

# EDDE JOURNAL

A Publication of the E-Discovery and Digital Evidence Committee  
ABA Section of Science & Technology Law

AUTUMN 2010 VOLUME 1 ISSUE 4

## Editor

[Thomas J Shaw, Esq.](#)  
Tokyo, Japan

## Committee Leadership

### Co-Chairs' Message

Co-Chairs:

[George L. Paul, Esq.](#)  
Phoenix, AZ

[Lucy L. Thomson, Esq.](#)  
Alexandria, VA

[Steven W. Teppler, Esq.](#)  
Sarasota, FL

[Eric A. Hibbard](#)  
Santa Clara, CA

[SciTech Homepage](#)

[EDDE Homepage](#)

[Join the EDDE Committee](#)

© 2010 American Bar Association. All rights reserved. Editorial policy: The *EDDE Journal* provides information about current legal and technology developments in e-Discovery, digital evidence and forensics that are of professional interest to the members of the E-Discovery and Digital Evidence Committee of the ABA Section of Science & Technology Law. Material published in the *EDDE Journal* reflects the views of the authors and does not necessarily reflect the position of the ABA, the Section of Science & Technology Law or the editor(s).



## Ten Things That Trouble Judges About E-Discovery

By [Craig Ball](#)

As counselor, consultant or court-appointed special master, my law practice revolves around electronically stored information (ESI)--seeking to salvage the wrecks others have made of e-discovery and helping parties to navigate unfamiliar shoals. The goal is to forestall or resolve conflicts with judges incensed by parties' failure to fulfill e-discovery duties. Judges frequently doubt that electronic discovery is as difficult or expensive as the lawyers before them claim. For the most part, the judges are right. E-discovery is not that hard and need not be so costly. That is, it's not that hard or expensive if counsel knows what he or she is doing, and that's a huge "if." Judges feel lawyers should know how to protect, marshal, search and [Read more](#)

## Managing the Unmanageable: E-Discovery and Electronic Health Information

By [Kenneth Rashbaum](#)

"We have accumulated stupendous know-how . . . Nonetheless, that know-how is often unmanageable . . . And the reason is increasingly evident: the volume and complexity of what we know has exceeded our individual ability to deliver its benefits correctly, safely or reliably." Medical information, as eloquently described above by Dr. Atul Gawande, is increasing at an accelerating rate. Physicians struggle to keep up with the latest developments. Doctors, like so many of us, are barely keeping their heads above water in the information tsunami of 2010. Medical information about patients in the form of electronic health records [Read more](#)

## Going Native – It's Really Not That Hard

By [Sonya Sigler](#)

Why are requesting parties letting producing parties off easy when it comes to producing electronically stored information (ESI)? This article will address several areas where requesting parties can pressure the producing party to do the right thing or in another light - be more demanding, precise, and knowledgeable when making their e-Discovery requests. Requesting parties do themselves a disservice in agreeing to produce data tied to using a keyword list, limiting produced meta data fields to "commonly used" meta data fields, not specifying a form of production (or caving in to a producing parties demands that they just want to produce TIFFs), and believing a producing party's statement that it will take six (6) months or more to review their documents for privilege. [Read more](#)

## E-Discovery Search: The Truth, the Statistical Truth and Nothing but the Statistical Truth

By [Nick Brestoff](#)

This article is a call to revisit Rule 26(g)(1) of the Federal Rules of Civil Procedure, which requires attorneys to certify "to the best of the person's knowledge, information, and belief formed after a reasonable inquiry" that disclosures are "complete and correct." Given the exponentially growing mountain of electronically stored information (ESI) and the incompleteness and statistical nature of search technologies, which this article will explain, no attorney can honestly so "certify." One day, this gap, a loophole between the law of yesterday and the technology of today, will cause a monumental waste of judicial, attorney, and client resources. Most of us know the meaning of a "loophole." [Read more](#)

## Ten Things That Trouble Judges About E-Discovery

By Craig Ball



*As counselor, consultant or court-appointed special master, my law practice revolves around electronically stored information (ESI)--seeking to salvage the wrecks others have made of e-discovery and helping parties to navigate unfamiliar shoals.*

*The goal is to forestall or resolve conflicts with judges incensed by parties' failure to fulfill e-discovery duties. Judges frequently doubt that electronic discovery is as difficult or expensive as the lawyers before them claim. For the most part, the judges are right. E-discovery is not that hard and need not be so costly.*

*That is, it's not that hard or expensive if counsel knows what he or she is doing, and that's a huge "if." Judges feel lawyers should know how to protect, marshal, search and produce the evidence in their cases or enlist co-counsel and experts with that know how. The judges are right about that, too. Lawyers must master modern evidence in the same way that doctors must stay abreast of the latest developments in medicine.*

The challenge to listing ten things that trouble judges about e-discovery is limiting it to *only ten* things. E-discovery exposes much that is not pretty about the state of the law practice, *e.g.*, wasteful, obsolete practices; poor management skills; conflicting interests between lawyers and clients; and unequal access to justice between the rich and the rest. E-discovery didn't create these problems, but like a hard rain on an old roof, it exposes failings too long ignored.

First and most intractable among these problems is:

### 1. Lawyer incompetence

The landscape of litigation has forever changed, and there is no going back to a paper-centric world. Too many lawyers are like farriers after the advent of the automobile, grossly--even stubbornly--unprepared to deal with electronic evidence

As lawyers' duties to supervise and direct clients' preservation and collection of ESI have broadened, their grasp of information systems, forms of ESI and effective search hasn't kept pace. This knowledge gap troubles judges who rely upon lawyers to police the discovery process and stand behind the integrity of that process. Lawyers cannot defend what they don't understand.

No lawyer wants to be thought incompetent; yet the skills developed to collect, assess and produce paper records do not translate well to a world steeped in ESI. Digital is different, and

neither clients nor the justice system can long afford the costly, cumbersome efforts lawyers employ to regress data to paper or images.

Other things that trouble judges about e-discovery are:

## **2. Misstatements of fact coupled with a lack of reliable metrics**

Perhaps because no lawyer wants to be thought incompetent, some resort to “winging it” when it comes to reporting the state of client ESI and status of discovery. The case law proves the folly of blind reliance on clients when gauging the true state of retention and collection. Lawyers must not parrot client claims without undertaking even minimal steps to establish their accuracy.

Often, the misstatements take the form of fanciful claims of burden or cost, advanced sans reliable metrics gained through measurement or testing. Judges expect more than histrionics and hand wringing. They demand competent, quantitative evidence of burden and cost supported by the testimony of knowledgeable people who’ve done their homework. It troubles judges to be asked to decide important issues on much less.

## **3. Cost and waste**

Judges are of one troubled mind about litigation today. They all feel it costs too much and worry that spiraling costs may crowd out legitimate cases or compel unjustified settlements. Recently, a distinguished panel of e-discovery experts surprised this writer by agreeing that about 70% of the money spent on e-discovery is wasted through poor planning and decision-making. Worse, they attributed about 70% of that waste to lawyer incompetence. If true, that suggests that about *half of every dollar spent on e-discovery is wasted* because lawyers don’t know what they’re doing with ESI. *Half!*

## **4. Delay in addressing ESI Issues**

Over time, data tends to morph, migrate and disappear. Employees join and leave, and machines are re-tasked or retired. Memories fade. Active data migrates to tape. Tape moves to warehouses. Old tape formats give way to newer formats, and old tape drives are discarded. With these changes, discoverable information grows more difficult and costly to access over time. It troubles judges when parties ignore ESI issues until little problems grow into big ones.

Judges expect parties and counsel to think and act in timely ways, identifying and preserving potentially responsive evidence when they *anticipate* a claim or lawsuit instead of waiting until a preservation demand surfaces or a lawsuit is filed.

Judges are also troubled when parties or counsel delay getting needed help from experts and vendors. When a lawyer waits until discovery is overdue to begin seeking such help, it's hard for a judge to impute good faith.

## **5. Lack of communication and cooperation**

One reason judges don't like discovery disputes is that they're often so unnecessary; that is, they concern issues the parties could have resolved if they'd simply listened and cooperated. It greatly troubles judges when parties and counsel exert little effort to resolve e-discovery disputes before filing motions and demanding hearings. It further troubles judges when lawyers mistakenly equate candor and cooperation with weakness, seeking to profit from pointless disputes and motion practice.

Judges don't abide trial by ambush or gamesmanship in e-discovery. The bench expects parties to be forthcoming about the volume and nature of discoverable ESI and to be reasonably transparent in, e.g., detailing preservation efforts or disclosing automated search methods. Because judges never forget that all lawyers owe duties to uphold the integrity of the justice system they serve, judges are troubled when advocates let the desire to win eclipse those duties.

## **6. Failing to get the geeks together**

Communication presupposes comprehension, but judges daily confront how working through intermediaries clouds the court's understanding of technical issues. Like lawyers, information technologists employ a language all their own. They speak geek.

Because lawyers rarely know what IT personnel are talking about, lawyers are often fearful of allowing technical personnel from opposing sides to talk to each other. Instead, counsel for the requesting party conveys questions from their technical expert to opposing counsel, who passes them on to in house counsel, who has the paralegal on the case talk to the IT person. The IT person responds to the paralegal who speaks to in house counsel who tells outside counsel who passes on his or her best understanding to opposing counsel or the court. No wonder so much gets misunderstood.

Judges expect clear, accurate communication about technical matters, and it troubles them when knowledgeable people aren't brought together to foster transparency and trust.

## **7. Failing to implement a prompt and effective legal hold**

Preservation is a backstop against error. Slipshod preservation pervades and poisons much of what follows, and the cost to resolve inadequate preservation is breathtakingly more than the cost of a reasonable and timely legal hold effort.

