 2006, said that “[t]his opinion, which probably strikes the reader as matter-of-fact, sensible, and routine, would have been extraordinary a scant six years ago,” when e-discovery disputes were handled as simple extensions of conventional paper discovery. Kenneth J. Withers, *Electronically Stored*

[35] For complex cases involving vast amounts of information, the new federal rules mandate a change in the practice of law. Clearly, parties will need to act in a more sophisticated and transparent fashion to disclose electronically stored information in their possession. In response to discovery requests, they will need to propose search strategies to be negotiated with opposing counsel, involving the scope of information to be searched, and the method utilized. As discussed below, new methods of search and retrieval will inevitably enter into the discussion. These may be necessary, but not sufficient; a new way of thinking about the process of discovery is in order.

B. EMERGING METHODS OF “SEARCH AND RETRIEVAL” OF INFORMATION

[36] Lawyers need to rethink how they perform “searches.” This means using computers and not just associates, contract lawyers, or outsourced offshore workers to search databases.

1. BEYOND KEYWORDS

[37] As a soon-to-be published Sedona Conference white paper on search and retrieval explains,⁹⁰ the *status quo* for the legal profession is to use “keywords,”⁹¹ without more, to ferret out electronically-stored information in large corporate and institutional databases. The legal profession has adopted keyword searching in light of its longtime familiarity with its use

Information: The December 2006 Amendments to the Federal Rules of Civil Procedure, 4 NW. J. of TECH. & INTELL. PROP. 171, ¶ 5 (2006), available at <http://www.law.northwestern.edu/journals/njtip/v4/n2/3>.

⁸⁹ See, e.g., *Johnson v. Kraft Foods N. Am., Inc.*, 2006 WL 3302684 (D. Kan. Nov. 14, 2006); *Seer Sys., Inc. v. Beatnik, Inc.*, 2006 WL 1180058 (N.D. Cal. May 3, 2006); *Balboa Threadworks v. Stucky*, 2006 WL 763668 (D. Kan. Mar. 24, 2006) (advising parties to meet and confer on the use of a search protocol, including keywords). See also *Medtronic Sofamor Danek, Inc. v. Michelson*, 229 F.R.D. 550, 552 (W.D. Tenn. 2003) (“Producing electronic data requires, at minimum, several steps: (1) designing and applying a search program to identify potentially relevant electronic files . . .”).

⁹⁰ The Sedona Conference paper, for which the present authors both serve as editors, is expected to be published in the 2007 Sedona Conference Journal. See www.thesedonaconference.org for further information.

⁹¹ For the purpose of this article, we intend the use of the term “keyword searching” to refer to set-based searching using simple words or word combinations, with or without Boolean and related operators, as described *infra*.

in connection with the offerings of the major online legal retrieval services, which allow for searches to be made of structured databases containing case precedent and statutory authority. To a greater or lesser extent, this familiarity also translates into a working knowledge of Boolean operators (“and,” “or,” and “not”);⁹² simple proximity operators (e.g., “w/3,” “w/s,” or “w/p,” meaning within three words, within the sentence, and within the paragraph, respectively); as well as various types of wildcard, truncation and/or stemming devices (e.g., “*” or “!,” standing in for all possible combinations of letters used between letters of a keyword, or after a given letter of a keyword), as a way to focus legal research queries. Lawyers also have some exposure to “natural language” queries to find best-matching case law, although it remains uncertain how widely this is used.⁹³ But lawyers also well understand the enormity of

⁹² A Boolean search is an exact-match engine in that a Boolean search engine will only return documents that exactly match the query, and the documents will be returned in no particular order. . . . If AND is used, then the engine will retrieve only documents which contain every term so joined. Such queries generally return too little. If OR is used, then the search engine will return any and every document which contains any one or more of the so joined terms. Such queries generally return too much. . . . Most such search engines permit proximity searches which enable an experienced researcher to form more complicated queries by stipulating that certain terms must be within a certain distance of each other.

J.C. Smith, *Machine Intelligence and Legal Reasoning*, 73 CHI.-KENT L. REV 277, 334-35 (1998).

⁹³ In best-match search engines, the documents do not have to exactly match the query, but are returned in a ranked order according to their similarity with the query. A best-match search engine that permits one to form the query in the way one normally writes or speaks, is often referred to as a natural language search engine because it permits the use of natural language queries.

Id. at 335. See Natural Language Processing, Wikipedia, http://en.wikipedia.org/wiki/Natural_language_processing (defining natural language processing as a subfield of artificial intelligence and linguistics); Stop Words, Wikipedia, http://en.wikipedia.org/wiki/Stop_words (defining stop words as common words filtered out as noise from natural language queries). Professor Paul Thompson at Dartmouth has noted that since the 1990s, users of the major online retrieval services have been given the choice of doing Boolean or ranked retrieval (i.e., natural language) queries, but that companies have found that the vast majority of their users have preferred to stay with Boolean retrieval. Paul Thompson, *Looking Back: On Relevance, Probabilistic Indexing and Information Retrieval*, INFO. PROCESSING AND MGMT. J. (submitted for publication).

the task in finding all relevant case law even in well-known, structured databases – a subject which has been explored in past research.⁹⁴ In contrast, the problems faced by the practitioner in e-discovery contexts are even more daunting by several orders of magnitude, and for many and varied reasons.

[38] First, and most importantly, there are profound issues of ambiguity and indeterminacy in human language, and thus in all texts in large, heterogeneous databases subject to discovery. Ludwig Wittgenstein pointed this out in mid-twentieth century in his *Philosophical Investigations*, by noting that words are living, elastic aspects of human behavior subject to constant change and only have meaning in their use. In short, language is a “form of life.”⁹⁵ Others have catalogued types of indeterminacy arising from this truth.⁹⁶ Thus, it is not surprising that lawyers and those to whom they delegate search tasks may not be particularly good at ferreting out responsive information through the use of simple keyword search terms.⁹⁷ Furthermore, people make up words on

⁹⁴ “While it is now possible to store enormous amounts of reported decisions in electronic databases, the retrieval of relevant cases remains an extremely difficult task.” Smith, *supra* note 92, at 333 (providing references at note 112 to what is described as a “major field of research”); Howard Turtle, *Natural Language vs. Boolean Query Evaluation: A Comparison of Retrieval Performance*, in PROCEEDINGS OF THE 17TH ANNUAL INT’L ACM SIGIR CONF. ON RESEARCH AND DEV. IN INFO. RETRIEVAL 212-22 (W. Bruce Croft & C.J. van Rijsbergen eds., 1994).

