AVOIDING AN E-DISCOVERY ODYSSEY

Note: This article first appeared in the Northern Kentucky Law Review, Volume 36 Number 4 (Summer 2009).

Roland Bernier[*]

INTRODUCTION: AVOIDING THE ATTENTION OF THE MODERN-DAY HOMER

> Sing to me of the man, Muse, the man of twists and turns driven time and again off course . . . .[1]

Pity poor Odysseus. Admire him, if you prefer. The heroic warrior, after all, travels through 12,109 lines of poetry in the form of dactylic hexameter, skirting monsters and whirlpools, suffering seven years of imprisonment by the nymph Calypso, fighting pirates, helplessly watching the transformation of half his men into swine by the evil Circe (always a nuisance), being driven half-mad by recklessly exposing himself to the songs of the Sirens, and suffering through at least two ship wrecks.[2] He endures all this after ten years at war, and makes good on his quest to return home to his beloved Ithaca, even if rather belatedly and with the complete loss of his crew.[3]

Odysseus faced outrageous challenges. The attorney navigating his own e-discovery ocean may occasionally feel justified in comparing his own situation to that described by Homer. After all, his own quest is no walk in the wading pool. The waters he must navigate may seem as deep and dark as any Odysseus crossed. The dangers lurking in the hidden eddies and islands, unseen and unknown, are just as treacherous. The navigation across this binary gulf is just as difficult. In response, he should seek to treat his own dangers with more caution and more foresight than did good Odysseus.

One can imagine Hermes, the messenger of the Gods, dispatched from the heights of Olympus to deliver a message of warning to the reluctant captains of pre-litigation. What would he say? Do not become as distracted and unfocused as Odysseus, but rather, be methodical and systematic. Do not give credence to the sirens who insist upon shortcuts. Use all of the devices available to you to

---

[*] Attorney and Consultant with Forensics Consulting Solutions LLP, Mr. Bernier works from his office in Houston where he helps other attorneys and corporate legal departments develop strategies for tackling large, complex discovery projects. Mr. Bernier has worked technology in legal settings for nearly 15 years, first as a computer programmer and database administrator, and now in his current capacity.

[1] HOMER, THE ODYSSEY 1.1-1.2 (Robert Fagles trans., Penguin Books 1996) [hereinafter THE ODYSSEY].

[2] *Id.*

[3] *Id.*

safely navigate your way across the uncharted depths. Unlike Odysseus, whose quarrel with the sea god Poseidon nearly caused his doom, Hermes would admonish those perceptive enough to listen to stay on the good side of the Judges and Magistrates, to whose good judgment, and graces, they entrust not only their own fate, but that of their client as well.

You may pity poor Odysseus for all of the troubles he endured, some of which he obviously brought upon himself, or you may admire him for enduring to the end and finding his way back to his family in Ithaca. Whatever you do, however, do not follow in his footsteps. Cast aside the follies of those who traveled through these treacherous waters before you. Learn from them. Understand what they did wrong, so that you will avoid the same traps. Begin not your own odyssey, but rather, plan a straightforward and uneventful trip; and do so down to the smallest detail possible. It may be tedious, and it certainly will not be easy. It may be boring, but excitement in this area is to be shunned. So feel free to admire Odysseus the adventurer, but do not emulate him.

This article will discuss how to build and implement a discovery process that is systematic and cyclical in nature. It will begin by briefly discussing the huge societal changes that affected discovery and forced changes by the legal profession. It will examine how communication between opposing counsel can thwart surprises and clear some of the challenges from one's path. It will discuss the benefits of obtaining competent technical assistance. The main body of the article then examines the fundamentals of searching and retrieval operations, how to define the term "success" and illustrates how a "process based" approach may be incorporated into the discovery project workflow. The article then concludes by noting the dangers of making decisions on discovery processes in an arbitrary manner.

This article's purpose is to discuss measures that may be taken to strengthen the defensibility of the search protocol. And to keep you off of Circe's island . . . squealing like a pig and rooting around in the mud.

INTERESTING TIMES

"We live in tumultuous times."[4] All around us, change is occurring at a pace heretofore not experienced.[5] The development of computer technology and the global reach of the internet have, among other things, transformed our way of

[4] FRANK WEBSTER, THEORIES OF THE INFORMATION SOCIETY 59 (Routledge Taylor & Francis Group 2d ed. 2002).

[5] *See id.* ("It is widely acknowledged that established relationships [between societal actors] are undergoing major change and that, in addition, the pace of change is quicker than at any time in history."); GEORGE L. PAUL, FOUNDATIONS OF DIGITAL EVIDENCE 3 (2008) (noting that changes attributed to new technologies developed over the last twenty years have "altered commerce, everyday communication, government, public discourse—indeed almost everything," including "civilization's system of writing").

life.[6] Nearly every industry has been affected by the revolution associated with ubiquitous computing[7] and data storage.[8] Though the true extent to the transformation to our society that increasingly powerful computing and global connectivity will reap is still obscured to us, we can find an analogy in our past.

When William Caxton brought the first printing press to England in 1476,[9] Edward IV, "sun of York," [10] was King of England,[11] giving the island kingdom a welcomed respite from the series of skirmishes known as the Wars of the Roses.[12] It would be seven years before Edward's brother, Richard, would ascend to (or usurp) the throne, clearing the way by murdering his two young nephews, if Shakespeare is to be believed. Two years after that, Henry Tudor would defeat the last Plantagenet King at Bosworth Field[13] to establish the Tudor dynasty and end the Wars of the Roses forever.

Edward's government could hardly be said to have foreseen danger or have objected to the press, because the first dated work printed in England by Caxton is said to have been The Dictes or Sayengis of the Philosophres ("Sayings of the Philosophers"), written by Earl Rivers,[14] who was Edward's brother-in-law. Later rulers would see danger in the power of the press.

The impact of the printing press on society is often under-appreciated:

---

[6] *See, e.g.*, Ben Schneiderman, *Universal Usability: Pushing Human-Computer Interaction Research To Every Citizen,* COMM. OF THE ACM, May 2000, at 88 ("[S]tabilizing forces of standard hardware, operating systems, network protocols, file formats, and user interfaces are undermined by the rapid pace of technological change."); Dale W. Jorgenson, *Information Technology and the U.S. Economy* 10 (Harvard Inst. of Econ. Research, Paper No. 1911, 2001), *available at* http://ssrn.com/abstract=257536 (noting that many technologies, "have progressed at rates that outrun even the dramatic pace of semiconductor development").

[7] *See generally* CONCISE ENCYCLOPEDIA OF COMPUTER SCIENCE 303 (Edwin D. Reilly ed., John Wiley & Sons 4th ed. 2004) (noting that a casual description of the term "ubiquitous computing" is "computers everywhere," and that the term is used to describe an environment in which mobile devices that communicate through wireless channels are used); *Id*. at 523-24 (continuing that these devices have functionality similar to that found in a desktop computer, but allow for freedom from geographic constraints to "allow a more effective, convenient, and timely use of computing and communication." Examples are "wearable computers," some of the newest PDA's and cell phones, and devices that use radio frequency identification (RFID) technology).

[8] S*ee generally* Julian Stuhler, *Managing the Data Explosion*, IT-DIRECTOR.COM, Jan. 22, 2009, http://www.it-director.com/technology/data_mgmt/content.php?cid=11025 (describing corporate IT facilities as "drowning in data" and citing a study that estimates that 45 GB of data exists in repositories for each person on earth); Jeremy Kirk, *Data Explosion Shakes up IT*, PC WORLD, Sept. 13, 2007, http://www.pcworld.com/businesscenter/article/137161/data_explosion_shakes_up_it.html (forecasting that in three years "the bytes of data generated by digital cameras, mobile phones, businesses IT systems and devices will equal the number of grains of sand on the world's beaches").

[9] CRAIG JOYCE ET AL., COPYRIGHT LAW 15 (6th ed. 2003).

[10] WILLIAM SHAKESPEARE, THE TRAGEDY OF KING RICHARD III act 1, sc. 1.