One need only peruse the opus opinions in *The Pension Committee of the University of Montreal Pension Plan, et al. v. Banc of America Securities, et al.*,<sup>1</sup> or *Rimkus v. Cammarata*<sup>2</sup> to appreciate the signal importance judges place on a prompt and effective legal hold of potentially relevant ESI and documents. Lawyers appear to have only two settings when it comes to implementing legal holds: "off" and "crazy." Either they ignore the need for a hold until challenged about missing data, or they issue so vague, paralyzing and impractical a retention directive, that responses run the gamut from doing nothing to pulling the plug and sitting in the dark.

It troubles judges when lawyers and clients fail to preserve information that bears on the issues. Judges rightly expect lawyers to promptly hone in on potentially responsive information when a claim or suit looms. Judges expect lawyers to identify fragile forms of information and take reasonable steps to protect the evidence against loss or corruption due to negligence or guile.

## **8. Overbroad requests and boilerplate objections**

In the bygone era of paper discovery, asking for "any and all documents touching or concerning" a topic was accepted. Information was generally stored on paper, paper was predictably managed and a company's documents were typically organized topically in a few easily-ascertainable locations.

But when information exploded into countless shards of messages and attachments strewn across a sea of accounts, servers, machines, media and devices, "any and all" became too many.

It deeply troubles--even antagonizes--judges when requests for information are unfocused and over-inclusive and when reasonable requests are met with a litany of generic objections. Both demonstrate a lack of care and judgment.

Judges want to see evidence that the discovery sought is proportional to the matters at issue. They expect objections to be asserted in good faith and narrowly drawn. Some judges are even exploring sanctions under Fed. R. Civ. P. 26(g) to address fishing expeditions and boilerplate objections. See, e.g., *Mancia v. Mayflower Textile Servs. Co.*<sup>3</sup>

## **9. Mishandling claims of privilege**

Ask a judge what percentage of documents claimed "privileged" actually prove to be privileged, and you'll probably hear, "ten percent, perhaps less." Yet more than one e-discovery expert has opined that finding, fighting about and redacting privileged documents accounts for a

---

<sup>1</sup> 2010 WL 184312 (S.D.N.Y. Jan. 15, 2010)

<sup>2</sup> 07-cv-00405 (S.D. Tex. Feb. 19, 2010)

<sup>3</sup> 253 F.R.D. 354 (D. Md. 2008)

sizeable share of the money spent on e-discovery. Whatever the percentages, it's clear litigants spend far too much money and time ginning the seeds of privilege from electronic evidence, even while overlooking privileged content through a paucity of quality assurance and control. See, e.g., *Mt. Hawley Ins. Co. v. Felman Prod., Inc.*<sup>4</sup> and *Victor Stanley, Inc. v. Creative Pipe, Inc.*<sup>5</sup>

Lawyers gravitate to error-prone tools, like seat-of-the-pants keyword search, to cull potentially privileged content, mischaracterizing much that's not privileged and much that is. Further, many lawyers forget (or ignore) their client's duty to generate a proper privilege log when material withheld from discovery as privileged happens to be ESI.

Finally, lawyers inexplicably fail to avail themselves of Fed. R. Evid. 502, which provides significant protections against waiver of privilege, including the near-impregnable shield of a R. 502(d) court order.

Last, but not least, any list of things that trouble judges about e-discovery is sure to include:

#### 10. Failing to follow the Rules

Judges value the rules of procedure, and they expect those who come to their courts to do so. So it troubles judges when the rules set forth a clear requirement that's ignored, especially when the failure to follow a rule triggers a superfluous motion and hearing.

A telling example is the Federal Rule of Civil Procedure requiring a producing party to object to a requested form of production and specify the form to be produced.<sup>6</sup> It's a rule observed more in the breach than in compliance; yet adherence to the rule would make many costly battles demanding alternate forms of production unnecessary. The rule sets out what to do--with the goal that conflicts be resolved *before* production in objectionable forms--but litigants just don't do it.

#### Heads in the Sand

Ironically, what most troubles judges about e-discovery also makes their lives easier: judges are astounded they don't see more efforts to discover ESI! The bench well understands that the dearth of e-discovery isn't an indicia of cooperation, but one of evasion. Though all evidence today is largely digital, too many lawyers still try to pretend otherwise and look where they've always looked for

---

<sup>4</sup> 2010 WL 1990555 (S.D. W. Va. May 18, 2010)

<sup>5</sup> 250 F.R.D. 251 (D. Md. 2008)

<sup>6</sup> Fed. R. Civ. P. Rule 34(b)(2)(D): **Responding to a Request for Production of Electronically Stored Information.** The response may state an objection to a requested form for producing electronically stored information. If the responding party objects to a requested form—or if no form was specified in the request—the party must state the form or forms it intends to use.

evidence. Increasingly, judges know this shouldn't be the case and that it can't last. They enjoy the calm, but are troubled that so few are prepared for the gathering storm.

*Craig Ball of Austin is a Board Certified trial lawyer, certified computer forensic examiner and electronic evidence expert. He's dedicated his globetrotting career to teaching the bench and bar about forensic technology and trial tactics. After decades trying lawsuits, Craig now limits his practice to service as a court-appointed special master and consultant in computer forensics and electronic discovery, and to publishing and lecturing on computer forensics, emerging technologies, digital persuasion and electronic discovery. Craig writes the award-winning Ball in Your Court column on electronic discovery for Law Technology News and is the author of numerous articles on e-discovery and computer forensics, many available at [www.craigball.com](http://www.craigball.com). Craig Ball has consulted or served as the Special Master or testifying expert in computer forensics and electronic discovery in some of the most challenging and well-known cases in the U.S.*



## Managing the Unmanageable: E-Discovery and Electronic Health Information

By Kenneth Rashbaum



“We have accumulated stupendous know-how . . . Nonetheless, that know-how is often unmanageable . . . And the reason is increasingly evident: the volume and complexity of what we know has exceeded our individual ability to deliver its benefits correctly, safely or reliably.”<sup>1</sup>

*Medical information, as eloquently described above by Dr. Atul Gawande, is increasing at an accelerating rate. Physicians struggle to keep up with the latest developments. Doctors, like so many of us, are barely keeping their heads above water in the information tsunami of 2010. Medical information about patients in the form of electronic health records (“EHR’s”), while presenting physicians with the capacity to learn almost everything they may need to know about a patient’s history and condition, is proliferating faster than bacteria in a warm Petri dish.*

Huffing and puffing behind the medical profession as it races to catch up with this information wave is the legal profession. Yet, what if the unthinkable were attainable; that is, could doctors learn something from lawyers? The answer, in the age of the Electronic Health Record is “Yes,” because the same skills which lawyers counsel their corporate clients to apply in e-discovery readiness – identification, preservation, collection, review and disclosure – can help lasso the ever-growing health information beast. And in turn, better management of digital health information could streamline the time and dollar costs in litigation discovery of that information.

### **Heresy Absolved By a Dollar Sign**

Health care and law have always been in the same room at the same time under duress. This is due in no small measure to the disconnect between the black-and-white of the law and the shades of gray inherent in diagnostic workup. “The witness will answer Yes or No,” the court frequently instructs the expert witness from a prestigious academic medical center, in contradiction to the philosophy of consideration of all diagnostic possibilities, with which she daily imbues her students. Mixed into this roiling stew is the architecture of EHR’s which was never intended to accommodate courts or government oversight agencies. In short, if health records are difficult to manage for caregivers, how will attorneys and courts fit these square pegs into the round holes of litigation discovery?

It’s amazing how even these who have the least in common can find common ground where money is dangled, provided that they are guided to this realization by counsel working in collaboration with in-house resources toward the congruent goals of efficient information governance and e-discovery readiness. In other words physicians would be well advised to set aside their reputed animosity

---

<sup>1</sup> Gawande, Atul, *The Checklist Manifesto* (Metropolitan Books 2009) at 13.



toward lawyers in the interest of efficient use of EHR's for quality of care, because lawyers e-discovery readiness skills can assist them in meeting the criteria for federal incentive payments, and perhaps aid quality of care in the process through efficient health information governance.

The HITECH Act (Health Information for Clinical and Economic Health), a component of the American Reinvestment and Recovery Act of 2009 (ARRA, colloquially known as "The Stimulus Package"), appropriated approximately 27 Billion Dollars to assist in and facilitate the transition to digital health records. In June 2010 the Centers for Medicare and Medicaid Services issued the Final Rule for Medicare and Medicaid Programs: Electronic Health Record Incentive Program.<sup>2</sup> These 864 pages set forth requirements for accessing the Incentive Payment funds, beginning in January 2011.

Eligible Hospitals must attest, by the time they file their application for funds, that they have met 15 Stage One Core Criteria. The Criteria are called "Objectives," and the metrics to demonstrate compliance are known as "Measures." One Core Criterion requires the demonstrated capability to exchange key clinical information with caregivers and "other patient-designated entities."<sup>3</sup> Another Core Criterion requires hospitals to report on a variety of care metrics to oversight agencies.<sup>4</sup> Compiling the information on a particular patient to send to other hospitals or physicians, or to "patient-designated entities" such as courts, attorney and insurance companies, is a task far more daunting than it may seem.

In the paper days, one would send the request for records to the Medical Records Department, which would be the final repository of all the pieces of paper that made up the "patient's chart." Of course, this wasn't the end of the inquiry: the record could also comprise radiology films, sonography videos (now in digital media), off-site laboratory reports, and notes from departments scattered throughout the facility yet beyond the formal chart. In 2010, physicians often communicate with each other, to a great degree, through such electronic media as email and text messages. Without appropriate protocols, these communications may not make it to the formal "chart."

Text messages, which are on the service providers' servers (or in the Cloud) and not the hospital organization's network, may not be recoverable at all without procedures and training of staff to save these frequently relevant communications. In addition, clinical departments in a hospital may communicate over their own remote "modules," or applications that do not link to the organization's network or the Medical Records Department (which was set up for paper, not digital records). Imaging specialties such as radiology, sonography and Nuclear Medicine often record their images and findings on their own systems, as do the clinical and pathology laboratories.