⁹⁵ [T]o imagine a language means to imagine a form of life . . . There are countless kinds [of sentences]; countless different kinds of use[s] of what we call “symbols,” “words,” “sentences.” And this multiplicity is not something fixed, given once and for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten.

WITTGENSTEIN, *supra* note 30, at §§ 19, 23.

⁹⁶ See BLAIR, *supra* note 6, at 294-301.

⁹⁷ See Eric Goldman, *Deregulating Relevancy in Internet Trademark Law*, 54 EMORY L. J. 507 (2005). In the context of conducting Internet searches for purposes such as determining trademark infringement,

[U]nfortunately, searchers do a poor job selecting keywords. Searchers with domain expertise on a topic generally do a better job selecting keywords, but because searchers routinely have low domain expertise, searches routinely choose keywords poorly. Specifically, most searches use no more than two keywords in a keyword search, and searchers almost never use advanced search methodologies like

the fly, including new codes that function as language. People in different parts of the country, in different parts of an organization, or in different age groups devise their own private languages for the context of their then current environment. For example, what does POS mean?⁹⁸ What is 1337?⁹⁹

[39] Next, searching by any means, including keywords, may be “fraught with technical difficulties,” especially where the physical location of data on tape or disk makes the search task impossible or impractical.¹⁰⁰ There is also a well-known error rate in scanning using optical character recognition (OCR) technology,¹⁰¹ as well as a less understood, but universal error rate associated with simple misspellings as a matter of human input.¹⁰² Thus, at a minimum, in searches of larger data collections, some consideration should be given to employing “fuzzy” search logic in connection with any search technique to be employed, including keywords.

[40] Accordingly, the assumption on the part of lawyers that *any* form of present-day search methodology will fully find “all” or “nearly all” available documents in a large, heterogeneous collection of data is wrong in the extreme. A leading study by Blair & Maron, where the legal teams only found 20% of the responsive documents in a large subway crash case,

Boolean logic or advanced searching functionality offered by search providers.

Id. at 515-16.

⁹⁸ Does “POS” mean “point of sale,” or “parent over shoulder” of someone typing on a keyboard? It all depends on what the “language game” of the moment intends POS to mean.

⁹⁹ Computer gamers and hackers have invented, and widely use, a new dialect called “leet speak” that is incomprehensible to those outside the relevant community. This dialect has taken hold in English, Chinese, and Russian, among other languages. 1337 means “leet” in leet, which is short for “elite.” See Leet, Wikipedia, <http://en.wikipedia.org/wiki/Leet>.

¹⁰⁰ Sasha K. Danna, Comment, *The Impact of Electronic Discovery on Privilege and the Applicability of the Electronic Communications Privacy Act*, 38 LOY. L.A. L. REV. 1683, 1719 (2005).

¹⁰¹ See Baron, *supra* note 50, at 241; Ross W. Kodner & Dale W. Cottam, *In Search of the Holy Grail: The Paperless Office*, 29 WYO. LAWYER 18, 19 (Aug. 2006); Shannon M. Curreri, Comment, *Defining “Document” in the Digital Landscape of Electronic Discovery*, 38 LOY. L.A. L. REV. 1541, 1576-77, n.169 (2005).

¹⁰² See BLAIR, *supra* note 6, at 304 (discussing misspellings).

has been widely cited as recognizing this inherent problem.¹⁰³ Others have recognized the problem for lawyers as well, especially in the age of the Web.¹⁰⁴

¹⁰³ The Blair & Maron study involved a manual review of 350,000 pages (40,000 documents) of online text captured in an IBM STAIRS (Storage and Information Retrieval System) database, for the purpose of finding responsive documents with particular content. As noted, *supra* note 8, the lawyers in the study greatly overestimated the effectiveness of their retrieval effort at finding relevant documents based on keyword searches. David Blair further describes the indeterminacy problem involved in the study at some length:

In the legal case in question, one concern of the lawyers was an accident that had occurred and was an object of litigation. The lawyers wanted all the reports, correspondence, memoranda, and minutes of meetings that discussed this accident. Formal queries were constructed that contained the word 'accident' along with the names of the [city] where it occurred. In the search for *unretrieved* relevant documents, the experimenters later found that the accident was not always referred to as an 'accident,' but as an 'event,' 'incident,' 'situation,' 'problem,' or 'difficulty,' often without mentioning the relevant proper name – the name of the city in which it occurred. The manner in which an individual referred to the accident was frequently dependent on his or her point of view. Those who discussed the event in a critical or accusatory way referred to it quite directly – as an 'accident.' Those who were personally involved in the event, and perhaps culpable, tended to refer to it euphemistically as, *inter alia*, an 'unfortunate situation,' or a 'difficulty.' Sometimes the accident was referred to obliquely as 'the subject of your last letter,' 'what happened last week was . . .,' or, as in the opening lines of the minutes of a meeting discussing the issue, 'Mr. A: We all know why we're here' [the words 'accident' and the name of the city were not used at any time in the meeting either]. Sometimes relevant documents dealt with the problem by mentioning only the technical aspects of why the accident occurred, but neither the accident itself no[r] the people or place involved. Finally, much relevant information discussed [contributing factors in] the situation *prior* to the accident and, naturally, contained no reference to the accident itself.

BLAIR, *supra* note 6, at 303 (quoting DAVID BLAIR, LANGUAGE AND REPRESENTATION IN INFORMATION RETRIEVAL 101 (1990)).

¹⁰⁴ See, e.g., Erich Schweighofer, *The Revolution in Legal Information Retrieval or: The Empire Strikes Back*, J. INFO. L. & TECH. 1 (1999), available at http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1999_1/schweighofer/ (providing overview and bibliography of legal informatics research).