[11] MIKE ASHLEY, BRITISH KINGS AND QUEENS 209 (2000).

[12] *Id.* at 21.

[13] PAUL MURRAY KENDALL, RICHARD THE THIRD 443 (W.W. Norton & Co., Inc. 1956).

[14] GERTRUDE BURFORD RAWLINGS, THE STORY OF BOOKS 116 (D. Appleton & Co. 1901); GEORGE HAVEN PUTNAM, 2 BOOKS AND THEIR MAKERS DURING THE MIDDLE AGES 103 (Hillary House 1962) (1896).

> It is almost impossible for us to appreciate the revolutionary impact on people, institutions and governments of the vastly enlarged power granted by the invention of printing to disseminate ideas and criticisms, including anonymous attacks on those in power, and to communicate to large numbers of people over extensive areas.[15]

The commentator contends that the capture of these abilities altered humans in a fundamental way. The technologies behind the information age, like the printing press, have been triggers for sweeping changes. They also are being credited for changes that have yet to occur. Jeffrey Cole, director of the Center for the Digital Future, a study on attitudes of users toward the World Wide Web, stated:

> "The internet has become an essential source of entertainment, information and communication... However, in 2006, we are beginning to measure real growth and discover new directions for the internet as a comprehensive tool that Americans are using to touch the world."[16]

The internet has become much more than a method for distributing or creating works: it is becoming part of the very fabric of life.[17] This revolution is still happening, and it is altering the world as dramatically as did the printing press. The difference is that in this day and age those alterations are taking place much more quickly.[18] It's as though we are watching events through time-lapse photography.

The spread of computers, the advent of ubiquitous computing, and the connectivity provided by the internet have combined to generate not only transformative, but also disruptive forces on society. The authors of *Unleashing the Killer App: Digital Strategies for Market Dominance*,[19] have a rule they call the Law of Disruption. This law postulates that technology progresses at an exponential rate while society changes at an incremental rate.[20] Because social norms and attitudes often lag behind the change, they outlast the very environmental and technological factors that fostered or sheltered them in the

---

[15] David Barnhizer, *Cyberpersons, Propertization, and Contract in the Information Culture: Propertization, Contract, Competition, and Communication: Law's Struggle to adapt to the Transformative Powers of the Internet*, 54 CLEV. ST. L. REV. 1, 12 (2006).

[16] BBC, *Virtual Pals Soar in Importance*, (Nov. 30, 2006), *available at* http://news.bbc.co.uk/2/hi/technology/6158935.stm.

[17] *See, e.g.,* JOHN HORRIGAN & LEE RAINEY, PEW INTERNET & AMER. LIFE PROJECT, THE INTERNET'S GROWING ROLE IN LIFE'S MAJOR MOMENTS, (2006), http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Major%20Moments_2006.pdf.pdf (examining the ways in which a growing number of people are using the internet to help them during important milestones in their lives).

[18] See Moore's Law, http://www.intel.com/technology/mooreslaw/index.htm (last visited April 19, 2009). The law, created by Intel icon Gordon Moore in 1965, predicted that the number of transistors on a chip will double about every two years and describes an exponential increase in computing power.

[19] LARRY DOWNS & CHUNKA MUI, UNLEASHING THE KILLER APP: DIGITAL STRATEGIES FOR MARKET DOMINANCE 8 (Harv. Bus. Sch. Press 1998).

[20] *Id.*

first place.[21] This is one explanation for the legal industry's attitude toward technology, which has been decried for being increasingly disconnected with needs of Information Age clients.

Nevertheless, the legal community has attempted to address the effect of technology on discovery issues. In August 2004, an advisory committee published a proposed set of amendments for the Federal Rules of Civil Procedure designed to guide courts and attorneys on issues associated with electronic discovery.[22] The committee passed a revised set, and ultimately these were adopted by the U.S. Supreme Court without any substantive modification.[23] The newly enacted Federal Rule of Evidence 502 was passed, in part, to stem concerns that productions of electronically stored information (ESI) were vulnerable to inadvertent production of privileged material and a resulting waiver of privilege.[24]

DETERMINING OBJECTIVES: SURVEYING THE LANDSCAPE

> We reached the Aeolian island next, the home of Aeolus, Hippotas' son, beloved by the gods who never die—a great floating island it was . . . .[25]

As did mariners in the time of Odysseus, attorneys seeking accurate maps to help guide their way may search in vain. Case-law is still evolving, while technology standards provide variety, even if they happen to be followed. Technological progress, while holding out the promise of empowerment, can also offer up surprises. Homer described the island Aeolia as floating freely (although it isn't clear if it floats in the air or the sea).[26] Odysseus thus had to not only concern himself with the very real difficulties of trying to navigate through difficult hazards in relatively uncharted territory, but had to deal with a landscape that actually shifted unpredictably. E-discovery offers similar opportunities for surprise. The territory underfoot can change suddenly, and the unprepared can be thrown off balance. Fortunately, there are ways to limit such situations. The most effective measure is the simple, though not always easy, act of communicating with the other side.

The lack of civility in the modern-day practice of law between opposing attorneys has been a source of running commentary in legal literature for some

---

[21] *Id.*
[22] Gil Keteltas & John Rosenthal, *Discovery of Electronic Evidence*, *in* ELECTRONIC EVIDENCE: LAW AND PRACTICE 10 (Paul R. Rice ed., 2008).
[23] *Id.*
[24] Shira Sheindlin & Jonathan M. Redgrave, *Special Masters and E-Discovery: The Intersection of Two Recent Revisions to the Federal Rules of Civil Procedure*, 30 CARDOZO L. REV. 347, 364 n.84 (2008).
[25] THE ODYSSEY, *supra* note 1, at 10.1.
[26] *Id.*

time.[27] E-discovery has been the impetus for the emergence of a growing body of literature which calls for not only early communication by the parties, as is required in the rules,[28] but actual collaboration between the parties and their attorneys.

The Sedona Conference[29] has recently released a memo urging greater cooperation between attorneys during the discovery process.[30] The main theme expressed in the memo is that zealous advocacy does not require counsel to engage in uncooperative behavior during the discovery process,[31] and that the Federal Rules of Civil Procedure mandate cooperation.[32] Other commentators have pursued the point,[33] while case law has begun to speak to the issue with greater assertiveness and frequency.[34]

---

[27] *See, e.g.*, Nancy B. Rapoport & Roland J. Bernier, (*Almost) Everything We Know about Pleasing Bankruptcy Judges We Learned in Kindergarten*, 27-6 AM. BANKR. INST. J. 16, 74 (2008) (stating "[c]ivility is still a problem, even after decades of complaints about the decline in civility in the legal profession."); Alex J. Hurder, *Lawyer's Dilemma: To Be Or Not To Be A Problem-Solving Negotiator*, 14 CLINICAL L. REV. 253 (2007) (discussing an attorney's difficulties in assuming the role of problem solver rather than strictly an advocate); Jeffrey P. Hopkins et al., *Lawyers in the Hot Seat: The State of Ethics & Professionalism*, 6 DEPAUL BUS. & COM. L. J. 557 (2008) (examining public perception of lawyers in general, and specific instances of uncivil behavior by attorneys in legal proceedings). *But see* Mark Neal Aaronson, *Be Just to One Another: Preliminary Thoughts on Civility, Moral Character, and Professionalism*, 8 ST. THOMAS L. REV. 113, 117-18 (1995) (arguing that the part played by uncivil behavior by lawyers on the public's declining opinion of lawyers has been exaggerated); David B. Wilkins, *Practical Wisdom for Practicing Lawyers: Separating Ideals from Ideology in Legal Ethics*, 108 HARV. L. REV. 458, 459 (1994).