---

<sup>2</sup> 45 C.F.R. parts 412, 413, 422 and 495 (June 13, 2010) ("Meaningful Use Final Rule")

<sup>3</sup> *Id.* at \*59; 229-231

<sup>4</sup> *Id.*

In order for a hospital to get its proverbial hands around all the clinical information subject to the Stage One Core Criteria, it must first ascertain what information it possesses and all the places where it may be located. E-discovery lawyers have been wrestling this issue for many years, ever since *Zubulake v. UBS Warburg*<sup>5</sup> set the standard that an organization must preserve relevant data when it reasonably anticipates litigation. Their first step toward corralling this data could be an Information Flow Assessment, or “Data Map,” and as hospitals find themselves the subject of more frequent and complex e-discovery demands, this Assessment or Map will serve a dual purpose of assisting the initiative to meet the stage One Core Criteria as well as e-discovery readiness.

Caregivers who receive and process so much electronic health information, are learning to adjust to the reality that, with health care as one of the most litigation-intensive industries in the U.S., their records will be sought early and often. There has, to date, been no “e-discovery exemption” for the medical disciplines when it comes to electronic discovery. Health care, in order to avoid devastating sanctions such as partial summary judgment and adverse inferences, should become familiar with principles of discovery, which by this time have been learned well by the litigation fire-trials of the securities, pharmaceutical, medical device and electronic industries.

#### **The E-Discovery Big Five: Identification, Preservation, Collection Review, and Disclosure**

These basic components of electronic discovery are, as seen above, transferrable between information governance to meet the Stage One Core Criteria and e-discovery readiness, but not yet in a seamless way. Definitions of “the legal patient record” will vary from jurisdiction to jurisdiction. Issues such as whether the record includes all, or some emails and texts between caregivers, voicemails, and metadata of the chart entries are still in an early state of evolution. Cost-shifting motions for data demands that may be broad-based, in the absence of pre-discovery depositions such as those permitted under Fed.R. Civ. Proc. 30(b)(6), have garnered inconsistent results.<sup>6</sup> Further, until national standards for EHR’s are fully implemented,<sup>7</sup> rarely will two institutions’ systems be alike. Definitions may differ among the plethora of applications, as will repositories, characterizations of data and standards for clinical documentation.<sup>8</sup>

Accordingly, the first step for requesting and producing parties is to identify the pertinent information. Failure to appropriately tailor the request may result in an incomplete production, sanctions or costs of search and production shifted to the requesting party.

---

<sup>5</sup> 229 F.R.D. 422 (S.D.N.Y. 2004)

<sup>6</sup> Compare *Cason- Merenda v. Detroit Medical Center* No. 06-15601, 2008 WL 2714239 (E.D. Mich. 2008), (denying cost-shifting) with *Makrakis v. Demelis*, \_\_\_\_ N.E..2d. \_\_\_\_ (Docket No. SUCV2009-00706-C, Sup. Ct. MA 2009), granting motion to shift costs to plaintiff in medical malpractice action)

<sup>7</sup> See, Generally, Office of National Clinical Health IT Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 45 CFR Part 170

<sup>8</sup> Brouillard, Chad P., *Emerging Trends in Health Record Liability*, Medical Liability and Health Care Law (Defense Research Institute July 2010)

Preservation is no small matter of concern for producing health care providers. Electronic information issues consume thousands of words and/or gigabytes in legal journals and court opinions for one intuitive reason: it is easily lost or altered. As there is as yet no set of best practices on documentation or the composition of the “legal” chart, health care providers are well-advised to map their information flows and evaluate how potentially relevant information is to be preserved. Failure to have a standard protocol for preservation has been called “gross negligence,” by a number of federal judges,<sup>9</sup> and health care providers have not escaped sanctions for failure to preserve potentially relevant electronic information.<sup>10</sup> The Litigation Hold Notice must be specific enough to capture all pertinent data types and all potential repositories of relevant data, including portable media of individuals subject to a Legal Hold who depart from the organization’s employ after receipt of a Legal Hold Notice.<sup>11</sup>

Preservation is perhaps the trickiest aspect of e-discovery – and information compliance with the Meaningful Use Stage One Criteria – because of the relative lack of experience of health care providers with disclosures of large amounts of electronic data, they “are still especially vulnerable to e- discovery requests due to failures to identify, locate, and produce all relevant data, failures to retain or store data, and failures to preserve data in its original form once a litigation hold has been issued, particularly in actively used or live EHR databases.”<sup>12</sup>

In this regard, discovery of digital patient data poses challenges and risks for requesting and producing parties. While the failure to narrow health information requests, due to overreaching or just a lack of understanding of the subject applications and systems may result in the requesting party being denied discovery or having the costs of their request imposed upon them (as in the old adage “Be careful what you ask for . . .”), incomplete collections can result in such sanctions as an adverse inference jury instruction, awards of attorneys fees, partial judgment or even referral for attorney discipline.<sup>13</sup>

As if the above were not sufficiently daunting, after review the producing party is under a further obligation to produce the pertinent data in a manner that comports with the Privacy and Security Rules of the Health Information and Portability Act of 2006 (“HIPAA”). These obligations include the requirement that the data disclosed be no more than the “minimum necessary” to meet the discovery request.<sup>14</sup> This may cast doubt on the ability of a hospital to respond, for example, to demands for

---

<sup>9</sup> *Keithley v. The Home Store, Inc.*, 46 A.D.3d 74, 842 N.Y.S.2d 558 (2d Dept. 2007). Verdict was reversed on other grounds (flawed jury instructions); *Treppel v. Biovail*, 249 F.R.D. 111 (S.D.N.Y. 2008); *Pension Committee of the University of Montreal v. Banc of America Securities, LLC*, 2010 WL 184312 (S.D.N.Y., 2010)(failure to issue written Litigation Hold as gross negligence)

<sup>10</sup> *United Med. Supply Co. Inc. v. United States*, No. 03-289C, 77 Fed. Cl. 257 (Fed. Cl. 2007)

<sup>11</sup> See, generally, *Cache La Poudre Feed, LLC v. Land O’ Lakes Inc.*, 244 F.R.D. 614 and *Rimkus Consulting v. Cammarata*, US Dist.Lexis 14573 (S.D.TX February 19, 2010). The term “portable media” should be broadly defined to capture all potential repositories of data subject to a Hold

<sup>12</sup> Brouillard, Chad P., *Emerging Trends in Health Record Liability*, Medical Liability and Health Care Law (Defense Research Institute July 2010) at 44.

<sup>13</sup> See *Qualcomm v. Broadcom*, 2008 US DISTLEXIS 911 (S.D.CA January 7, 2008)

<sup>14</sup> 45 C.F.R. §164.502(b)

“any and all records on Gladys Smith for all in-patient and out-patient admissions,” if only one admission is the subject of the law suit and the requesting party cannot show relevance to production of all records which may possibly span many years. Further, the health care provider must protect the confidentiality of the disclosure by administrative and technical means, pursuant to the HIPAA Security Rule.

This may well mean, for example, encrypting the transmission of electronic information,<sup>15</sup> to the consternation of a technically unsophisticated requesting party who cannot open an email attachment. The provider who has undergone the required HIPAA Security Risk Analysis pursuant to the Meaningful Use Final Rule,<sup>16</sup> however, will have already put into place protocols for such disclosures and will be prepared to respond to objections with regard to protection of the information disclosed in discovery. The Security Risk Analysis, with its requirements for documentation of health information security, will have the added benefit of assisting the producing party in admitting its electronic information into evidence, in that it will be able to document safeguards for the information and chain of custody, thereby establishing authenticity.<sup>17</sup>

### **Bilateral Health E-Discovery Management**

Discovery of electronic health information entails a significant risk of the tail wagging the dog; that is the devolution of the litigation into disputes about thoroughness of searches, which repositories should be searched and when, choice of search terms,<sup>18</sup> and of course, timeliness and manner of preservation. Lost in all this may be the merits of the underlying dispute and issues regarding the role of EHR's in quality of care. Digital health information, partly as a result of the availability of the HITECH Act Incentive Payments, will increase significantly in the next five years. Like day follows night, discovery disputes over that information will skyrocket, and will continue to proliferate until greater experience with digital records results in (relatively) uniform standards, including guidelines that take e-discovery into account. Until that happy day arrives the parties, in the interest of savings of time and resources, would be well advised to take seriously the Fed. R Civ. Proc. 26(f) discovery conference mandate to raise all pertinent issues regarding requests and production of health information before these issues reach the court, and to take into consideration the number of judges who have signed onto The Sedona Conference® Cooperation Proclamation,<sup>19</sup> a statement of principles that embody the concept that discovery is a cooperative component of an adversarial process.

In this way inroads can be made into making health information in litigation more manageable.

---

<sup>15</sup> 45 C.F.R. §164.312(a)(2)(iv)

<sup>16</sup> Meaningful Use Final Rule at \*220-221

<sup>17</sup> An excellent overview of the admissibility foundation requirements for electronic information may be found in the opinion of Magistrate Judge Paul Grimm in *Lorraine v. Markel Insurance Co.*, 241 F.R.D. 534 (D.MD. 2007)

<sup>18</sup> See *In Re Fannie Mae Litigation*, 2009 WL 21528 (C.A.D.C. 2009), in which a perhaps precipitous agreement to utilize proffered search terms resulted in a search costing over Six Million Dollars.