2. NEW “INFORMATION CONCEPTS” IN THE PRACTICE OF LAW

[41] Arguably, at least a partial solution to the search problem in litigation, including possible over-reliance on techniques focused solely on keywords, will be found by embracing alternative forms of search methods, tools, and techniques. As a first step; however, lawyers need to be more comfortable with a set of technical concepts and methods drawn from the fields of computer and information retrieval science, so as to be in the position to reliably evaluate or “benchmark” competing solutions.¹⁰⁵ Such basic terms include “recall,”¹⁰⁶ which is a measure of *completeness* – namely, how well one has done in retrieving *all* of the potentially responsive documents from a candidate universe; and “precision,”¹⁰⁷ which is a measure of *efficiency* – namely, how well one has done in retrieving responsive documents as a percentage of the total number of documents retrieved, including all “false positives.”

[42] Next, lawyers must have a better understanding of information retrieval science, and the ways in which information can be searched. Even before the emergence of the Web, information retrieval science has constituted a vast and growing field, any full-scale discussion of which is beyond the scope of this article.¹⁰⁸ However, broadly speaking, information retrieval methods fall into three broad classes: *set-theoretic* (Boolean strings, supplemented by fuzzy search capabilities), *algebraic*

¹⁰⁵ See Baron, *supra* note 50, at 238.

¹⁰⁶ “Recall” is defined as the fraction x/y , where x = the number of relevant documents retrieved and y = the number of relevant documents in the overall collection. See Information Retrieval, Wikipedia, http://en.wikipedia.org/wiki/Information_retrieval (last visited March 6, 2007).

¹⁰⁷ “Precision” is defined as the fraction x/z , where x = the number of relevant documents retrieved, and z = the number of documents retrieved, including both relevant and not relevant documents. See *id.*

¹⁰⁸ See generally RICARDO BAEZA-YATES & BERTHIER RIBEIRO-NETO, MODERN INFORMATION RETRIEVAL (1999); READINGS IN INFORMATION RETRIEVAL (Karen Sparck Jones & Peter Willett eds., 1997); GERARD SALTON, AUTOMATIC INFORMATION ORGANIZATION AND RETRIEVAL (1968); G. SALTON & MICHAEL J. MCGILL, INTRODUCTION TO MODERN INFORMATION RETRIEVAL (1983); C.J. VAN RIJSBERGEN, INFORMATION RETRIEVAL (2d ed. 1979). See also Bates’ Bibliography of Works on Information Seeking, Indexing, and Information Retrieval System Design, <http://www.gseis.ucla.edu/faculty/bates/bibliography.html> (last visited March 6, 2007) (providing bibliography of references).

(premised on the mathematical idea that the meaning of a document can be derived from the constituent terms in a document, and thus weighting retrieval by the proximity of a document's terms in the form of two or higher dimensional maps, as in vector space modeling),¹⁰⁹ and *probabilistic* (using language models¹¹⁰ and Bayesian belief networks,¹¹¹ the latter of which involves making educated inferences about the relevance of future documents based on prior experience in reviewing documents in a given collection).¹¹²

[43] In thinking about retrieval problems, one can also supplement all of these methods by focusing on the language used by the creators of the records, which will include using *taxonomies*¹¹³ and *ontologies*,¹¹⁴ essentially synonyms of words and relevant classes of related words to be developed and built in at the front end of a search process to better refine the search, and to maximize both recall and precision. In contrast to strict set-based Boolean techniques, the above algebraic and probabilistic categories of search methods are often broadly termed under various forms of the heading "concept searching." All three of these search method categories ultimately involve words and word strings contained in source texts. Still, other emerging information retrieval strategies go the next

¹⁰⁹ According to Wikipedia, a "vector space model (or *term vector model*) is an algebraic model used for information filtering, information retrieval, indexing, and relevancy rankings. It represents natural language documents (or any objects, in general) in a formal manner by the use of vectors (of identifiers, such as, for example, index terms) in a multi-dimensional linear space." See Vector Space Model, Wikipedia, http://en.wikipedia.org/wiki/Vector_space_model (last visited March 6, 2007).

¹¹⁰ See Language Model, Wikipedia, http://en.wikipedia.org/wiki/Language_modeling (last visited March 6, 2007).

¹¹¹ Generally speaking, Bayesian belief networks are models which calculate conditional probabilities from combinations of observed events and prior probabilities. For a good explanation of Bayesian networks, see an application of Bayes' theorem at <http://www-128.ibm.com/developerworks/library/wi-robot15/> (last visited March 6, 2007).

¹¹² See generally Information Retrieval, *supra* note 106.

¹¹³ "Taxonomy" is the practice and science of classification. "Almost anything, animate objects, inanimate objects, places, concepts, and events, may be classified according to some taxonomic scheme." See Taxonomy, Wikipedia, <http://en.wikipedia.org/wiki/Taxonomy> (last visited March 6, 2007).

¹¹⁴ An ontology is "a *data model* that represents a set of concepts within a *domain* and is used to *reason* about the objects within that domain" and the relations between them. Ontology, Wikipedia, [http://en.wikipedia.org/wiki/Ontology_\(computer_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science)) (last visited March 6, 2007) (emphasis added).

step of mining various forms of metadata, including audit trail information with respect to the original creators and recipients of documents, as well as the relationships between users in a given corporate entity, all in aid of the quest of determining potential relevance.¹¹⁵

[44] Many hybrid forms of proprietary software, which utilize search methods borrowed from Boolean methods, concept searching, and beyond, are available in the legal tech marketplace.¹¹⁶ Whatever form of new technology is used, however, lawyers will need to pay attention on the front-end of any discovery process to mapping out how best to approach the unique discovery problem at hand by evaluation of hardware, software, and the scope or intellectual content of the search problem in the context of the particular case.¹¹⁷

[45] Just as there was no explicit reference in case law to the recently emergent, important concept of “metadata” prior to 1998,¹¹⁸ as of early

¹¹⁵ Beyond the full use of Boolean, natural language, and “concept search” techniques supplemented with taxonomies and ontologies, a feature-rich set of new information retrieval methods are being discussed in the academic literature and employed in selected real-world contexts, and thus may soon be on the horizon for use in future litigation. Such techniques make exhaustive use of various forms of metadata, and are referred to by various umbrella terms, including social networking analysis, links analysis, visualization techniques, and cognitive information behavior. See, e.g., Diane Kelly, *Implicit Feedback: Using Behavior to Infer Relevance*, in NEW DIRECTIONS IN COGNITIVE INFORMATION RETRIEVAL (A. Spink & C. Cole eds., 2005); Adam Perer et. al., *Using Rhythms of Relationships to Understand Email Archives* (2005) (discussing visualization techniques), available at <http://www.cs.umd.edu/hcil/emailviz/workshop>; Clive Thompson, *supra* note 63 (discussing networking and links analysis).