[28] FED. R. CIV. P. 26(F)(1) (calling for parties to meet and "confer as soon as practicable"). The rule also includes a list of items parties should consider during the conference. *Id.*

[29] The Sedona Conference, http://www.thesedonaconference.org/. The Sedona Conference is an organization dedicated to exploring issues on "antitrust law, complex litigation, and intellectual property rights." *Id.* Its publications have been commonly cited in court opinions on electronic discovery or ESI. *See generally* John B. v. Goetz, 531 F.3d 448, 459 (6th Cir. 2008); Regan-Touhy v. Walgreen Co., 526 F.3d 641, 649 (10th Cir. 2008); Mancia v. Mayflower Textile Servs. Co., 253 F.R.D. 354, 364 (D. Md. 2008).

[30] THE SEDONA CONFERENCE, THE SEDONA CONFERENCE COOPERATION PROCLAMATION (July 2008), *available at* http://www.thesedonaconference.org (follow "Cooperation_Proclamation.pdf" hyperlink).

[31] *Id.* at 1.

[32] *Id.* at 2.

[33] *See* Why E-Discovery is Ruining Litigation in America and What Can be Done About It, http://ralphlosey.wordpress.com/?s=why+e-discovery+is+ruining+litigation (Dec 7, 2008, 19:46 EST) (stating that one requirement for ensuring "reasonable" and "affordable" e-discovery projects is "strategic cooperation [between opposing attorneys] in technical areas of discovery to avoid unnecessary disputes").

[34] *See* Mancia v. Mayflower Textile Servs. Co., 253 F.R.D. 354, 359 (D. Md. 2008); Securities & Exchange Commission v. Collins & Aikman Corp., No. 07 Civ. 2419(SAS), 2009 U.S. Dist. LEXIS 3367, at *30 (S.D.N.Y. Jan. 13, 2009); Aguilar v. Immigration & Customs Enforcement Div., 255 F.R.D. 350, 358-59 (S.D.N.Y. 2008). *But see* Cason-Merenda v. Detroit Med. Ctr., No. 06-15601, 2008 U.S. Dist. LEXIS 94028, at *17-18 (E.D. Mich. Nov. 12, 2008) (arguably punishing a party for early cooperation by effectively voiding an agreement to defer motions on cost-shifting). It is hoped that such cases represent an anomaly in the general trend lines.

The bottom line is that by negotiating on production format,[35] search methodologies,[36] and provisions of production protocol as early as possible, many of the perils of e-discovery can be avoided by both parties. Parties can discuss such items as native file productions vs. tiff or pdf files; whether or not to allow for de-duplication or near de-duplication; what specific legal database application load files to include in productions; what search criteria should be used; what processes may be used to verify search results; provisions for "clawing back" documents inadvertently produced. Parties can reap great advantages by agreeing to such terms early in the process.

A "collaborative" relationship between opponents in a legal matter is probably too much to expect. It is however, "an affirmative duty to engage in pretrial discovery in a responsible manner that is consistent with the spirit and purpose of" the relevant rules of procedure.[37] By opening the lines of communication, and taking reasonable stances, and refraining from "obstructionism," both sides can keep a trial's discovery phase on course, and keep the terrain from shifting and islands from floating away.

FINDING A NAVIGATOR

> Now bright-eyed Athena sent them a stiff following wind blustering out of a clear sky, gusting on so the ship might run its course through the salt sea at top speed . . . .[38]

Homer relates in his lost poem, *Cypria*, that Odysseus was so reluctant to leave his home to go to war at Troy that he feigned madness in an unsuccessful attempt to opt out. Likewise, attorneys are often dragged, kicking and screaming, into e-discovery.

This reluctance is not surprising. Attorneys are not known as particularly computer literate, and the case law makes this area volatile in comparison to other substantive areas. Furthermore, there is constant pressure from clients to

---

[35] "Production format" refers to the manner and format in which documents and electronically stored data is produced to opposing counsel. It may involve the conversion of electronic data to a paginated, image based format that simulates traditional paper-based exchanges, an exchange of electronic data, an exchange of paper documents, or some combination of all three.

[36] Negotiating the elements of any search criterion can remove the specter of a court finding that search criteria executed prior to production was "not reasonable," and therefore expose the client to possible additional costs or penalties. *See* Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 259-60 (D. Md. 2008) (questioning attorneys' and client's qualifications to formulate a valid search criterion); United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) ("Given [the] complexity [involved in crafting searches designed to return relevant documents], 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. This topic is clearly beyond the ken of a layman…."). *But see* Rhoads Indus. v. Bldg. Materials Corp. of Am., 254 F.R.D. 216, 220-21 (E.D. Pa. 2008) (criticizing *Victor Stanley, Inc.* for its reliance on hindsight).

[37] FED. R. CIV. P. 26(g) advisory committee's note.

[38] THE ODYSSEY, *supra* note 1, at 15.325-28.

reduce costs while still serving up a defensible process. The problem is, though, that there really is no escape. ESI is becoming a standard format for evidence in even "small" cases. Electronic records are literally everywhere, and in order to practice today, an attorney needs to understand where these sources of data (read: of evidence) are. A litigator wouldn't dream of going into a courtroom without a good knowledge of the rules of evidence and procedure. More and more, ESI is a large component of both bodies of rules.

At least one commentator has argued that attorneys today are incompetent to ply their trade due to the combination of their general computer illiteracy and the pervasive nature of electronic evidence.[39] Judicial opinions have expressed distrust of protocols formulated by attorneys, and even judges.[40] Even as younger litigators emerge from the generation some label "digital natives," and hiring practices are altered to reflect the evolving environment of digital evidence, attorneys will need to rely on knowledgeable people to help them navigate what can be treacherous waters. Even those attorneys well-versed in the basics of information technology will need experienced consultants on a fairly frequent basis.[41] Finding help is often necessary, but the manner in which the decision to hire a consultant or service provider is also a key point in the discovery process. Because a consultant's or service provider's qualifications and opinions may be subjected to scrutiny, it makes sense for a competent and diligent attorney to do his or own vetting.[42] Furthermore, should a consultant's qualifications appear to be lacking, then doubt is cast upon the process undertook to hire that person, and an attack on the attorney's vetting process becomes a distinct possibility.

---

[39] Why E-Discovery is Ruining Litigation in America and What Can be Done About It, *supra* note 33. Losey states "[t]o put it bluntly, most of us trial lawyers are not fully competent to practice law in today's digital age of terabytes of potential evidence. Most of us do not know how to do e-discovery in an efficient and cost effective manner in that kind of an environment, much less employ effective quality control procedures." *Id.* He continues by excoriating those who "spin a web of pseudo-competence." *Id.*

[40] *See Victor Stanley, Inc.*, 250 F.R.D. at 256 (questioning the qualifications of counsel and their client to design "a search and information retrieval strategy that could be expected to produce an effective and reliable privilege review"); Am. Nat'l Bank & Trust Co. v. Equitable Life Assurance Soc'y, 406 F.3d 867, 879 (7th Cir. 2005) (finding fault with a magistrate judge's "sampling procedure for sanctioning" one of the parties for failure to produce some documents during discovery); *O'Keefe*, 537 F. Supp. 2d at 24 (concluding that certain types of analyses of search criteria falls within the area of expert witness, governed by FED. R. EVID. 702); Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008) ("[D]etermining 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 requires expert testimony . . . .").

[41] One reason for this is that the nature of e-discovery often requires that expertise from several different technical disciplines. In one discussion, a computer forensics expert may be required, but in a different point, the skill possessed a database administrator, programmer, or storage expert may be needed.

[42] Of course, should counsel be required by a client to work with a specific consultant or service provider, the responsibility for vetting should be, at the very least, diminished.

Hiring a consultant does provide certain advantages. The most obvious (and expected) consequence is that a competent advisor may assist in conducting a well-designed and properly executed discovery project. The presence of a consultant or expert may work to shift the burden of competence from attorneys and clients to the expert.[43] However, the mere fact that an expert makes a statement or employs a certain protocol does not necessarily mean that the court need accept it without some examination.[44]

The lesson is to hire an expert that can be trusted; one who can help attorneys and their clients avoid pitfalls. So like Odysseus, the attorney is forced to undertake the journey. The hope is that his will be a safer, shorter journey than was accomplished by our hero from Ithaca.