<sup>19</sup> Available at [thesedonaconference.org](http://thesedonaconference.org). Over 140 judges have signed the Proclamation

*Kenneth N. Rashbaum, Esq. is Principal of Rashbaum Associates, LLC in New York. A litigator in health care for over 25 years, Mr. Rashbaum's practice is focused upon health information management counsel for health care providers, compliance with HIPAA and state privacy laws, and e-discovery counsel to health care providers and multinational corporations. Ken is a member of the Steering Committee of the ABA Information Services, Technology and Data Protection Committee and a Vice-Chair of the ABA International Litigation Committee. More information on his practice may be obtained at [www.rashbaumassociates.com](http://www.rashbaumassociates.com)*

## Going Native – It's Really Not That Hard

By Sonya Sigler

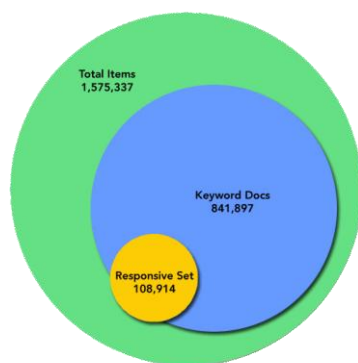


*Why are requesting parties letting producing parties<sup>1</sup> off easy when it comes to producing electronically stored information (ESI)? This article will address several areas where requesting parties can pressure the producing party to do the right thing or in another light - be more demanding, precise, and knowledgeable when making their e-Discovery requests.*

*Requesting parties do themselves a disservice in agreeing to produce data tied to using a keyword list, limiting produced meta data fields to "commonly used" meta data fields, not specifying a form of production (or caving in to a producing parties demands that they just want to produce TIFFs), and believing a producing party's statement that it will take six (6) months or more to review their documents for privilege. All of these issues can be avoided completely by going native! Going native here means reviewing data in its native format as well as producing that data in its native format.*

### Keyword Lists

Agreeing to use a key term or keyword list to find relevant data is a fool's errand. No keyword search is perfect – some searches are just more effective than others, but none is perfect. The key words thought to return relevant documents may return thousands of other documents too, which is over-inclusive (that is referred to as false positives in information retrieval); or the key words used in a search may not return the relevant documents at all, which is under-inclusive (that is referred to as a false negative in information retrieval).



In this Diagram, keyword search reduced the document set by only 47% and more troubling, 88% of the documents returned by this keyword search were not responsive (Over-inclusive).

Unfortunately, 8,553 responsive documents were missed by the keyword search, which is almost 8% of the responsive documents (Under-inclusive).

<sup>1</sup> There are too many instances where a party is both a requesting party and a producing party to just address this article to the Plaintiff's bar alone.

A case that illustrates this point succinctly is *Gross Construction*<sup>2</sup> in the Southern District of New York, where Judge Peck specifically berated lawyers who thought they could come up with an effective key word list (their key word list numbered in the thousands) without even consulting the people whose data they were searching to see what terms they actually used:

“This case is just the latest example of lawyers designing keyword searches in the dark, by the seat of the pants, without adequate (indeed, here, apparently without any) discussion with those who wrote the emails. Prior decisions from Magistrate Judges in the Baltimore-Washington Beltway have warned counsel of this problem, but the message has not gotten through to the Bar in this District.”

The proposed *Gross Construction* key word lists included words like “sidewalk,” “change order,” “driveway,” “access,” “alarm,” “budget,” “build,” “claim,” “delay,” etc. These words, as the parties pointed out would return the entire email database – not very effective. In fact, most terms were redundant and not effective at refining the search results. The court was left in the uncomfortable position of having to craft search terms without much help from the parties. As Judge Grimm stated in his *Victor Stanley*<sup>3</sup> opinion, “Selection of the appropriate search and information retrieval technique requires careful advance planning by persons qualified to design effective search methodology. Other judges are pressing parties to put together a reasonable method to find the relevant data. In one of the *AstraZeneca* cases,<sup>4</sup> even though it is relegated to a footnote, the court highlights the importance of effective search methodologies - Examples include omitting Seroquel's generic name, acronyms for diabetes, hyperglycaemia spelled the British way; and endocrine. The search method apparently failed to include common misspellings or the singular forms of words and failed to make allowance for spaces or dashes.”<sup>5</sup>

If these cases mean anything, it is that keyword lists should not be relied upon to find relevant data.<sup>6</sup> Courts are declining to evaluate search terms and are suggesting that search experts be used.<sup>7</sup> It is a disservice to clients everywhere. So what other search methods are out there? Other search and retrieval methods are being used – clustering, ontologies, data modeling.<sup>8</sup> ALL of these search

---

<sup>2</sup> *William A. Gross Constr. Assocs., Inc. v. American Mfrs. Mut. Ins. Co.* No. 07 Civ. 10639, 2009 WL 724954 (S.D.N.Y. March 19, 2009).

<sup>3</sup> *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 260, 262 (D. Md. May 29, 2008) (Grimm, M.J.).

<sup>4</sup> *In re Seroquel Products Liability Litigation* 244 F.R.D. 650 (M.D.Fla. Aug. 21, 2007)

<sup>5</sup> *Id.* Footnote 7.

<sup>6</sup> *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 260, 262 (D. Md. May 29, 2008) (Grimm, M.J.). Judge Grimm states, “The implementation of the methodology selected should be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented.”

<sup>7</sup> *Eurand, Inc. v. Mylan Pharms., Inc.*, 266 F.R.D. 79, 84 (D. Del. Apr. 13, 2010) (Thynge, Mag. J.).

<sup>8</sup> For more information on these other methodologies see The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery and Sigler, [Are Lawyers Being Replaced by Artificial Intelligence? Moving Beyond Keyword Search: An Introduction to Advanced Search & Retrieval Technologies](#).



methods (as well as keyword search) work more effectively on native data that has not been converted or truncated in terms of meta data being eliminated completely, or limited to commonly used fields, or limited to 10 chosen data fields. Native data includes the entire document, including all of the meta data. Most searches applied to converted data miss searching meta data completely. This can be important.

Using native data for whatever search methodology is used (not just keyword search) provides the best opportunity to actually find the relevant data. Ask for native data to be searched for more effective searches. Ask for native data to be produced in order to run your own searches on native data.

### **Meta Data**

Meta data is information about information or data about data. It is an integral part of any document. Likening an electronic document to a paper document gives a false sense of security that the page is produced as if it had been printed, then the requesting party is getting “the document.” That is a false assumption.<sup>9</sup> The meta data associated with most office documents and emails contains valuable information that may be of use to the requesting party (and in some cases, the producing party). For example, meta data fields can show who reviewed a particular document and when.

It drives me crazy when opposing counsel says – “Oh, we don’t need to produce the meta data, we just need to produce the document itself.” What drives me crazy about this line of argument is that it is missing the difference between paper documents and electronically stored information – they are not the same. Let me repeat that because it is an important point – paper documents are not the same as electronically stored information. Electronically stored information is much more robust in its details – and those details include all of the meta data.

I don’t know of any attorney who would think that it is OK to turn over a paper document having whited-out a hand-written comment on it; or having whited-out the bcc: line or the subject line of a memo. This is exactly what is happening if all of the meta data is not turned over. Meta data is an integral part of the electronically stored information, just as a handwritten comment on a memo is an integral part of that document.<sup>10</sup>

The Arizona Supreme Court got it right when they said: “A public entity is not required to spend “countless hours” identifying metadata; instead, it can satisfy a public records request merely by

---

<sup>9</sup> [The Sedona Principles](#). Even the Sedona Conference Principles have clarified this false assumption about meta data. The prior version of its Sedona Principles said that most meta data was useless. That well may be true about some meta data fields, but when meta data is useful it is very useful. This fact was acknowledged when the revised Sedona Principles were published in 2007. See Principles 12 and 14.

<sup>10</sup> *Lake v. City of Phoenix*, 220 Ariz. 472, 207 P.3d 725 (App. 2009). It would be illogical, and contrary to the policy of openness underlying the public records laws, to conclude that public entities can withhold information embedded in an electronic document, such as the date of creation, while they would be required to produce the same information if it were written manually on a paper public record.

providing the requestor with a copy of the record in its native format.”<sup>11</sup>

I have heard parties propose to each other – “let’s just agree to produce a TIFF of the documents plus the text searchable file along with commonly used meta data fields; all of this requires conversion, and the attendant additional expense, rather than just (cost-effectively) producing the native document.

Recently, a lawyer at a plaintiff’s law firm asked me what meta data fields he should ask for if he was going to agree with the opposing counsel to just produce 10 meta data fields. I told him that’s a ridiculous proposition; why not just get the native document? He said the opposing counsel was pushing him to limit the number of meta data fields produced. A different plaintiff’s lawyer asked me, “Why not just agree to commonly used meta data fields?” I looked at him like he was crazy and said “common to whom?”

Limiting meta data fields has two flaws: first, as I described above, fundamentally, electronically stored information is different than paper documents, with the primary difference being that ESI contains meta data, paper doesn’t. Producing ESI without any meta data is like whiting out the handwriting on a paper document before turning it over. The second flaw with limiting meta data fields is that you never know which meta data fields are going to be important until you start reconstructing what happened. A pet’s name in an address book field may be the key to breaking a password protected document open with a password cracker that can use pet’s names, kids names, and birth dates to crack password protected documents. Also, meta data tells the real story with regard to the lifecycle of ESI – who saw it, when, did they make changes to it and send it around again – all kinds of useful information that is specific to ESI (not paper).

The two plaintiff’s lawyers I mentioned above are both pushing for native document productions now that they see the fundamental difference between paper documents and ESI. Native productions give the requesting party the meta data, unaltered, unadulterated, and unchanged. Native data productions give the receiving party ESI, as the producing party used it. So, please, stop agreeing to convert data to other forms and stop agreeing to limit meta data fields – go native – as the ESI occurred AND, more importantly, as it was used by the producing party, meta data and all.

### **Form of Production**

The most common objection I hear about native as the form of production is that “I can’t Bates stamp native data. How will I know it is the same document I produced?” I have news for producing parties – the same way you did in the past – authenticate it!