¹¹⁶ See, e.g., Anne Kershaw, *Automated Document Review Proves Its Reliability*, 5 DIGITAL DISCOVERY & E-EVIDENCE 11 (2005).

¹¹⁷ See forthcoming Sedona Conference paper on search and retrieval. See also Nicolas Economou, *Of Litigators and Butterflies: The Quest for a Quantum Leap in Large-Scale Document Review*, 6 DIGITAL DISCOVERY & E-EVIDENCE 7 (2006). Additionally, there has been some indication of interest in the adoption of “six sigma”-type business improvement methods to more effectively model the discovery process as an aid in litigation, including on search and retrieval issues. See KPMG Forensic Advisory, *Six Sigma in the Legal Department: Obtaining Measurable Quality Improvements in Discovery Management* (2005); Six Sigma, Wikipedia, http://en.wikipedia.org/wiki/Six_Sigma (last visited March 6, 2007).

¹¹⁸ As part of a presentation at the 1998 Cohasset Managing Electronic Records Conference, one of the authors made a survey of the term “metadata” on WESTLAW, finding as of that time no explicit references in any reported case or law review, but

2007 a search of WESTLAW revealed no references in the reported case law to any alternative form of search method.¹¹⁹ Thus, for example, while the concept of “Bayesian” inference has been addressed both as a matter of academic scholarship¹²⁰ and by courts in certain limited evidentiary contexts,¹²¹ real-world applications of Bayesian belief networks used as

nevertheless noting the emergence of the concept as applied in such cases as *Armstrong v. Executive Office of the President*, 1 F.3d 1274,1283 (D.C. Cir. 1993) (holding that the paper and electronic versions of e-mail were mere “kissing cousins” due to the presence of sender and transmission metadata not routinely captured in hardcopy printouts, and thus the electronic as well as the paper versions must be managed under applicable federal records law), and *Public Citizen v. Carlin* 2 F.Supp.2d 1,14 (D.D.C. 1996) (holding that spreadsheets contain formulas that are not printed out), *rev’d on other grounds*, 184 F.3d 900 (D.C. Cir. 1999), *cert. denied*, 529 U.S. 1003 (2000). See Jason R. Baron, *An Emerging Law of Metadata* 198,

http://www.merconference.com/history_sessions.html (on file with author). In the intervening time, over 700 references to “metadata” have appeared just in federal case law and law reviews, in substantial part due to the comprehensive treatment of the subject found in the Technical Appendix (Appendix D) of THE SEDONA GUIDELINES: BEST PRACTICE GUIDELINES & COMMENTARY FOR MANAGING INFORMATION & RECORDS IN THE ELECTRONIC AGE (2004), available at <http://www.thosedonaconference.org/dltForm?did=RetGuide200409.pdf>.

¹¹⁹ A WESTLAW survey undertaken by the authors as of January 15, 2007 revealed no hits for any cases discussing alternative search methods beyond keyword searching in connection with an adjudicated issue in civil discovery.

¹²⁰ See generally DAVID H. KAYE, *What is Bayesianism?*, in PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM 1 (Peter Tillers & Eric D. Green eds., 1988); Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); Eric D. Green, *Symposium, Probability and Inference in the Law of Evidence*, 66 B.U. L. REV. 377 (1986); Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977); Neal C. Stout & Peter A. Valberg, *Bayes’ Law, Sequential Uncertainties, and Evidence of Causation in Toxic Tort Cases*, 38 U. MICH. J. LAW REFORM 781 (2005); Laurence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

¹²¹ See, e.g., *U.S. v. Davis*, 200 F.3d 1053, 1054-55 (7th Cir. 2000); *Jordan v. Riley*, 26 F.Supp.2d 173, 179 (D.D.C. 1998); *U.S. v. Shonubi*, 895 F.Supp. 460, 484 (E.D.N.Y. 1995), *vacated*, 103 F.3d 1085 (2d Cir. 1997), *remanded to*, 962 F. Supp. 370 (E.D.N.Y. 1997); *People v. Nelson*, 48 Cal. Rptr. 3d 399, 416-17 (Cal. App. 3d. 2006). The *Shonubi* case, involving an estimate of the total amount of drugs a smuggler brought into the U.S. using statistical methods, has attracted much interest. See, e.g., Joseph L. Gastwirth et al., *The Shonubi Case as an Example of the Legal System’s Failure to Appreciate Statistical Evidence*, in STATISTICAL SCIENCE IN THE COURTROOM, 405-13 (Joseph L. Gastwirth ed., 2000); Peter Tillers, *Introduction: Three Contributions to Three Important Problems in Evidence Scholarship*, 18 CARDOZO L. REV. 1875, 1879-89 (1997).

searching methodologies have not yet occurred. Against this backdrop, the profession is facing a sea change, as information inflation continues to vex practitioners.

[46] Clearly, there will be a growing demand on the part of both proponents and recipients of document requests for new and better ways of finding responsive information. So too, a need exists for objectively benchmarking the efficacy of alternative or hybrid means of approaching search tasks, as compared to more traditional forms of searching, by keyword or otherwise, in order to assist the trier of fact in determining what is reasonable.¹²² Accordingly, a new jurisprudence on the quality of search methods employed, governing how one must conduct more advanced searches, is likely to evolve. At the same time, the courts will almost always be strongly encouraging litigants to collaborate – to stipulate to search methodologies in advance – so that adjudications over this issue can be avoided. Otherwise, courts will increasingly need to plunge into this area.