CHARTING THE COURSE

> The rest of the winds she stopped right in their tracks, Commanding them all to hush now, go to sleep.[45]

Recent court decisions have placed a greater level of scrutiny upon search protocols used during electronic discovery projects.[46] These decisions call for better crafted search protocols, and specifically call for the use of sampling techniques and methods for confirming the accuracy of initial search criteria.[47] In other words, there's no forgiveness for merely plotting your course and blindly sailing on without ever taking your bearings and making needed adjustments. You need to be able to make sure you stay on course, and make corrections when you find that you've strayed, or when the terrain proves not what was previously assumed. This article will propose a workflow derived from principles gleaned from court decisions that have cast baleful eyes upon inadequate attempts at building search criteria, and appropriate principles now need to be adopted into the e-discovery space. It will help you construct a road map, and adjust to terrain that was previously uncharted territory.

---

[43] Rhoads Indus. v. Bldg. Materials Corp. of Am., 254 F.R.D. 216, 221 (E.D. Pa. 2008).

[44] Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 261 (D. Md. 2008) ("Moreover, if the court is to be given scientific or technical information to resolve a contested discovery matter, what standards should govern its evaluation? Should the court ignore a purported ESI expert's lack of qualifications if that shortcoming is demonstrated by the party opposing his opinion? Should the court accept opinions shown to be unsupported by sufficient facts or based on demonstrably unreliable methodology? The answer is obviously 'No.'").

[45] THE ODYSSEY, *supra* note 1, at 5.422-23.

[46] *See e.g., Victor Stanley, Inc.*, 250 F.R.D. at 261; United States v. O'Keefe, 537 F. Supp. 2d 14, 23-24 (D.D.C. 2008); *Rhoads Indus.*, 254 F.R.D. at 221.

[47] *See Victor Stanley, Inc.*, 250 F.R.D. at 256-57 (noting that no assertion was made that any sampling was conducted to defend producing party's search protocol, which the court ruled was not reasonably undertaken, causing a waiver of privilege for any privileged documents erroneously produced to opposing counsel as a result of the search filter).

THE SEARCH PROTOCOL: CALIBRATING YOUR COMPASS

> if you only knew, down deep, what pains are fated to fill your cup before you reach that shore, you'd stay right here . . . .[48]

The reader of Homer's Odyssey is introduced to the adventure medias res; that is, he or she is dropped into the middle of the action. This is not unlike the feeling often accompanying a lawyer who must begin a case that promises a large discovery phase . . . the action starts immediately. Or at least, it should. Getting a handle on the scope of e-discovery associated with a case is one of the first important accomplishments; it helps to reduce the intensity of the inevitable scramble as deadlines approach.

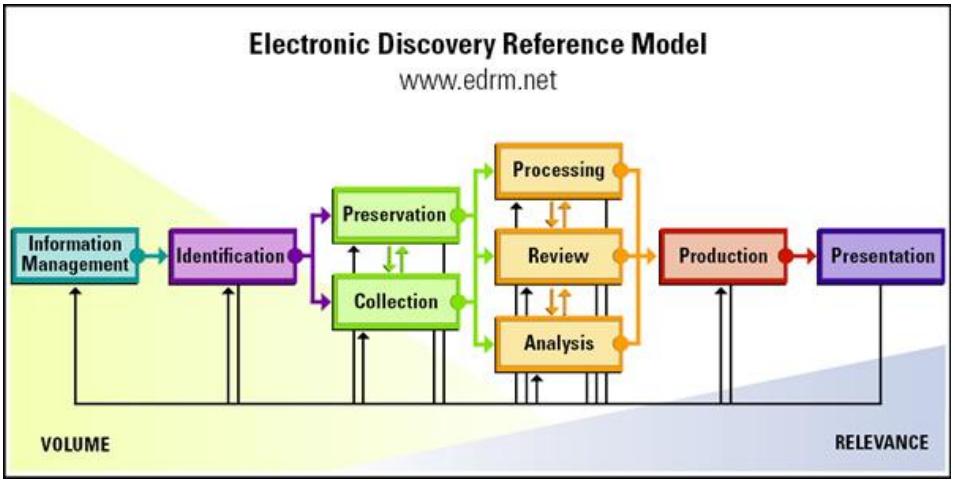
The search protocol, sometimes referred to as the "filter," possesses a central place in the e-discovery workflow. In the most common scenarios, it is the search criteria that determine which documents will be reviewed, and which will not. Thus, the "filter" makes an essential determination before review ever occurs: it says that the excluded documents are either not relevant to the case, or they are relevant, they are inaccessible. In some situations, the search protocol is solely responsible for identifying documents that are reviewed for production to opposing counsel. In others, the search protocol substantially determines what documents will be produced, as no, or only limited, review is conducted. Any component making such a significant contribution to the process must certainly be subjected to careful design considerations, periodic verification as the project progresses, and scrutiny by the court and opposing counsel. As Odysseus would see it, this is your compass, your way of keeping tabs over which seas you sail, which locales you visit.

COMPONENTS OF THE SEARCH PROTOCOL

The search protocol is more than just a list of keywords or date ranges. This article uses the term to encompass all steps used by which one begins to examine a data universe for discoverable documents. Thus, the protocol has elements in most of the pre-review phases of e-discovery. A common model often adopted to illustrate e-discovery phases, created by the Electronic Discovery Reference Model (EDRM) is below.

---

[48] THE ODYSSEY, *supra* note 1, at 5.228-31.



EDRM (chart used with permission of EDRM, at http://www.edrm.net/)

Elements of a search protocol may be applied at any point from the "Identification" stage through the "Processing, Review and Analysis" block illustrated in

EDRM (chart used with permission of EDRM, at http://www.edrm.net/)

. The first elements of a search protocol could therefore be said to consist of those steps instrumental in ascertaining sources of data likely to contain relevant documents. Some of these elements are discussed below.

WHAT IS SUCCESS?

What is meant by the success of an e-discovery project, and particularly that of its search protocol? Culling rates are often used by vendors to define, or at least illustrate, success. Clients and attorneys tend to like this measure, as it identifies, in an easy way, a cost savings. If you culled 60% of documents from a population, then that is, roughly, a 60% savings in attorney review time, with attendant reductions on certain costs associated with production and related processes. It must be emphasized, however, that purely discussing the success of a search protocol in terms of how much data is filtered from the eyes of reviewers is a grossly inaccurate method, especially if a significant quantity of documents relevant to the case could be found in the excluded document set. Even clients would be unhappy to learn that relevant documents favorable to their cause had been "left on the table," sacrificed on the altar of economy. Furthermore, unless given clearance by the court or agreed to by opposing counsel, limiting a document review or production strictly due to volume is impermissible.

To truly define what we mean by success, we need to define the ultimate goal. The goal from an operational standpoint is to supply counsel with the most "accurate" data set possible. What is meant by the term, "accurate?" One could

define the term thusly: a set of documents including all those relevant to the matter whether privileged or not, and excluding all those not relevant to the matter. This is, of course, misleading in its simplicity.

Because the process leading to the formation of a data universe for review or production occurs in multiple steps, this article takes an expansive view of the definition of search protocol, going well beyond the meaning implied by the term "filter." The search protocol, for the purposes of this article, includes all those components designed to remove all irrelevant material from the review corpus, while preserving that data which is relevant to the matter. It will be useful to examine traditional models defining search accuracy.