---

<sup>11</sup> *Id.*

Bates numbering is no longer necessary on documents. Each document that is processed and reviewed has a unique hash value associated with it. Duplicates are determined by matching hash values among documents. Produced documents carry those same hash values with them when they are produced as ESI. The use of hash values to match documents is a comparable method to Bates stamping that enables parties to make certain that the documents produced and the documents used by the other party are one in the same.<sup>12</sup>

A requesting party can specify the format of ESI to be produced. Otherwise Rule 34 states that the producing party can produce ESI as it was kept in the usual course of business.<sup>13</sup>

Unfortunately, when it comes to producing ESI, “wink and a nod” games are still being played. “Don’t ask me for native data and I won’t ask you for native data.” This is just another version of a hide the ball tactic, or, more likely, just plain ignorance in the profession.<sup>14</sup>

Producing native data should not even be an issue. Producing data in native format is cheaper than converting data to TIFFs or PDFs. It is less likely to be corrupted or changed if it is produced in native format than if it is converted. Disturbingly, I am only seeing corporations insist that native data be produced because they do not want to pay for converting it to another format. I am not seeing law firms ask for native data. I am not sure if law firms are clinging to TIFF format productions because they are comfortable with that form of data or if they feel they have to use their Concordance or Summation licenses, or if they just don’t understand the benefits that can be realized with native data.

The second most common objection I hear about producing native data is that you can’t redact native data. That is just plain wrong. There are software tools available that enable native data to be redacted. As a second choice, only the documents to be redacted can be converted to TIFF or PDF and then redacted (there are many more redaction tools for these converted formats than there are tools for actual native redaction).

So, requesting parties, stop letting the producing party choose how to produce ESI; Ask for native data in your production requests, and be ready to refute opposing counsel’s objections.

---

<sup>12</sup> Losey. [HASH: The New Bates Stamp](#), 12 Journal of Technology Law & Policy 1 (June 2007).

<sup>13</sup> FRCP 34. (E) *Producing the Documents or Electronically Stored Information*. Unless otherwise stipulated or ordered by the court, these procedures apply to producing documents or electronically stored information:

- (i) A party must produce documents as they are kept in the usual course of business or must organize and label them to correspond to the categories in the request;
- (ii) If a request does not specify a form for producing electronically stored information, a party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms; and
- (iii) A party need not produce the same electronically stored information in more than one form.

<sup>14</sup> See Ralph C. Losey’s Book: *Electronic Discovery: New Ideas, Case Law, Trends and Practices*, 2010 ed., which contains [Ralph Losey’s Blog](#) from 4/27/2009 and 5/17/2009 on incompetence when it comes to ESI and the legal profession.

## Privilege

One of the loudest and most frequent complaints I hear is that “We can’t possibly produce data in that amount of time because we have to review for privilege.” Reviewing for privilege is certainly an important concern that cannot be understated, but there are many ways to address this issue. There are many new tools on the market that help segregate privilege information for manual review and there are even tools available that can automate privilege review and generate privilege logs. Granted, generating privilege logs can be tedious, painful, and time-consuming, but they are important to do well. Thankfully, generating privilege logs can even be automatically using new tools that have appeared on the market.

All of this data administration for privilege review and privilege log generation is done most effectively with native data that has not been altered – including any normalization of names or email addresses. Normalization is the process of replacing text blocks in the data so that searching can be sped up. Unfortunately, this can also mask some of the pertinent information that may be useful to reviewers. A simple example is email addresses may be substituted with a person’s name so that the email address is no longer seen because the name is what is visible in a data field after normalization.<sup>15</sup> Finding privilege material is more effective with the native data rather than with normalized data.

It is not a stretch to say that the use of these types of software tools is encouraged under FRE 502 to be able to manage the growing mound of ESI. Specifically with regard to privilege a claw back or quick peek arrangement might be made before hand.<sup>16</sup> But do not rely on this as a safety net, it is merely a precaution. Privilege review is the safest route but can be expensive.

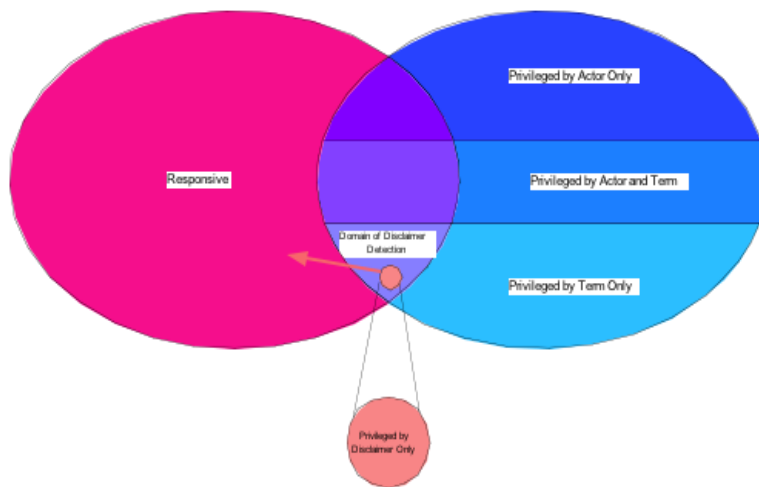
The Notes to FRE 502(b) state that the use of “advanced analytical software applications and linguistic tools in screening for privilege and work product may” be considered in determining whether a party has taken “reasonable steps to prevent inadvertent disclosure.” As the diagram illustrates data can be pinpointed and segregated for privilege review by using ontologies to locate documents that are potentially privileged by a term alone (i.e. the word privileged) or by a person or by both. This allows for automated and quicker privilege determination. This graph is one example of how technology can be used for privilege screening. Other tools exist to actually generate privilege logs. These tools use native data.

Requesting (and producing) parties should make use of technology to simplify and automate privilege review and request native data to make them the most effective. The more data that is culled out, by

---

<sup>15</sup> I can’t emphasize this enough – that attorneys should know what their vendors are doing with regard to normalization in the data processing step. If you don’t understand what is happening to your ESI, please do not be afraid to ask vendors what is being done in the data processing step.

<sup>16</sup> A claw back agreement is where the parties agree to return any inadvertently produced documents and that such a production will not result in a privilege waiver. A quick peek agreement is where documents are produced without review then the other party says which documents they want, and then the producing party reviews only those desired documents for privilege. Neither of these arrangements is ideal.



use of normalization, conversion to TIFF or PDF, or by limiting meta data fields, the less effective software tools will be.

Going native, or rather, using native data for document review and production can lessen the time it takes to review for privileged ESI; it can make searching for relevant data easier and quicker; and with meta data included, it can make the searches more effective. So, the bottom

line is, ask for native data to be reviewed and produced.

*Sonya L. Sigler is the Vice President, Operations & General Counsel at Cataphora, Inc. She is a frequent writer and speaker on electronic discovery issues and other topics. Ms. Sigler is a member of the Association of Corporate Counsel, the American Bar Association, and The Sedona Conference Working Group 1 on Electronic Document Retention and Production. She is active in her community and is on the Governance Council for the San Carlos Charter Learning Center and is the Chairman for the Chickens' Ball Steering Committee. Past board work has included the Women in Interactive Entertainment Association, Women in Technology Advisory Board and the Nova Vista Symphony. When not working she enjoys playing trombone in several Bay Area groups and coaching her sons little league and soccer teams. Sonya holds a JD from Santa Clara University and a BA in Philosophy with a music minor from UC Berkeley and is a member of the California State Bar. She can be reached at [sonya.sigler@cataphora.com](mailto:sonya.sigler@cataphora.com).*

## E-Discovery Search: The Truth, the Statistical Truth and Nothing but the Statistical Truth

By Nick Brestoff



*This article is a call to revisit Rule 26(g)(1) of the Federal Rules of Civil Procedure, which requires attorneys to certify “to the best of the person’s knowledge, information, and belief formed after a reasonable inquiry” that disclosures are “complete and correct.”<sup>1</sup> Given the exponentially growing mountain of electronically stored information (ESI), and the incompleteness and statistical nature of search technologies, which this article will explain, no attorney can honestly so “certify.” One day, this gap, a loophole between the law of yesterday and the technology of today, will cause a monumental waste of judicial, attorney, and client resources.*

Most of us know the meaning of a “loophole.” These days, when one seeks a definition, or perhaps an example, we look online and, more often than not, we turn to Wikipedia. According to Wikipedia, “[a] loophole is a weakness or exception that allows a system, such as a law or security, to be circumvented or otherwise avoided. Loopholes are searched for and used strategically in a variety of circumstances, including taxes, elections, politics, the criminal justice system, or in breaches of security.”<sup>2</sup>

Wikipedia mentions the “criminal justice system.” But to this entry we must add our system of “civil justice,” and, in particular, the giant middle of every lawsuit, discovery. As most attorneys are now aware, what used to be thought of as “discovery” is now dominated by e-discovery.

But e-discovery is a hybrid, a confluence of slowly changing laws and rules, on the one hand, and rapidly changing computer-based technologies, on the other. In this dynamic context, which besets every system of justice in the world, loopholes may be expected. Here we explore a rather large disconnect (or loophole) in the U.S. system of justice which comes as a result of the new complexities of e-discovery.

### Loopholes

In the context of e-discovery, lawyers have attempted to exploit what they thought were loopholes right from the start. Examples abound. In one case, for example, when the format for producing ESI was not specified and emails (and only emails) were requested, they were produced, but they were

---

<sup>1</sup> Rule 26(g)(1)(B) applies the certification to discovery responses, and requires a certification that is “consistent” with the rules, which includes Rule 26(g)(1)(A).

<sup>2</sup> The Wikipedia entry for “Loophole,” as modified on 27 July 2010, was viewed by the author on August 27, 2010.