3. GREATER USE OF SAMPLING

[47] In addition to using alternative search methods and to understanding search and retrieval concepts, lawyers will wish to make much greater use of statistical sampling techniques for the purpose of best meeting discovery obligations, both on their own initiative and as part of requests made by opposing parties. While court-mandated sampling has occurred prior to 2006,¹²³ newly modified Rule 34(a) has elevated (and legitimized)

¹²² See generally Baron, *supra* note 50. In 2006 a new “legal track” was introduced as part of the 15th Annual Text Retrieval Conference (TREC) sponsored by the National Institute of Standards and Technology (NIST), aimed at further evaluation and study of competing search methodologies as used in a hypothetical setting modeled on real world litigation. The results of the 2006 TREC legal track are expected to be available after March 2007 at <http://trec.nist.gov>, and on the TREC legal track home page at <http://trec-legal.umiacs.umd.edu/>, and will be the subject of additional scholarship.

¹²³ See *McPeck v. Ashcroft*, 202 F.R.D. 31, 34-35 (D.D.C. 2001). In this Title VII action, the magistrate judge ordered limited restoration of certain designated backup tapes as a “test run,” for the purpose of determining whether a further search of backups was justified. Upon further review of the sample obtained, the court failed to order further restoration. *McPeck v. Ashcroft*, 212 F.R.D. 33, 37 (D.D.C. 2003). See generally THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION 44, Rule 11, Comment 11.b (2005),

“sampling” to the same status as inspection, copying, and testing.¹²⁴ The Advisory Committee’s notes to the 2006 Amendments further add that:

Rule 34(a)(1) is also amended to make clear that parties may request an opportunity to test or sample materials sought under the rule in addition to inspecting and copying them. That opportunity may be important for both electronically stored information and hard-copy materials. The current rule is not clear that such testing or sampling is authorized; the amendment expressly permits it. As with any other form of discovery, issues of burden and intrusiveness raised by requests to test or sample can be addressed under Rules 26(b)(2) and 26(c).¹²⁵

[48] In *Zurich Am. Ins. Co. v. Ace Am. Reinsurance Co.*, plaintiff filed a motion to compel defendant to produce certain claim information from its information system, which the defendant opposed on grounds of burdensomeness.¹²⁶ The court recognized that the “volume of data accumulated” by defendant made a “search of its entire database infeasible.”¹²⁷ Nevertheless, in finding that “a sophisticated reinsurer that operates a multimillion dollar business is entitled to little sympathy for utilizing an opaque data storage system, particularly when, by the nature of its business, it can reasonably anticipate frequent litigation,” the court ordered that the parties “propose a protocol for sampling” defendant’s

available at http://www.thesezonaconference.org/dltForm?did=RFP_plus_july05ver.pdf (encouraging parties to use electronic tools and processes such as data sampling).

¹²⁴ Federal Rule of Civil Procedure 34(a) provides that any party may serve on any other party a request,

[T]o inspect, copy, test or *sample* any designated documents or electronically stored information – including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations stored in any medium from which information can be obtained – translated, if necessary, by the respondent into reasonably usable form, or to inspect, copy, test, or *sample* any designated tangible things

FED. R. CIV. P. 34(a) (emphasis added).

¹²⁵ FED. R. CIV. P. 34 (COMM. NOTE).

¹²⁶ *Zurich Am. Ins. Co. v. Ace Am. Reinsurance Co.*, No. 05 Civ. 9170 RMB JCF, 2006 WL 3771090 (S.D.N.Y. Dec. 22, 2006).

¹²⁷ *Id.* at *2.

claim files, to obtain examples of claims files in which issues of the allocation of policy limits has been addressed.¹²⁸

[49] While the use of various statistical techniques as proof of evidentiary propositions may remain subject to some degree of controversy, the blessing given to sampling embodied within Rule 34(a) should encourage more sophisticated approaches to satisfying e-discovery obligations, especially when real individuals in a given litigation setting are faced with enormous volumes of evidence.¹²⁹ For a recipient of large production requests, the use of sampling techniques may prove useful in showing burden, overbreadth, or unreasonableness, especially where samples reveal no responsive documents. Sampling may also, as discussed below, be a prime technique in encouraging more structured dialogue and cooperation among party litigants.

4. “VIRTUOUS CYCLE” ITERATIVE FEEDBACK LOOPS

[50] Greater use of sampling necessarily implicates another emergent phenomenon under the new Rules: the need for more structured, iterative complexity built into the Rule 26(f) “meet and confer” negotiation cycle.¹³⁰ Whereas parties may have, to date, approached their “meet and confer” obligation under Rule 26(f) in a fashion equivalent to hosting a one-of-a-kind, global “summit meeting,” change is in the air: inefficiencies in processing electronically stored information for purposes of responding to discovery requests fairly dictate new, more structured approaches to the meet and confer obligation.

[51] Thus, in response to the problem of searching large data sets, one can expect “virtuous cycles”¹³¹ in the form of iterative feedback loops where

¹²⁸ *Id.*

¹²⁹ See *supra* notes 115-116 and accompanying text.

¹³⁰ See FED. R. CIV. P. 26(f).

¹³¹ See Virtuous Cycle and Vicious Cycle, Wikipedia, http://en.wikipedia.org/wiki/Virtuous_circle_and_vicious_circle (describing the assumption in the field of economics that “a complex system of determinants will tend to lead to a state of equilibrium,” and introducing the notion of “virtuous” and “vicious” cycles to describe when an unstable pattern of events emerges) (last visited March 6, 2007). “Both systems of events have feedback loops in which each iteration of the cycle

multiple, iterative meet and confer sessions occur for information exchange and discussion of issues to research, negotiate, and agree. This should be handled in the first instance by parties without need for judicial intervention.¹³² Following on Judge Scheindlin's recommendations in *Zubulake V*, the authors here suggest the following "archetype scenario" for iterative discussion of search and retrieval and of obligations of preservation and access generally, in a case of roughly similarly situated parties each with large volumes of electronically stored information:

[52] **Step 1:** The parties meet and confer on the nature of each others' computer hardware and software applications. Proposals are exchanged on the scope of search obligations, in terms of databases and applications to be searched, what active and possibly legacy media are to be made subject to search, and any limitations on scope keyed to particular individuals within an institution, particular time periods, or other ways to limit the scope of the search obligation. Keywords are proposed as a basis for conducting searches, with attention paid to negotiating appropriate Boolean strings of terms, with a full range of proximity operators, wildcard, truncation and stemming terms (to the extent any or all such techniques can be utilized). Alternative concept-based search methodologies are discussed, to the extent either party has experience in using and has found to be efficacious in finding documents. A timetable is agreed upon for conducting initial searches.