RECALL AND PRECISION

Traditionally, information retrieval is divided into two steps: 1) formulating a question aligned with a "specific information need;"[49] and 2) translation of language describing that need into a query appropriate for a particular retrieval technology.[50] Two terms often used to measure the success of a retrieval transaction are precision and recall.[51] Precision may be defined as the proportion of documents that are retrieved that are relevant.[52] Recall may be defined as the proportion of relevant documents that are retrieved.[53] Perhaps the concept can best be described by inventing an example. A user executes a search. The search returns 10 documents, which the user codes as either "R," for "relevant" or "N," for "Not Relevant." The user ends up categorizing 6 documents as "R," and 4 as "N." The precision of the search is 60%, because that's the proportion of the documents returned by the query that were relevant. Recall measures the effectiveness of the search in getting all the relevant documents from the enterprise. Let's return to our search. We returned 10 documents, 6 of which were relevant. If there are 20 total documents in the enterprise, and 12 of them were relevant, then our query found one-half of the documents we were looking for. Recall = 50% (6/12). I have seen the concepts of precision and recall illustrated in a fashion similar to Figure 2.[54]

---

[49] ROBERT R. KORFHAGE, INFORMATION STORAGE AND RETRIEVAL 191 (1997).
[50] *Id.* at 192.
[51] *Id.* at 194.
[52] *Id.*
[53] *Id.*
[54] Patrick Oot, Director of Electronic Discovery and Senior Litigation Counsel, Verizon Legal Department, Panel Discussion and Presentation at LegalTech West 2008: Searching and Sampling ESI (June 26, 2008).

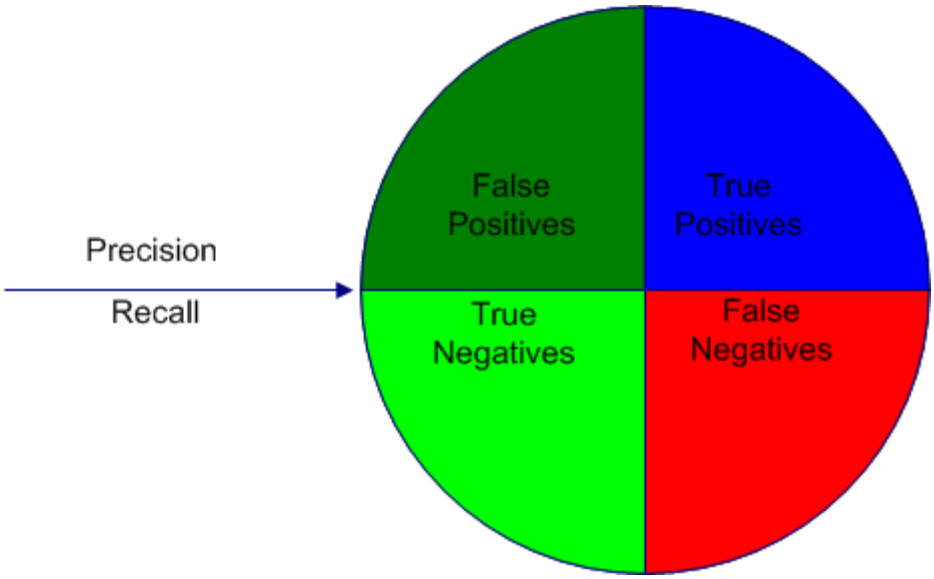


Figure 1: Precision and Recall as Often Depicted

In this diagram, the term "false positives" describes those documents returned that are not relevant. "True Positives" denote those documents returned by the query that the user hoped to see. The six documents our user above coded as "R" would fall into this sector. This sector would be associated with the 60% in our description above. The "False Negatives" label here is used to describe documents not returned that were relevant to the user's need. The "true negatives" were those rightly excluded. I feel these labels are slightly misleading. I would tend to associate the term "False Positives" with a document returned by the search despite not meeting the criteria. In other words, a search for a document containing the word "dog," returns a document without the term. To me, *that* constitutes a false positive; a technical error based on a faulty index or discrepancy within the search engine component itself. If the document is correctly returned by the search, yet isn't Relevant to the matter, then the reason lies within imperfections in the criteria, and our terminology should reflect this. So, although I hesitate to get lost in the minutiae of syntax, I propose to use different terms. These would be Returned and Relevant (RR), Returned and Not Relevant (RNR), Not Returned and Not Relevant (NRNR), and finally Not Returned and Relevant (NRR). A diagram using this these terms appears below.

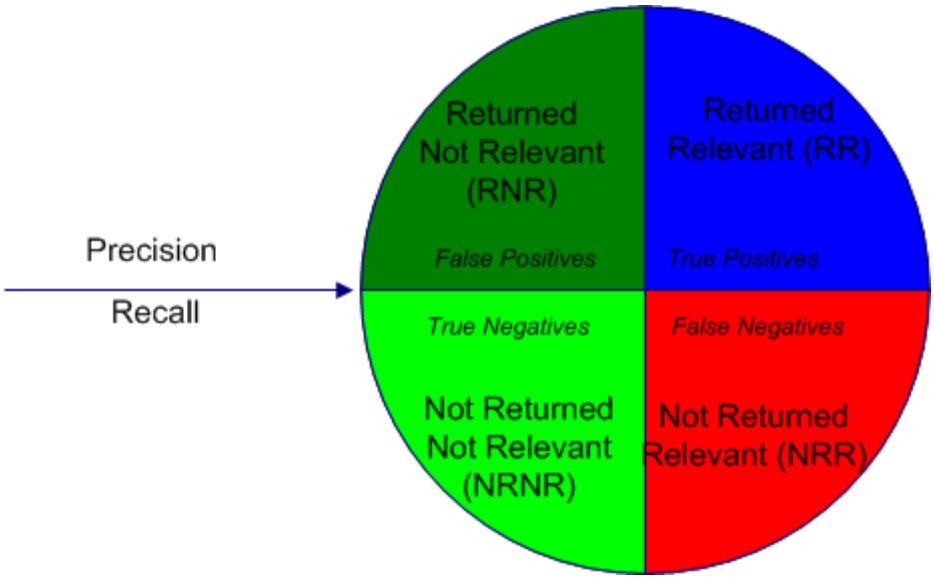


Figure 2: Precision-Recall, Slightly Reinvented

It is often stated that the goal of Precision is to reduce the number of documents that are returned not relevant, and the goal of Recall is to reduce the number documents that are not returned and relevant.[55] Another way to illustrate the concepts above is by use of a matrix.

	Not Retrieved	Retrieved	**Total**
Relevant	3	7	**10**
Not Relevant	5	5	**10**
Totals	**8**	**12**	

Table 1: Precision-Recall Matrix

Precision and Recall are generally expressed in fractional units between 0 and 1.[56] Precision would be expressed mathematically as 7/(7+3) = .7. Recall would be 5/(5+5) = .5. Two other measures should be mentioned as well. Fallout is the measure of non-relevant documents existing in the data universe

[55] *See id.* (Mr. Oot, however, used the terminology of false positives, true negatives, etc.).
[56] KORFHAGE, *supra* note 49, at 195.

that were not returned.[57] That would be 3/(20-8) = .25. Finally, Generality measures the proportion of relevant documents existing within the data universe.[58] Here, there are 12 relevant documents in a universe of 20. 12/20 = .6.

These concepts are, obviously, best suited for environments that are more controlled and are better organized than most commercial IT enterprises. Nevertheless, by understanding them, an attorney can become more effective when managing the technical sphere of discovery.

PLANNING THE PRODUCTION

Among the producing party's foremost concerns should be drafting a plan that is sufficiently thorough to satisfy valid document requests from the opposing counsel.[59] Avoiding undue conflict and the prospect of court-imposed sanctions due to "missing" or "untimely produced" documents is necessarily an important goal. A particular point of possible contention is in determining the nexus between the legal obligation to produce relevant documents and the effectiveness of the document review process itself.