“divorced” from their attachments, which were not produced.<sup>3</sup> In another case, a producing party converted searchable documents into nonsearchable TIFF files before producing the ESI.<sup>4</sup>

These gambits revealed certain weaknesses in the system, and some of them have been addressed. Now, for example, the federal rules provide that when a party is seeking documents from an opposing party, or from a third party pursuant to a subpoena, the requesting party may specify the *form or forms* of the documents when they are produced.<sup>5</sup> California’s e-discovery statutes also provide that the requesting party may specify the “form or forms in which each type of [ESI] is to be produced,” but, like the federal rules, the requesting party has only an opportunity to specify the forms *once*.<sup>6</sup>

Even though its growth-rate is prodigious, the hallmark of e-discovery is the immense *volume* of ESI that must be addressed. In 2003, researchers at UC Berkeley published an update to their study, *How Much Information?* At that point in time (and now hopelessly outdated), they explained that each year almost 800 megabytes of recorded information was produced per person, and that 92% of that information was stored on computers or a computer-based storage system.<sup>7</sup> Eight hundred megabytes is enough to fill a set of books stacked 30 feet high. Today, if each person generated only 25% more information than in 2003, or 1,000 megabytes, then each person would generate a gigabyte of data per year, and that amount is roughly equivalent to 75,000 pages, if printed.<sup>8</sup> It is easy to imagine that today we generate much more than that. Indeed, it is often said that 98% or 99% of all the information generated today, by everyone in the world, is generated as ESI. Why? Because today the digital universe includes not only servers, desktops, laptops, cell phones, hard drives, flash drives, and photocopy/fax machines,<sup>9</sup> the digital universe includes data from TV and radio transmissions, telephone calls received as emails, surveillance cameras, datacenters supporting “cloud computing,” and, of course, social networks.<sup>10</sup>

So, in lawsuits, parties and attorneys must often deal not just with gigabytes of data, but with several terabytes of data, and a single terabyte is roughly equivalent to 75 *million* pages, if printed. Even if a

---

<sup>3</sup> See *PSEG Power N.Y., Inc. v. Alberici Constructors, Inc.*, No. 1:05-CV-657 (N.D.N.Y. 2007) (producing party ordered to re-produce ESI at its cost).

<sup>4</sup> See *Goodbys Creek, LLC v. Arch Ins. Co.*, No. 3:07-cv-947-J-34 HTS (M.D.Fla. 2008) (conversion held improper; producing party order to re-produce ESI); *L.H. v. Schwarzenegger*, 2008 U.S. Dist. LEXIS 86829 (C.D.Cal. 2008) (sanctions were imposed for the untimely (late) production of non-sortable PDFs).

<sup>5</sup> Federal Rules of Civil Procedure, Rules 26(f)(3)(C) [discovery plan] and 34(b)(1)(C) [content of the request].

<sup>6</sup> California Code of Civil Procedure §2031.030(a).

<sup>7</sup> Lyman, Peter and Varian, Hal, *How Much Information?* (2003); see <http://www.sims.berkeley.edu/how-much-info-2003> (reviewed on August 28, 2010).

<sup>8</sup> *Id.*

<sup>9</sup> Keteyian, Armen, “Digital Photocopiers Loaded with Secrets: Your Office Copy Machine Might Digitally Store Thousands of Documents That Get Passed on at Resale,” CBS News (New York, April 15, 2010); See <http://www.cbsnews.com/stories/2010/04/19/eveningnews/main6412439.shtml?tag=mncol;txt>.

<sup>10</sup> Gantz, et al., *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011* (March 2008) (Executive Summary). See <http://www.idc.com>.



requesting party asks for readily accessible data, meaning data in native format with metadata intact, there is still the problem of how to search through a much, much bigger haystack than lawyers ever faced when, e.g., 10,000 boxes of documents were produced.

### Key Words and Boolean Searches

Now, how can anyone get their arms around this much data? They can't. The volume of data today is far greater than those times when parties attempted to hide the needle in the haystack by producing truckloads, or worse, warehouses full of boxes stuffed with papers. E-discovery expertise is partly the domain of an information technologist and partly the domain of lawyers. The technologist's approach is to cull the data by removing exact duplicates (de-duping) and system files. Culling will certainly reduce the size of the data set. Now the lawyer's task is to query that data set with key words and "field" terms, just as they did when searching opinion databases for applicable case law. Because they are familiar with key words, the receiving attorneys include key words describing the subject matter of the dispute and the names of the key players and employees who had "any involvement with the issues raised in the litigation or anticipated litigation."<sup>11</sup> An oft-used field term is a date or a range of dates.

Indeed, in the context of online legal research, teams of lawyers and other law firm denizens have become "power users" of key words and field terms. It was not always so. Fifty years ago, lawyers relied on their memories and library tables populated with books. Their search technique was non-linear and depended on a more personal skillset. But once the published cases were uploaded and computers could be used to hunt through databases, key words, date ranges, and Boolean connectors (e.g., AND, OR, NOT, term X "within 7 of" term Y, etc.) were deployed. Lawyers have been using this technique for over 35 years.<sup>12</sup>

But now the scope of the data is vastly increased and the problem is different. The problem is different because we are not querying databases of published opinions in which courts use familiar legal terms. In the e-discovery context, we are working within the context of the law, but we are not looking for it. We are trying to find the facts, and we are trying to find them in a mountain of data that is not only enormous, it is contained in numerous places. In this endeavor, opposite sides have different goals, especially because they treat the discovery process as an adversarial adventure and notwithstanding the platitudes spoken about cooperation.

---

<sup>11</sup> See *Pension Comm. Of the Univ. of Montreal Pension Plan v. Banc of Am. Sec., LLC*, 2010 WL 184312 (S.D.N.Y. Jan. 15, 2010; as amended May 18, 2010) (Scheidlin, J.)

<sup>12</sup> Here is an example of a "power key word search" using Boolean operators (which were borrowed from computer programming): (successor /5 corporation) /p (toxic or hazardous or chemical or dangerous /5 waste) /p clean! and da(aft 1/1/90). In plain language, this search is for cases where a successor corporation is liable for the cleanup of hazardous (toxic) waste. The sample Boolean search looks for the combination of *successor* within five words of *corporation*, in the same paragraph as the combination of *toxic* or *hazardous* or *chemical* or *dangerous* within five words of *waste*, within the same paragraph as *clean* or *cleanup* or *cleans* or *cleaned* or *cleaning* (the exclamation mark in *clean!* causes the computer to search for all words with *clean* as a root). Cases are limited to those dated after January 1, 1990.

For example, a requesting party may attempt to use key words to over-collect documents. In one recent case, for example, where, pursuant to a stipulated order the defendant had sole discretion to specify search terms, the defendant submitted 400 search terms. Over the producing party's objections based on cost (\$6 million), which the court denied because of the stipulated order, these 400 terms yielded 660,000 documents.<sup>13</sup>

On the other hand, a producing party may attempt to under-produce. They may use key words to narrow the scope of the documents they must produce. In a different case, the plaintiff requested documents from the hard drives of 26 employees. The defendants used de-duplication to narrow the documents to be produced down from 423,835 to 129,000, and then used search terms to narrow the actual production down to 4,000 documents. The plaintiff objected, and wanted more, but the magistrate dismissed the plaintiff's objections, stating "To the extent Plaintiff contests the adequacy of the search terms, it has not set forth an alternative search methodology; moreover, no specific challenge to the search terms has been brought and briefed before the Court."<sup>14</sup>

Ah, now there's a rub. Is there an alternative search methodology? Yes. But before describing it, let's stay with key words for a moment. The goal, after all, is to use automated, computer-based searches to find as many of the potentially relevant documents as we can. All non-privileged information relevant to a claim or defense must be produced.<sup>15</sup>

But just how successful are key word searches? Test yourself. Here's the proposition: Key words using Boolean connectors will find only about 25% of the relevant documents. True or false?

True! One of the founders of the "information retrieval" field, M. E. Maron (now professor emeritus, UC Berkeley) reported as long ago as 1985 that attorneys were over-estimating the efficacy of their searches. The attorneys thought they were identifying 75% of the relevant documents, but they were wrong: they were finding *only about 20%*.<sup>16</sup> More recently, studies show that key word searches are, even today, only a little more successful. Tomlinson and others reported in 2008 that Boolean searches identified only 22% of the relevant documents,<sup>17</sup> while Oard and others reported in 2009 that Boolean searches pinned only 24% of the relevant documents.<sup>18</sup> (These reports come from the Legal Track of the Text Retrieval Conference (TREC), which is administered by the U.S. National Institute of Standards and Technology.)

---

<sup>13</sup> See *In re Fannie Mae Secs. Litig.*, 552 F.3d 814, 818-819 (D.C.Cir. 2009).

<sup>14</sup> See *In re CV Therapeutics, Inc. Sec. Litig.*, 2006 WL 2458720 (N.D.Ca. Aug. 22, 2006).

<sup>15</sup> Federal Rules of Civil Procedure, Rule 26(b)(1); see *Zubulake v. UBS Warburg LLC*, 217 F.R.D. 309, 316 (S.D.N.Y. 2003); *SEC v. Collins & Aikman Corp.*, 256 F.R.D. 403, 417-418 (S.D.N.Y. 2009) (over objections based on cost, SEC ordered to produce emails; parties required to establish a reasonable search protocol).

<sup>16</sup> Maron, M. E., An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval Sys., 28(3) Comm. of the ACM 289 (1985).

<sup>17</sup> Tomlinson, Stephen, et al., Overview of the 2007 TREC Legal Track (April 30, 2008).

<sup>18</sup> Oard, Douglas W., et al., Overview of the 2008 TREC Legal Track (March 17, 2009).

Now for attorneys used to key word searches, these reports are not good news. As previously noted, in the process of “early disclosure” and responding to document requests, an attorney must certify that, “to the best of [their] knowledge . . . formed after a reasonable inquiry,” their response to a document request is “complete and correct.”<sup>19</sup>

Is there an alternative methodology to key words and field terms? Yes. We come to it now: concept search.

### Concept Search

What is concept search? Concept search is a way of finding patterns in unstructured data sets. Its sounds technical, doesn’t it? Yes, it is. It involves matrix algebra, formulas you don’t want to see (ever), and statistical concepts you don’t want to know about, but will be forced to learn anyway (note: more on this point, later).

Let’s stick with key words for a moment. Key words approach a document collection in a simplistic way; either a document contains the key word (or a variation of it) or it does not contain that word. Let’s say we have only two key words, w1 and w2, for our query, and that we find w1 in document 1, which we’ll call d1, and w2 in document 2, or d2; but we do not find w1 in d2 and we do not find w2 in d1. In the four-square box at the end of this sentence, a “1” means that the word in question is present, while a “0” means that the same word is not present:

Dox ---			
>	d1	d2	
w1	1	0	
w2	0	1	

This simple “picture” is a hypothetical word-document matrix. It is clear that using w1 as “input” will result in d1 as “output,” but not d2. If we use w2 as input, we will get d2, but not d1. But if we are looking for a document with both w1 AND w2, we will get nothing.