[53] **Step 2.** In the interval between meet and confers, parties conduct searches in accordance with the representations made at the initial meet and confer. Based on sampling techniques or other methods employed, estimates are gathered on the volume of data potentially to be made

reinforces the first (positive feedback). The difference between the two is that a virtuous cycle has favorable results" *Id.*

¹³² Judge Scheindlin anticipated adoption of such a strategy where she stated in the *Zubulake* case that "[i]t might be advisable to solicit a list of search terms from the opposing party . . . so that it could not later complain about which terms were used." *Zubulake v. UBS Warburg LLC*, 229 F.R.D. 422, 432 at n.75 (S.D.N.Y. 2004). The opinion in *Zubulake V* further advises that counsel might itself run a broad list of search terms to preserve what may be an overbroad number of documents, and thereafter negotiate a more restrictive set of search terms for the purpose of reviewing documents in response to actually propounded discovery. *See id.* at 432.

subject to search in light of the wording of opposing parties' search requests.

[54] **Step 3.** Returning to the meet and confer table, the parties describe how initial searches were conducted and what are the preliminary results. Based on a finding that either too few or too many files were retrieved corresponding to particular specific requests, search protocols are adjusted accordingly for a second round of searching. If some form of open discovery measures are agreed to, an exchange of actual documents found as the result of the initial searches takes place at this juncture, so as to provide the opposing party with the opportunity to essentially request "more like this" (or not).¹³³ Even, however, absent fully open discovery, more limited reporting is made of search results, in order to narrow or expand search requests as appropriate.

[55] **Step 4.** The process continues in iterative fashion as agreed to by the parties, until a mutually agreed time, or a mutually agreed cap on numbers of responsive documents is reached.

[56] As is well known in the field of economics, the art of war, and elsewhere, cooperative behavior can be encouraged and will rationally arise within an otherwise adversarial paradigm (including outright state of hostility between parties), where a continuing relationship exists and there is a modicum of goodwill existing in the form of trust.¹³⁴ No reason exists

¹³³ The envisioned process is one application of the idea of "relevance feedback," a well-known concept in the information retrieval field. "Traditional relevance methods require that users explicitly give feedback by, for example, specifying keywords, selecting and marking documents, or answering questions about their interests. Such relevance feedback methods force users to engage in additional activities beyond their normal searching behavior." Diane Kelly & Jaime Teevan, *Implicit Feedback for Inferring User Preference: A Bibliography*, SIGIR FORUM 18, 18(2003), available at <http://www.ils.unc.edu/~dianek/kelly-sigir-forum03.pdf>.

¹³⁴ See ROBERT AXELROD, *THE EVOLUTION OF COOPERATION: AGENT-BASED MODELS OF COMPETITION AND COLLABORATION* (1997). Axelrod's contributions to game theory have included identification of "TIT FOR TAT" as a dominant, stable meta-strategy for cooperation among adversaries over a multiple-stage, Prisoner's Dilemma type game. Robert Axelrod, *The Emergence of Cooperation Among Egoists*, 75 AM. POL. SCI. REV. 306 (1981). As explained in John Setear's law review article analyzing Axelrod's theories in the context of civil discovery abuse:

In considering the dynamics of the Prisoner's Dilemma, Axelrod begins with the implicit assumption that, whether through conscious choice or genetic programming, individuals choose their strategy in a particular move of the multi-move game through the use of what is known as a 'meta-strategy.' This meta-strategy does not vary from move to move, though the strategies that the meta-strategy determines for each move can change. Axelrod expressly assumes that each member of a community can identify all other members of the community and can perfectly remember her previous interactions with them.

A meta-strategy generally chooses a strategy for a particular move by applying its rules to the history of interactions between the two parties to that move. For example, a meta-strategy might simply specify that the choosing individual will cooperate if the comrade she encounters cooperated in either of their previous two interactions, and will otherwise defect. Alternatively, a meta-strategy might specify that the choosing individual cooperate if the current opponent has cooperated more frequently in the past than it has defected. A meta-strategy can employ a random element, even to the extent of ignoring past interactions completely and simply choosing randomly whether to cooperate or defect.

John Setear, *The Barrister and the Bomb: The Dynamics of Cooperation, Nuclear Deterrence, and Discovery Abuse*, 69 B.U. L. REV. 569, 594-95 (1989). As further explained, TIT FOR TAT won all rounds of a round-robin tournament, where its meta-strategy consisted of (i) always cooperating the first time it encountered a given comrade, (ii) cooperating in the current play within a round if the comrade had cooperated in their previous encounter; and (iii) defecting in the current play if the comrade had defected in their previous encounter. *Id.* "In other words, TIT FOR TAT offered its cheek for the first encounter, but otherwise adhered to standards of Old-Testament justice." *Id.*

As applied in the area of meet and confer obligations, arguably cooperative strategies between two or more parties, in the form of providing greater openness and transparency from the beginning of the meet and confer process, and continuing in subsequent rounds (or iterations), ideally will end up optimizing overall effectiveness, as measured by the richness of the document set to be identified (parties maximizing recall), as well as decreasing the amount of noise or false positive documents to be looked through (maximizing precision). "Forced" cooperation through timely judicial intervention may also be employed with great beneficial effect. Of course, no litigation ever goes perfectly smoothly, and non-cooperation which is otherwise tolerable within the framework of the rules should be anticipated (hence, a litigator's resort to the flip-side of TIT FOR TAT, at least for one round).

not to similarly employ models of cooperation in attempting to narrow the search task in light of information inflation.¹³⁵

III. EVOLUTION IN THE LAW OF PRIVILEGE

[57] Besides new collaborative models and new search and retrieval methodologies, the law of privilege must also adapt as a consequence of information inflation. Primarily at issue is the expense imposed on litigants by privilege review, which now almost always dictates a manual process because of “death penalty” waiver doctrines that evolved long ago when information was still manageable. The huge numbers of files which must be reviewed will almost always mean that privileged information inadvertently will be disclosed, even absent negligence by the producing party. This argues for an evolution of privilege law regarding inadvertent disclosures. There simply is too much information now, for old standards of inadvertent waiver to apply.