The requesting party will often press home the point that the legal obligation is not excused merely because an otherwise responsive document fell outside of the boundaries of the document search plan. While no court expects 100% of all relevant documents to be found and produced, it will certainly expect that any plan for searching for and reviewing documents across a data enterprise is competently formulated, is reasonably effective, and can be defended by its creators.[60] This is one area where an agreement at the beginning of discovery may help avoid disputes later.[61]

Although not the only model available, the basic planning model used in this article is based on a typical sequence of activities for discovery projects in which the author has been involved. This type of production is based on the concept that documents found in identified sources and returned by the filter will be reviewed for production. The search protocol in this model is designed to find documents to produce. As documents are categorized by reviewing attorneys, the size of the production grows. The accumulation of responsive, non-

---

[57] *Id.* at 196.
[58] *Id.*
[59] At least to an objective party. This is not to suggest that attorneys go out of their way to placate opposing counsel; that would be a self-destructive goal. Rather, the aim is to comply with the objectives of the discovery rules in creating an effective plan, competently executing that plan, and being able to defend both the plan's design and implementation.
[60] *See* Balboa Threadworks, Inc. v. Stucky, No. 05-1157-JTM-DWB, 2006 U.S. Dist. Lexis 29265, at *15–17. (D. Kan. Mar. 24, 2006) (discussing the formulation of a search protocol and suggesting counsel be prepared to discuss any disputes arising over the creation of a keyword list).
[61] This is sometimes easier said than done. A party with little in the way to produce provides its opponent with little leverage. If, however, both sides will be forced to preserve, collect, and produce significant volumes of material, then it is often easier to find common ground.

privileged documents as discovery progresses can be illustrated visually as a triangle (**Figure 3**).

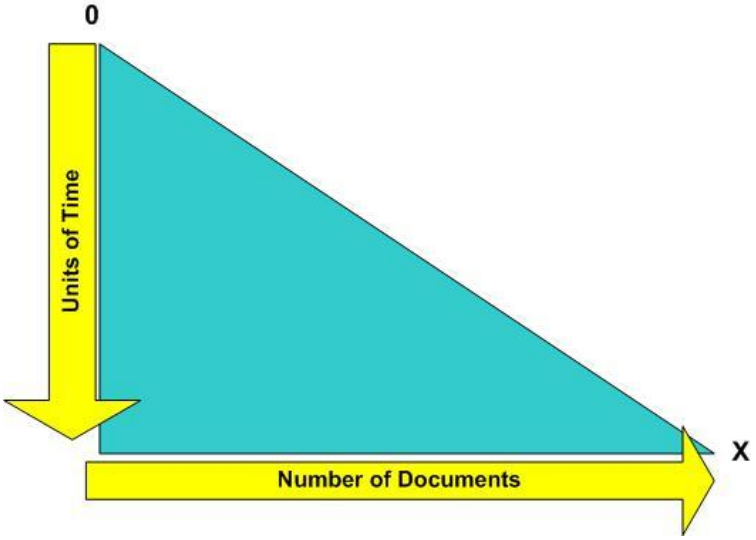


Figure 3: Increasing Production Volume as a Function of Time

The opposite approach begins with the expectation that all files from specific sources will be produced except for those that are returned by the filter ultimately found to be privileged. In that case, the filter exists to find documents that should be removed from the pending production. This process is illustrated below (in Figure 5).

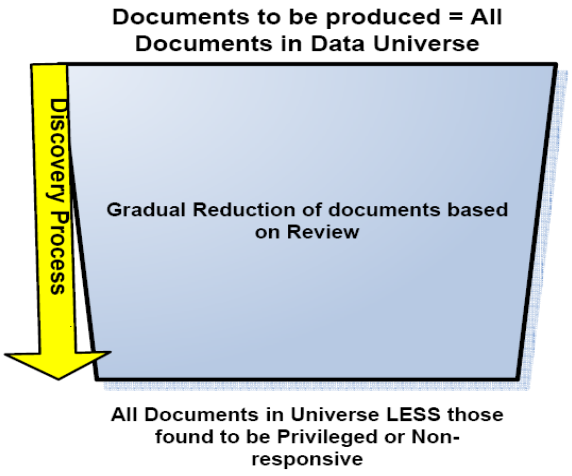


Figure 4: Removing Documents with the Filter

Upon receiving requests for documents associated with current or potential litigation, the initial step in extracting the data is the identification of those sources where relevant data resides.[62] At this point preservation policies have been implemented, so many, if not all, of the sources have been identified. Concurrent to source identification will be the formulation of a search criterion. Primarily consisting of keywords and date ranges,[63] the criterion acts as a filter that "catches" documents falling within its boundaries while allowing all other data to remain outside of consideration. After formulation of the criterion, the filter is then applied to the data, resulting in a review concentrating on documents whose contents fall inside the boundaries of the filter. During the review, attorneys and legal assistants categorize the returned documents for relevance and privilege. Periodically, the results of the review process should be examined for additional sources and modifications to the criteria. Finally, those documents chosen as relevant and not privileged are packaged into an appropriate format and produced to opposing counsel. The process as typically implemented is illustrated below (in Figure 5).

Another common approach occurs when a larger universe is subjected to a query designed to responsive documents. The resulting subset (call it presumptively responsive) is then searched by a second query looking for privileged documents. Documents returned by that query (call them presumptively privileged) are then reviewed or withheld, and what's left of the presumptively responsive set is then produced without being subjected to "eyes on" review. In this case, only the presumptively privileged document set is subjected to human review and categorization. The initial stage consists of activities such as learning of and understanding the dispute, identifying the potential issues of concern, formulating strategies and goals.

Identification of sources is necessary for both implementing a proper document preservation protocol and for targeting documents for possible production.[64] In the "top down" production model, where review is used to

---

[62] These data sources are sometimes referred to as "custodians."

[63] Keywords and dates are the most common components in search criteria, although there are additional advanced techniques. There are commentators who urge that "more creative" techniques be looked upon favorably by courts. *See* George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J.L. & TECH. 10, 2-3 (2007), *available at* http://law.richmond.edu/jolt/v13i3/article10.pdf (concluding that "litigation, as we have known it, is threatened by information's new hyperflow[,]" while urging that "lawyers and judges…be far more tolerant of using [advanced] techniques . . . as part of a reasonable search process" even in the absence of "adequate" testing).

[64] One of the obligations is to identify "key players" and look at the data they create and store. *See, e.g.*, Zubulake v. UBS Warburg, 229 F.R.D. 422, 432 (S.D.N.Y. 2004) (stating that becoming familiar with the client's information system "involve[s] communicating with the 'key players' in the litigation, in order to understand how they stored information" and citing "[c]ounsel's [d]uty to [l]ocate [r]elevant [i]nformation"); Jeffrey S. Follett, *Hold Everything? Litigation Response In The Electronic Age*, 747 PLI/LIT 237, 249 (2006) (emphasizing that data belonging to "key players" should be most carefully scrutinized, and that "key employees" should be made aware of their

accumulate responsive documents, this is the first tool in filtering out non-responsive documents.

Again, formulation of the search criteria is usually based on key words and dates. The process may also involve utilizing special types of queries or search protocols. The scope of the search can be narrowed or widened using various techniques, depending on the search technology employed.

The review stage is used to categorize the documents in various ways. It necessarily identifies responsive documents, and will usually be used to compile a "privilege log." It may also be used to identify other aspects, such as "hot docs," or other customized categories.

Note that the process detailed in Figure 5 does not contain a process for adjusting either search terms or sources from the review process. Although this is often incorporated by informal methods, it is also often ignored. Formally including a process by which review informs the upstream processes and allows for adjustments to those processes allows for a more accurate and flexible production plan. It assists the analytical processes involved in discovery by making a determination on the effectiveness of the initial choices for sources and criteria, and may provide information necessary for modifications. It also may make the plan more defensible to the court, and may help to defuse disputes, or perhaps gain leverage, with opposing counsel.

---

obligations"); Mafé Rajul, "*I Didn't Know My Client Wasn't Complying!" The Heightened Obligation Lawyers Have To Ensure Clients Follow Court Orders In Litigation Matters*, 2 SHIDLER J.L. COM. & TECH. 9 (2005) (describing the act of "speaking with every key player involved in the litigation" as an "affirmative reasonable step").