But wait. This matrix is too simplistic. It consists of only two key words and only two documents. The documents in our collection, which will likely consist of gigabytes and terabytes of data, are certain to have many more than one word each. Here is the key to understanding what concept search engines do: they find with “co-occurrences” of words that are *not* used as search terms.

If a picture is worth many words, a bigger matrix should help. You can see what co-occurrence means by looking at the next matrix.

<sup>19</sup> Federal Rules of Civil Procedure, Rule 26(g)(1)(A).

Dox	d1	d2	d3	d4	d5	d6	d7	d8
Word								
w1	1	0	0	0	0	0	0	0
w2	0	1	0	0	0	0	0	0
w3	1	1	0	0	0	0	0	0
w4	0	1	0	1	1	1	1	1
w5	0	1	0	0	1	1	1	1
w6	0	1	0	0	0	1	1	1
w7	0	1	0	0	0	0	1	1
w8	0	1	0	0	0	0	0	1

It starts in the upper left hand corner with the simple four square matrix of (w1, w2) and (d1, d2) that we first described. But then this matrix adds more words (w3 through w8) and more documents (d3 through d8).

Let's begin with w3. It appears in both d1 and d2. When we were considering the four-square matrix, inputting w1 AND w2 did not result in either d1 or d2; it resulted in nothing. In the matrix below, if we input w3, we will get d1 *and* d2, because it is contained in both documents.

Now look at w4. It is contained in d2 and d4, d5, d6, d7, and d8. Similarly, w5 is in d2 as well as in d5, d6, d7 and d8 (so one less; w5 is not in d4). And so on. Now we can make some observations about our collection (or corpus).

First, note that neither w1 nor w2 are in any of the other documents, d3 through d8, which is why, for the w1 and w2 rows, there are nothing but "0s" in the columns after d2. For both the rows for w1 and w2, the columns d3 through d8 are all zeros.

Also, no matter what word we use to query this matrix, will we ever get back d3? No. It has none of the words on the list.

Now let's look at words w4, w5, w6, w7, and w8. Notice that w4 shows up in d4 through d8. Fine, that word is used frequently. But *frequency* is not the test.

The big idea of concept search is to find documents (as output) that are responsive to a query (using key words as input), based on co-occurrences. As output, we want documents that have key words *in them*, but also the documents that do not contain *any* of the key words but which are nevertheless potentially related and, thus, potentially relevant. We are looking for patterns.

In this regard, patterns can be strong or weak. Which document exhibits the strongest pattern? It's d8. Although d8 does not have our input key words, w1 or w2, column d8 has five of the same content words contained in d2; that is, both d2 and d8 have words w4 through w8 in common. The weakest pattern involves the most documents but the weakest link: d4 through d8 all share only one word – w4 – with d2.

Computers do not understand “patterns.” They go through a process (a series of steps) which eventually leads to a measurable threshold, a cut-off point. To scholars in the field of Information Retrieval, such steps, including the mathematical scissors, is called an “algorithm.” In our simplistic hypothetical, if we want all documents that are potentially relevant, we might choose a cut-off where there is only *one* matching co-occurrence, a low threshold. If we want to find a “smoking gun,” we might search again, this time adjusting our process (algorithm) to find only the strongest co-occurrences. In this example, if we want, say, more than four (4) co-occurrences, the search output would be *only* d8.

See how this works? With concept searching, computers are going through gigabytes and terabytes of data consisting of documents and words, using a strictly mathematical approach.

This search methodology is called Latent Semantic Indexing or LSI. This term is best understood “inside out.” The “Index” part is simple. You have seen indexes before. They are at the end of nearly every book. Indexes indicate which words are on which page. Here, the computer ingests all of the documents and all of the words, and creates an index of each word that is contained in each document. We have just done this with two hypothetical matrices, one with two words and two documents, the other with eight words and eight documents.

What does “Latent” mean? Roughly speaking, it means “hidden.” And “Semantic” means, again roughly, “meaning.”

So, the phrase is actually descriptive of what we are trying to accomplish: find the hidden meanings (patterns) in a collection of documents, not because of the specific words we choose as input, but because of the *other* words in the documents containing the words we *did* choose and their “co-occurrence” with words in other documents, documents which do *not* contain our search terms.

Let's deepen our understanding. As we did with the documents themselves, culling out exact duplicates and system files, let us cull our words. In LSI, we discard articles (like “a” and “an”); prepositions, conjunctions, common verbs (like known, see, do, be); pronouns (e.g., it, they); common adjectives (like big, late, and high); pointer or frilly words (like thus, therefore, however, and albeit); any words that appear in *every* document; and any words that appear in only *one* document. Now we are down to the core words that have semantic value; they have “content.” It is with these words that we form the word-document matrix.

Now we do some “weighting” (think “handicapping”). Some content words appear more than once in a single document. They are given greater weight; and the process of giving them more weight is called “local weighting.” Still other words show up frequently throughout the entire set, and because of this, they are “commonplace.” Words that appear in only a small handful of documents may have special significance. They get greater weight. This is “global weighting.” And there is a scaling step, called “normalization,” which is just like handicapping in golf. Some documents may be long ones and have many key words. To keep them from overwhelming the shorter documents, the larger ones are penalized a bit, so that every document has, approximately, equal significance.

Because LSI is mathematical, it is a search engine that “likes” addressing large collections of data. The more words and documents in the set, the better LSI performs at finding documents responsive to a query. And, after a fruitful search puts some documents into a “shopping cart,” a human being can learn from the initial results and *iterate* the process. With this feedback, the input terms are more focused and the LSI search engine is likely to produce even better results.<sup>20</sup>

LSI was not conceived to address the problem of search in the e-discovery context.<sup>21</sup> But it has found application in the world of e-discovery. Moreover, because many business and governmental endeavors involve more than one language, LSI is useful because it does not pretend to understand anything about the words it is considering. The words are, in a sense, digitized; then LSI creates the word-document matrix, and seeks out the patterns based on statistical co-occurrences. It is therefore as functional with words in Chinese, Korean and Japanese (or Arabic) as it is with words in English. Using LSI, “hot” documents across different languages can be identified. The next step is machine translation, which is *not* known for precision. So, the step after that is human review. And if certain documents appear to a human to be suitable for use in deposition, in a motion, or at trial, the final step is human translation, so that the translated documents can be certified and offered into evidence.

In the e-discovery context, you have likely seen LSI in action. You just didn’t know what was “under the hood.” Simply put, concept search based on LSI, or a variant of LSI, is now at the heart of programs that are offered by a number of different vendors, each of which has provided different “bells and whistles” to differentiate themselves.<sup>22</sup>

---

<sup>20</sup> Two tests are “recall” and “precision.” Recall is the proportion of relevant documents identified out of the total number of relevant documents that exist. If the total number of relevant documents is 100, but a search identified 80, the recall rate is 80%. Precision is the percentage of identified documents that were actually relevant. If 100 documents were identified but only 75% of them were relevant, the precision would be 75%. Using LSI, recall and precision rates just under 90% have been achieved. Source: Content Analyst Company, LLC (“Content Analyst”) in Reston, Virginia (<http://contentanalyst.com>). Content Analyst is the original patent-holder of LSI.

<sup>21</sup> See Landauer, T. K. and Dumais, S. T., “Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge,” *Psychological Review*, 104(2), 211-240 (1977).

<sup>22</sup> There are at least three hosted review platforms that have integrated an LSI solution from Content Analyst: Relatively (by kCura), iCONNECT, and Eclipse by IPRO. In addition, a variation of LSI called Probabilistic LSI is “under the hood” of Axcelerate by Recommind.

Why is LSI powerful? Because, when LSI is used on unstructured data, such as business communications, LSI returns documents that may be highly relevant that even power key word searching would miss. Here's an example. In a stock option back-dating case, an LSI-based search returned documents whose common denominator (pattern) was the phrase "Let it roll." Why return these documents? Remembering that LSI is designed to seek out hidden meanings, the consultants involved in the case called the "Let it roll" group to the attention of the litigators. Sure enough, this phrase turned out to be the "go" signal the executives were using to authorize the back-dating. Unless a power key word searcher made a lucky guess, the "Let it roll" documents – the key needles in a very large haystack – would have gone undetected.

So LSI has proven to be more efficient than key words, even though key words are still used in the queries that are framed. But could you have explained LSI to a court, in case you were challenged by opposing counsel to do so?<sup>23</sup>

### **It's All Statistical**

Now, finally, we come back around to whether an attorney can honestly sign off on the Rule 26 certification concerning the documents he or she has disclosed or produced. With a new appreciation for what goes into searching a collection for potentially responsive documents, the answer is "no." We have a loophole. Attorneys are, by rule, being forced to certify to a degree of certainty that just is not there; and they put their licenses on the line when they sign.<sup>24</sup>

Suppose we have collected 100 million documents; how many should be produced? A suitably sized random sample will accurately reflect the number of responsive documents to be produced, no matter how large the set may be.<sup>25</sup> For a confidence level of 95%, with an error of plus or minus 5%, a random sampling of 1,537 documents must be examined. For a confidence level of 99%, with an error of plus or minus 1%, a sampling of 66,358 documents is needed. Thus, "if we have 100 million documents in the unretrieved set, we need to examine only 1,537 documents to determine within 95% confidence that the number of responsive documents in the unretrieved set is within the margin of error. If we find that there are 30 documents that were responsive in the unretrieved set, we can state that we have 95% confidence that the number of responsive documents in the sampled set is between 28 and 32 (rounding up the document count on the high end, rounding down on the low end). Extending that to the 100 million population, approximately 1,951,854 plus or minus 97,593 are responsive in the unretrieved set. [Para.] In the case of a review where errors are expensive (such as a review for privilege), 99% confidence with 1% error condition would require 66,358 samples. If we identify 200

---

<sup>23</sup> For that matter, could you differentiate LSI from still other computer-based search approaches, including taxonomies, ontologies, and Bayesian classifiers? These topics are beyond the scope of this article.