[58] One of the essential aspects of the 2006 e-discovery rules amendments is the lack of any change in substantive privilege waiver law.¹³⁶ Because of the perceived quantity of information attendant to e-discovery, the drafters of the new rules, as well as those who testified and commented, were keenly aware of the need for protection against privilege waiver. However, because waiver is a substantive legal issue, the Federal

¹³⁵ Alternatively, there is always the possible remedy of cost shifting, which has come into vogue regarding searching information that is “not reasonably accessible” because of the complexities of modern information systems. See FED. R. CIV. P. 26(b)(2)(B). See generally *Zubulake v. UBS Warburg LLC* (Zubulake I), 217 F.R.D. 309, 324 (S.D.N.Y. 2003) (allowing for cost shifting in case of inaccessible data). One possible scenario to a Rule 26(f) meet and confer request to search through a billion e-mails is to allow the search, but at the expense of the searcher. In other words, “if you wish to search through my billion documents, ‘Make my day’ – you can pay for it and spend the time doing so.” Courts already have this authority under the proportionality principle, now found in Rule 26(b)(2)(C).

¹³⁶ See PAUL & NEARON, *supra* note 2 (stating that it is undisputed that the changes to Rule 26(f) involving clawbacks and quickpeek agreements, and to Rule 26(b)(5) involving the new retrieval procedure are procedural only).

Rules of Civil Procedure are powerless to address the issue, which varies from jurisdiction to jurisdiction.¹³⁷

[59] Accordingly, to save on the enormous costs of a review (where missing even one document can trigger a cascade of consequences), an innovation in the substantive law of privilege waiver must occur. One possible solution is for the parties to execute a court-endorsed “inadvertent disclosure agreement” which covenants there will be no privilege waiver by means of inadvertent disclosure, and for such “private law” to control even as to third parties.¹³⁸ Otherwise, the risk of privilege waiver will force parties to continue hugely expensive privilege reviews, or to forego the attorney-client privilege or work-product privilege altogether.

[60] The problem is that: (1) currently, it is often state law which governs privilege waiver issues; and (2) this is often a court-made, common law rule of decision (slow to evolve and dependent on presentation of a discrete dispute).

[61] There are currently two initiatives that might institute a new rule of decision. As of publication of this article, Senator Arlen Specter has introduced the “Attorney-Client Privilege Protection Act of 2007” (ACPPA).¹³⁹ The ACPPA would prohibit the government from forcing organizations into: (1) disclosing information protected by the attorney-client privilege or work product doctrine; (2) refusing to contribute to the legal defense of an employee; (3) refusing to enter into a joint defense

¹³⁷ George L. Paul & Dawn Bergin, “Clawbacks,” “Sneak-Peeks,” and the “Retrieval Procedure” under the New Federal Rules: *The Risks of the New Inadvertent Disclosure Procedure*,” CORP. COUNS., Dec. 2006, http://www.corpcounsel-digital.com/corpcounsel/sample/templates/pageviewer_print?pg.

¹³⁸ Another way to view such an agreement is within the context of a protective order. See *The Sedona Guidelines: Best Practices Addressing Protective Orders, Confidentiality & Public Access in Civil Cases*, 6 SEDONA CONF. J. 183 (2005).

The purpose of the order would be to facilitate the cooperative exchange of voluminous discovery . . . Attorneys should cooperate in efficiently exchanging discovery in civil litigation. Such cooperation includes an early, full discussion of the scope of discovery and the treatment of potentially discoverable materials that the parties deem confidential or private, to avoid later pre-trial litigation of this issue.

Id. at 203-04.

¹³⁹ S. 186, 110th Cong. (2007).

strategy with an employee; (4) refusing to share relevant information with an employee; and (5) terminating or disciplining an employee.¹⁴⁰ However, it is too early to predict whether this newly introduced legislation will be enacted.

[62] Next, there is proposed Federal Rule of Evidence 502. Hearings on this proposed new federal rule governing privilege waiver were recently held in Phoenix, Arizona and New York City. The proposed rule among other things would make court-endorsed “clawback” type agreements enforceable against third parties.¹⁴¹ It is currently unclear how broad the rule might become, or to what extent it would apply to state court actions. The drafters currently intend that it apply only in federal litigation or arbitration caused disclosures, whether later arising in federal or state court.¹⁴² The usefulness of the proposed new rule is limited because disclosures originating in state court actions would not be affected. Nor would the rule apply unless there was an agreement.

[63] Indeed, one can question whether immunity from privilege waiver, through an inadvertent disclosure of information, should be dependent on an agreement at all, since in many instances there will be no agreement, yet the same cost-saving policies regarding lack of punitive privilege waiver rules should apply.

¹⁴⁰ *Id.* Among other commentators, both the ACLU and Edwin Meese III of The Heritage Foundation have commended the bill, primarily as a reaction to the exercise of power by the Executive Branch. Press Release, American Civil Liberties Union, ACLU Welcomes Attorney-Client Privilege Protection Act, Says Bill Would Safeguard Constitutional Right to Counsel (Dec. 7, 2006), <http://www.aclu.org/crimjustice/gen/27637prs20061207.html>; Press Release, Edwin Meese III, The Heritage Foundation, Meese Praises Approach of Attorney-Client Privilege Protection Act (Jan. 7, 2007), <http://www.heritage.org/Press/NewsReleases/nr120706a.cfm>.

¹⁴¹ See *Advisory Committee on Evidence Rules: Hearing on Proposal 502*, (2006), available at http://www.uscourts.gov/rules/EV_Hearing_April_2006.pdf.

¹⁴² The drafters have concluded that they could, constitutionally under Article I commerce clause power, legislate rules in state courts as well because of the effect on interstate commerce. They currently do not intend to do this, however, because of state court objections and reasons of comity. See *id.*

IV. A VIEW TO THE FUTURE

[64] No one knows if, or when, civilization's new inflationary period will end. Are we merely in the early years of an inflation phenomenon? Will the present-day rate of acceleration continue apace, or, as might even be possible, will the rate of acceleration increase? Will things flatten out or level off?