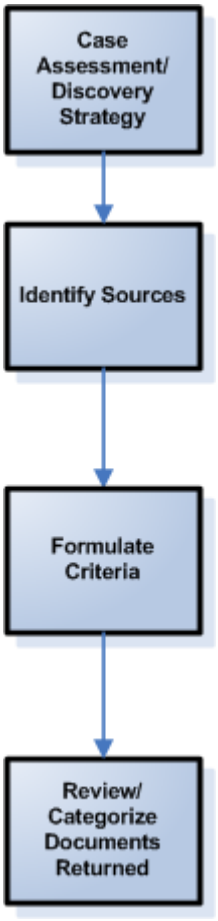


Figure 5: A typical production planning process

Below, we tinker with this model to try to render it in a more defensible form, including adding some elements not present from that above that allow for some verification of the effectiveness of parts of the plan, and pointing out where modification may increase the efficiency of the plan.

HITTING THE BULL'S EYE

Measuring accuracy within the controlled environments of library bibliographic systems is challenging in itself. When we discuss those same measures in less structured environments that make up the IT systems of corporations and government agencies, and similar institutions, the challenge becomes ever more daunting.

Fortunately, responding parties[65] are not expected to match levels of accuracy comparable to a modern library. They should, however, be able to justify the elements incorporated into the design of the search protocol, whether those elements were reasonable, and whether the protocol as a whole was designed with sufficient care and diligence. Failing to address areas in the process leaves it vulnerable to attack, and an attack by opposing counsel is a much more serious affair than is an irritated reader who can't find what he wants in a library's archives.

Another way to view the components making up the steps to production is to see the process in layers. It is possible to perceive that there is a universe of relevant documents existing on the IT infrastructure. This universe of documents is the initial target, and is our first layer. We can imagine it as a circle, resembling Figure 7 below.

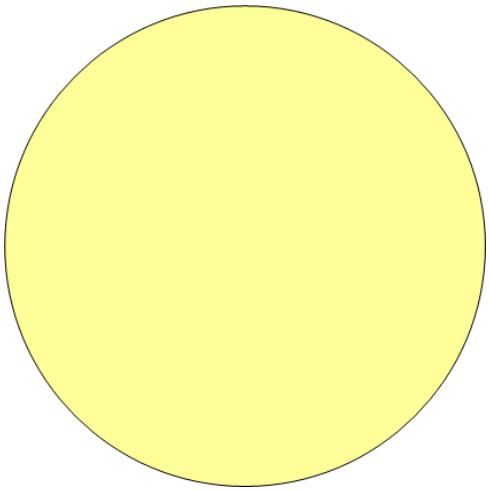


Figure 6: The Universe of Relevant Data

The next layer is comprised of the universe of documents that is responsive to opposing counsel's discovery requests. It is illustrated in such a manner to reflect the fact that these requests, while encompassing much of the relevant universe, is often overbroad, and strays beyond those boundaries. This is accounted for in the territory beyond the dotted line. So,

---

[65] This article uses the term "responding party" to denote the party required by the court to turn over documents and data to the opposing side. This allows us not to get caught up in non-sequiturs involved with identifying parties as plaintiffs or defendants, and tells readers that this is the party who has the obligation in E-Discovery. These obligations, of course, are often so divorced from the underlying issues in the case that the titles dependent upon those issues are usually rendered meaningless for the purposes served here. For the same reasons, I usually label the party who has asked that the data be produced the "requesting party."

Figure 7 shows the presence, really, of three layers: the total universe of relevant material; the population of material in discovery requests that is relevant, and the set consisting of material in discovery requests that is not-relevant (or is overbroad).

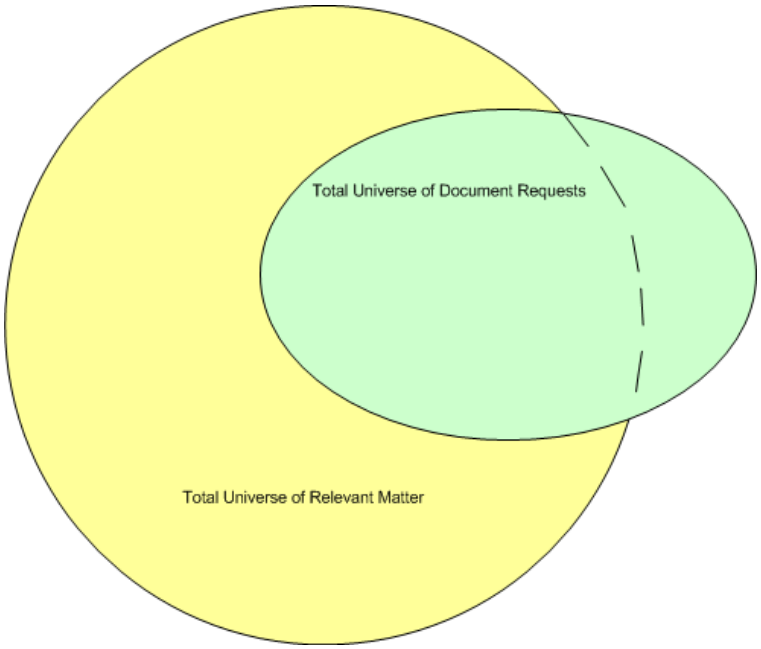


Figure 7: Overlap between Relevance and Discovery Requests

Notice the use of the term "relevant" rather than "responsive," is the latter being a favorite term for counsel and e-discovery practitioners. Figure 8 illustrates why I have done so. This particular objective forms the foundation while its alignment and stability affect the entire process.

The duty to produce documents arises in response to valid requests by opposing counsel.[66] These requests outline the scope of the duty. The term "responsive" is used to denote documents that fall within the bounds of these requests. Thus, the universe of responsive documents is dependent upon the requesting party's view of the matter, and its understanding of the producing party's data enterprise. However, other duties exist that are broader in scope and likely to be more accurate than the duty to produce. The duty to preserve

[66] *See* Infinite Energy, Inc. v. Chang, No. 1:07CV23-SPM/AK, 2008 U.S. Dist. Lexis 88084, at *5 (N.D. Fla. Aug. 29, 2008) (stating that the producing party "must respond to **each** discovery request served in **this** case and identify **each** responsive document").

encompasses that data that is relevant and is not inaccessible.[67] Also, the scope of opposing counsel's requests tends to become more refined, and expands as the knowledge of what exists becomes better known. This means the duty expands as well. Use of such an uncertain, volatile criterion as the foundation for production potentially subjects the entire process to initial error, and to adjustments of greater magnitudes than should be required. While the bounds of the duty to preserve may be subject to modification, they are unlikely to be as malleable as those of the duty based on outside counsel's understanding of the case and the data infrastructure of their opponents. Finally, as the owner of the data enterprise, it is the producing party who has a better, more certain, foothold concerning the content of the enterprise, and it is this knowledge that should be incorporated into the design of the search protocol, rather than basing it on the requesting party's best guesses.

After the relevant and responsive universes are plotted, more layers can be included. Documents that are not "accessible" might be excluded from the production. Documents that are relevant, but are privileged, may be excluded from the production. Duplicate documents can also be removed. There are a number of possible layers that could be inserted, but the idea is that the final production is a set of documents generated by a culling process that creates a "shrinking universe." See Figure 9 below.

---

[67] *See* RMS Servs.-USA v. Houston, No. 06-15585, 2007 U.S. Dist. Lexis 25536, at *4 (E.D. Mich. Apr. 4, 2007) (reminding counsel and their clients that it is their duty to "preserve evidence that may be relevant to this action[,]" and that this duty "extends to documents, data, and tangible things in the possession, custody, and control of the parties . . . [in the] action"); W.E. Aubuchon Co. v. Benefirst, LLC, 245 F.R.D. 38, 42 (D. Mass. 2007) (citing *Zubulake*, 217 F.R.D. at 318 in support of its statement that "the time and expense required to retrieve documents and electronic data depends primarily on whether such information is 'kept in an accessible or inaccessible format'").