<sup>24</sup> See *Qualcomm, Inc. v. Broadcom Corp.*, No. 05 Civ. 1958-B, 2008 U.S. Dist. (S.D.C al. Jan. 7, 2008); and *id.*, Order Declining to Impose Sanctions, Etc. (Document 998; filed Apr. 2, 2010).

<sup>25</sup> Search Guide, Electronic Discovery Reference Model Draft v.1.17 at p. 79 of 83 (May 7, 2009).



privileged documents in such a sample, you will have 99% confidence that the number of privileged documents in the sample is between 198 and 202 privileged documents.”<sup>26</sup>

### Some Proposals and a Grand Conclusion

As previously mentioned, responding attorneys must currently certify that, “to the best of [their] knowledge . . . formed after a reasonable inquiry,” the disclosure or response to a document request is “complete and correct.” But in this digital era, attorneys must face up to understanding some of the math they hoped to avoid (forever) by going to law school, because attorneys are ill-equipped to flatly certify the “completeness” of their disclosures or responses. “[T]he 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.”<sup>27</sup> So how can attorneys vouch for “completeness”? Clearly, attorneys who continue to sign off on Rule 26(g) certifications are over-promising. They are venturing into areas where an expert’s opinion is warranted, if not necessary.<sup>28</sup> If a client is prejudiced when a court agrees, after some future battle over the alleged impropriety of an attorney’s certification, that “completeness” was promised but not achieved, will that attorney have fallen below the standard of care? Having likely ventured beyond his or her competence, will that attorney have violated a rule of professional conduct? Is a malpractice lawsuit in that attorney’s future?

We come now to four concrete proposals for change, and one grand conclusion:

- Rule 26(g)(1)(A) should be changed to indicate (for example) that, with the assistance of experts, the document production is complete and correct, with a 95% confidence level and an error rate of plus or minus 5%;

<sup>26</sup> *Id.*

<sup>27</sup> *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 262 (D.Md. 2008) (Grimm, J.) (quoting from “Information Inflation: Can the Legal System Adapt,” 13 Rich. J. L. & Tech. 10 (2007), at \*38, 40). See *Mt. Hawley Ins. Co. v. Felman Prod., Inc.*, 2010 WL 1990555\*10 (S.D.W.Va. May 18, 2010) (failure to sample in order to identify and remove privileged documents was “imprudent”).

<sup>28</sup> In several recent cases, courts have made statements supporting the proposition that a certification of completeness of a large document product by an expert should replace certification by an attorney. For example, in *United States v. O’Keefe*, 537 F.Supp.2d 14, 24 (D.D.C. 2008), the court stated, “Whether search terms or ‘keywords’ will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics . . . . Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread.” In *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008), the court stated, “Determining whether a particular search methodology, such as keywords, will or will not be effective certainly requires knowledge beyond the ken of a lay person (and a lay lawyer) . . . .” And in *In re Seroquel Prods. Liab. Litig.*, 244 F.R.D. 650, 660 n. 6, 662 (M.D.Fla. 2007), the court criticized the defendant’s use of keyword search to select ESI for production, in particular because the defendant failed to provide information “as to how it organized its search for relevant material, [or] what steps it took to assure reasonable completeness and quality control,” and noting that “while key word searching is a recognized method to winnow relevant documents from large repositories . . . [c]ommon sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness.” (Emphasis added.)

- Attorneys would be wise (as a matter of best practices) to sample for privileged documents, so that they are withheld with a 99% confidence level and an error rate of plus or minus 1%;
- Malpractice insurers should be actively revising their applications for errors and omissions insurance to force attorneys to disclose the level of their e-discovery competence, and insurers should be monitoring, if not mandating, the continuing education of attorneys in e-discovery matters.
- Besides being able to choose the format for the production of ESI, requesting parties should be able to designate the search methodologies used by the responding parties to search for potentially relevant documents. Otherwise, responding parties may use key words and search methodologies that under-produce to the requesting party.

The grand conclusion brings us back to loopholes. In an adversarial system, attorneys will exploit loopholes. And now you know that a large technical loophole besets our system. It besets every discovery-based judicial system, and we have not yet faced up to it.

We seek the *truth*. But now that there's so much data, the best we can say about the truth is this: it's *statistical*.

*After graduating with a B.S. in engineering systems from the University of California at Los Angeles (U.C.L.A.), Nick Brestoff earned an M.S. in environmental engineering science from the California Institute of Technology (Caltech) and graduated from the Gould School of Law at the University of Southern California (U.S.C.) in 1975. For the next 35 years, Mr. Brestoff litigated business, employment, environmental, and other civil disputes in state and federal court. He is currently a consultant to businesses and attorneys through International Litigation Services ([www.ilsTeam.com](http://www.ilsTeam.com)). Mr. Brestoff's email address is [nbrestoff@ilsTeam.com](mailto:nbrestoff@ilsTeam.com).*

## Committee Co-Chairs' Message

Dear EDDE Committee Members,

After adventures of the summer, the EDDE committee is poised for action this 2010-2011 season. Congratulations to Eric Hibbard, CTO (Security and Privacy) at Hitachi Data Systems. Eric was named an EDDE Co-Chair this summer, and we expect that his contributions to our committee will continue unabated.

Some old news, followed by an overview of events, activities and meetings planned for this EDDE year.

As many of you know, EDDE held its first eDiscovery and Digital Evidence Practitioners' Workshop at Kent Law School in Chicago in May. This two-day sell-out conference was a first-of-its-kind event, and focused on current and emerging issues facing intermediate to advanced practitioners in the electronic discovery and digital evidence arenas. Our unique format included hands-on workshops, case study presentations, panel discussions and nearly a full day mock hearings involving ESI discovery and digital evidence before nationally renowned federal magistrate judges (thank you Magistrate Judges Frank Maas, Karla Spaulding, Nan Nolan, and Sidney Schenkier). We had more than 20 attorney practitioner presenters, each of whom added their unique practice perspective.

Special thanks to our expert non-attorney conference faculty: Bob Thibadeau (Wave Systems, Inc.), Liam Ferguson (Mesirow Financial), and Lawrence Wolfenden (FBI). Really special thanks to those who took time out of their busy schedules to help organize as well as present: Steve Teppler, Hoyt Kesterson II, Serge and John Jorgensen, Lauren Topelsohn, Eric Hibbard and Lucy Thomson. Many thanks also to our sponsors Autonomy, Clearwell Systems, Sylint, and Hitachi Data Systems. Of course, this event would not have been possible without the support and tireless efforts of our wonderful ABA SciTech staff, Shawn Kaminski, Maria Gamboa, and Julia Passamani, whose tireless efforts made this outstanding program possible. The workshop was conducted both live at Chicago-Kent Law School and streamed over the Internet by our technology partner, ALI-ABA. Thanks also to Chicago-Kent law school for providing the educational locale so uniquely suited to our format.

Our annual pre-RSA Security Conference EDDE meeting in February was a resounding success, and graciously hosted by Joe Burton, who leads Duane Morris' San Francisco office. Attendees included U.S. Magistrate Judges John Facciola and Andrew Peck, and included presentations and panel discussions addressing emerging issues in ESI discovery, ESI search, Fourth Amendment search and seizure, long-term data retention, computer and digital evidence forensics, and a judicial roundtable.

Members of the EDDE committee also presented a variety of ESI discovery and digital evidence track sessions at the RSA Security Conference. This conference is co-hosted each year by the ABA Science and Technology Law Section. We are always honored to present at this conference alongside our elder sister, SciTech Information Security Committee.

The *EDDE Journal* was launched this year with a thought-provoking article by Judge John Facciola. Our most enthusiastic thanks go to Thomas Shaw for his tireless efforts producing a series of outstanding issues. EDDE members contributed many excellent articles on a variety of e-discovery issues. Steve Teppler continued his prolific work producing the *EDDE Digest* so that EDDE members will be the first to learn about the implications of the latest judicial decisions.

*Coming This Year:*

*Second EDDE Practitioners' Workshop* - We're in the planning stages for our second practitioner's workshop, which may be held in both east and west coast law school locations. Stay tuned for announcements. If you're interested in participating, let us know.

*RSA-EDDE Meeting* – We are planning our next RSA-EDDE committee in San Francisco following the RSA Conference (February 14-18); the ABA mid-year meeting in Atlanta immediately precedes RSA this year (February 9-14). Stay tuned for more details. The EDDE tracks for the RSA panels will be announced soon, and we expect to have good representation at the conference.

*Work groups* – We will be announcing the work groups formed at our last meeting. As you may recall, the work groups include:

1. Fed.R.Civ.P. Rule 26(f) Best practices or guidance
2. Fed.R.Civ.P. Rule 26(f) Ethical obligations imposed by the CMC provisions. Analysis of potential divergent approaches (negotiations/puffery permitted similar to settlement meetings, or candor to tribunal/expediting litigation/competency requirements requiring transparency)
3. Model eDiscovery Requests with guidance and commentary
4. Gaming the Search – Requests vs. stipulated key word searches, - identifying and/or disclosing pseudonyms, synonyms, and acronymic substitutes for agreed upon search terms. In other words, while key words may refine the search, do they delimit the search universe to the exclusion of potentially relevant information not picked up as hits?
5. e-Discovery in the Cloud
6. Others being developed

We will select work group chairs to organize and coordinate respective work group efforts, and keep us productive. Please indicate your interest in serving in a leadership position for one or more of these work groups.

*EDDE Member Presentations* – We're extremely proud that our members present on cutting-edge issues at various ESI discovery-related conferences and/or CLE programs. Let us know, and we'll post to the rest of the committee. If any EDDE committee member is teaching a law school course, also please let us know.

Send us your ideas, and stay tuned for updates.

The EDDE Chairs, Eric Hibbard, George Paul, Steve Teppler and Lucy Thomson