[65] Some computer scientists forecast essentially more of the same, which itself should give pause.¹⁴³ Others imagine science fiction-like futures,¹⁴⁴ where computer power in the form of artificial intelligence has approached or exceeded the capacity of the human mind, and/or has been harnessed by trans-human beings with machinery incorporated into their living circuits.¹⁴⁵

[66] However, whatever may be the limits of machine or artificial intelligence, in the near term future lawyers must not be afraid to embrace creative, technological approaches to grappling with the problem of knowledge management. Emerging solutions to lawyers' search problems

¹⁴³ See Johannes Gehrke, *Monitoring the Data Tsunami*, COMPUTING RES. ASS'N, GRAND RES. CHALLENGES (2002), <http://www.cra.org/Activities/grand.challenges/gehrke.pdf> ("In 2020, we will live in a world that is networked to unprecedented scale, and computer networks have become more pervasive in scope and mission-critical to businesses, scientific endeavors and other computing applications.").

¹⁴⁴ What is science fiction to some, like human flight, can become accepted reality in a few years once the future unfolds.

¹⁴⁵ See RAY KURZWEIL, *THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY* 8 (2005) (arguing that "within several decades information-based technologies will encompass all human knowledge and proficiency, ultimately including the pattern-recognition powers, problem-solving skills, and emotional and moral intelligence of the human brain itself," at which point humans and machines will have approached a "singularity"). *But cf.* Stephanie Schorow, *Hard Drives Will Evolve Into Soft Hearts . . . Or Not*, MIT TECH TALK, Dec. 6, 2006, <http://web.mit.edu/newsoffice/2006/print/creativity-2-print.html> (reporting on debate held at MIT, in which Ray Kurzweil and Professor David Gelernter of Yale participated, to celebrate 70th Anniversary of Alan Turing's 1936 groundbreaking paper, "On Computable Numbers," where Gelernter took an opposing, "anti-cognitivist" viewpoint on the subject of whether machines will ever achieve consciousness). *See generally* Artificial Consciousness, Wikipedia, http://en.wikipedia.org/wiki/Artificial_consciousness (last visited on March 6, 2007) (summarizing debate and providing artificial intelligence literature review).

over the coming decade and beyond could likely include a synthesis of “intelligent search engine” applications culled from the areas of artificial intelligence,¹⁴⁶ neural networks,¹⁴⁷ and other forms of information-filtering and machine-learning techniques.¹⁴⁸ Nor should lawyers discount the possibility of one day employing even more advanced science fiction-sounding search techniques, derived from current research in the fields of nanotechnology, including quantum computing and bioinformatics.¹⁴⁹

¹⁴⁶ THE SEDONA CONFERENCE GLOSSARY: E-DISCOVERY & DIGITAL INFORMATION MANAGEMENT (May 2005), <http://www.thosedonaconference.org/dltForm?did=tsglossarymay05.pdf>. The glossary defines Artificial Intelligence (AI) as,

The subfield of computer science concerned with the concepts and methods of symbolic inference by computer and symbolic knowledge representation for use in making inferences- an attempt to model aspects of human thought on computers. It is also sometimes defined as trying to solve by computer any problem once believed to be solvable only by humans. AI is the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience. It attempts to approximate the results of human reasoning by organizing and manipulating factual and heuristic knowledge. Areas of AI activity include expert systems, natural language understanding, speech recognition, vision, and robotics.

Id. at 2-3.

¹⁴⁷ *Id.* at 30 (defining Neural Networks as “made up of interconnected processing elements called units, which respond in parallel to a set of input signals given to each.”).

¹⁴⁸ See *supra* note 111 and accompanying text. See also PAT LANGLEY, ELEMENTS OF MACHINE LEARNING (1996); Douglas W. Oard, *The State of the Art in Text Filtering*, 7 USER MODELING AND USER-ADAPTED INTERACTION J. 141 (1997) available at <http://www.springerlink.com/content/k040468m41264111/?p=b08b424d424e42839367a3498051919d&pi=0>.

¹⁴⁹ See Drew Harris, *Shrinking the Battlefield: A Review of Nanotechnology and Homeland Security*, 1 NANOTECH. L. & BUS. 116, 121 (2004) (referencing DNA computing, molecular computing, and quantum computing as new computing architectures that could process massive amounts of data (such as a database of faces), or quickly crack encrypted communications); Frank Murowski, *The Market for Nanoelectronics*, 1 NANOTECH. L. & BUS. 364, 365 (2004) (stating that “second generation developments are currently at theoretical or basic research stages and are expected to emerge in the post-2014 timeframe,” where “their anticipated ultra-high capabilities [will be] ten times more powerful than first generation nanoelectronics,” and that “we expect a quantum leap in the electronics market over the next ten years followed by another massive increase in capabilities in the ten to twenty year horizon.”).

[67] Information inflation reflects the fact that civilization has entered a new phase. Human beings are now integrated into reality quite differently than before. They can instantaneously write to millions. They engage in the real time writing of instant messages, wikis, blogs, and avatars. Accordingly, the flux of writing has grown exponentially, with resulting impact on cultural evolution.

[68] All this affects litigation. Vast quantities of new writing forms challenge the legal profession to exercise novel skills. This means litigation must become more collaborative. It means more use of computer technology. It means there will be new legal rules. And the future of litigation as we know it is at risk unless law and its practice co-evolve with information.¹⁵⁰

¹⁵⁰ See CHARLES SEIFE, *DECODING THE UNIVERSE: HOW THE NEW SCIENCE OF INFORMATION IS EXPLAINING EVERYTHING IN THE COSMOS FROM OUR BRAINS TO BLACK HOLES* 262 (2006).

Life too is shaped by information. All living creat[ur]es are information-processing machines at some level; intelligent, conscious creatures are processing that information in their minds as well as in their cells. But the laws of information put limits on the processing of information. There are a finite (if enormous) number of ways information can be arranged in our Hubble bubble, so there are a finite (and smaller, but still enormous) number of ways information can be arranged and processed in our heads. While humans might be able to contemplate infinity, we can only do so in a finite number of ways.

The universe might be infinite, but we are not.

Id. at 262. See generally JOHN ARCHIBALD WHEELER, *AT HOME IN THE UNIVERSE* 298 (1994) (“Tomorrow we will have learned to understand and express all of physics in the language of information.”); HANS CHRISTIAN VON BAEYER, *INFORMATION: THE NEW LANGUAGE OF SCIENCE* (2004).