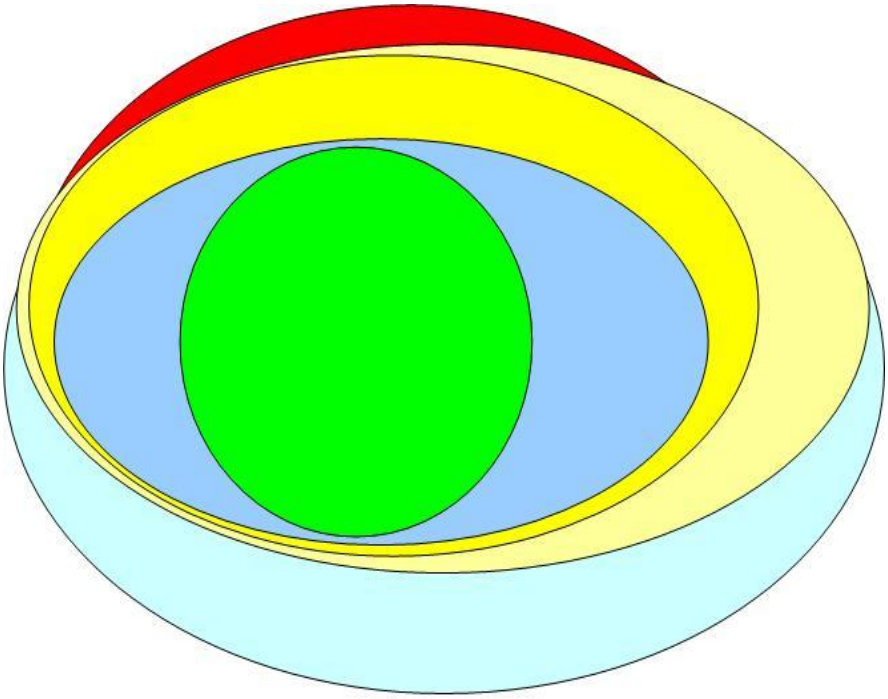


Figure 8: The Shrinking Universe of Production Documents

To return to our question of what it means to engineer and conduct a successful search protocol, we can see that from an operational standpoint, the goal of the search protocol is not to cull as many documents as possible from the review universe, but rather, to cull as many irrelevant documents from the collection while including as many of the relevant documents as possible.

Additional elements that remove duplicate documents (an element included here), or even near-duplicates,[68] are also pieces of the puzzle. Ultimately, the universe of documents shrinks down to something that we can imagine resembles a bull's eye.

CONSIDERATIONS FOR BUILDING THE SEARCH PROTOCOL: PLOTTING THE INITIAL COURSE

The review set[69] must be derived from a data universe by formulating a criterion that is consistent with certain objectives. These objectives, in turn, are constructed from the substantive content of the matter (or matters) at issue in the case. The process looks something like that seen in Figure 10.

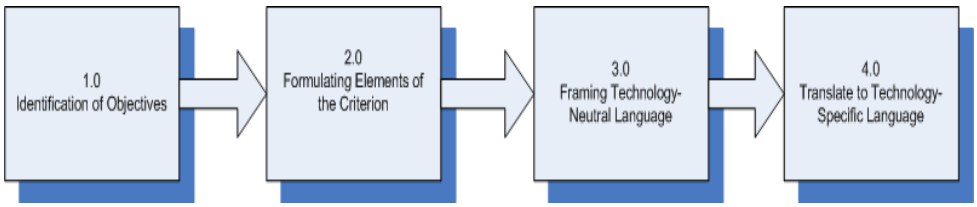


Figure 9: Developing the Search Protocol

Initially, the substantive points of the litigation are analyzed, and objectives are identified with respect to the types of documents that will be relevant that exist within the party's IT enterprise. Those objectives are then transformed into specific elements that will serve as the foundation for the criterion. Next, those elements are used to craft a series of technology neutral questions or queries that, when answered, will return the appropriate documents. Finally, these "statements" are translated into appropriate queries and instructions using the technology at hand. At this point you might be getting concerned with the possibility of scope creep. Where we at one point had a single sentence, we now have found the task expanded to a two part, complex sentence, and finally to a diagram containing four steps.

While the analysis may be more in-depth than some would expect, it should be noted that the effects of more extensive up-front planning often are the reduction of inefficiencies in resulting processes. This leads to cost savings over the whole term of the project. Furthermore, it helps to fortify the process against sanctions by the court. I would suggest that, therefore, it is worth engaging in an appropriate level of planning.

---

[68] Near duplicates are defined differently by different applications. We do not bother to define the term here, because it is tangential to the basic point being made.

[69] Again I clarify my terminology. Here, the article refers to the data set derived from all the steps discussed earlier as the "review set," because the most common use for this data set is to be presented to attorneys for categorization. This is not, as was mentioned earlier, the only scenario.

Thus, the initial stage would involve examination of the issues comprising the lawsuit, and determining what types of documentation would be relevant. So it is at this early stage that relevance must be identified and its effect on the document search defined. Counsel, whether it be in-house attorneys or lawyers from an outside firm, should first extrapolate the issues raised by the complaint. An initial list should be relatively straight-forward, but items may be added as counsel conducts interviews with department heads, section leaders, and others involved with the activities associated with the dispute.

Once a comprehensive list of topics is generated, a collaborative effort should be engaged in an effort to map these topics to:

> 1. A description of what type of personnel would be considered "key players;"[70]
> 2. A list of what projects, or categories, might be associated with relevant documents;[71]
> 3. A list of terms, including those of a general nature, those that are specific to the industry or occupation, and those that are native to the project, department or company.[72]

Once the initial descriptions of sources and document types have been generated in a general format, more precise descriptions should be formulated. These should be neutral to technology involved, although specific custodians and data repositories should be identified. Key word lists should be generated and date ranges identified.

Finally, the formal translation of queries to search syntax, and the generation of formal search protocol documents for each specific data source type should occur. Depending on the workflow used, the generation of multiple queries or protocols isn't necessarily concurrent. One might first generate a plan for a certain type of custodian or data source before addressing the next. For example, a protocol for the collection of data from custodians' PC's might be drafted before looking at plans for collecting data from sources such as network shares or databases.

---

[70] Zubulake v. UBS Warburg, LLC, 220 F.R.D. 212, 218 (S.D.N.Y. 2003) (noting that "Key players" are those [data custodians] likely to have relevant information).

[71] This will depend upon the business structure of the company and the nature of the lawsuit. For instance, an employment discrimination case would look at documents associated with hiring and firing practices typically associated with a company's human resources department, as well as the department in which the complainant worked. In contrast, a patent infringement case may require an examination of documents associated with a certain project that extends across various departments. Some companies have a departmental hierarchy, while others are driven by a project-based model, and knowledge of such structural details will inform the analysis.

[72] These terms will help to form the core of the key words part of the search protocol. The last type of term mentioned are those that the staff itself coined. Individuals often come up with inventive acronyms or names for elements or inventions associated with their projects. These should also be addressed during custodian interviews.

Once the discovery team members have interviewed the appropriate people, and the initial search protocol has been drawn up, the project is ready to go live. The ship is ready to launch. What must be remembered, though, is that no matter how thorough and well-designed the initial protocol, no matter how true the course is at first, the job is not done. You cannot simply point the ship and fall asleep. At least, not if you want to end up safely on the right shore.

CHECK YOUR PROGRESS

It is essential to be able to plot your current position on a map in order to determine whether or not you've wandered, or have been swept, off course. This is not always an easy matter for those managing complex e-discovery projects, for guidance is often lacking. Neither the law, in case or statutory form, nor technical texts, provide easy metrics and techniques to make determinations. In the (sometimes frustrating) absence of bright line rules, we must instead adopt guidelines that are less direct and satisfying, but also afford the flexibility that is often necessary. In Figure 5, a common process is outlined. Although it is one that is frequently used, it is a defective model. This type of process relies on the belief that initial assumptions on key players, dates and search elements are accurate and never in need of modification. We have already seen that courts are not impressed with such workflows. Therefore, a suitably robust process requires the ability to examine the accuracy of the elements of the search protocol and make adjustments. Such a process would look like Figure 11.

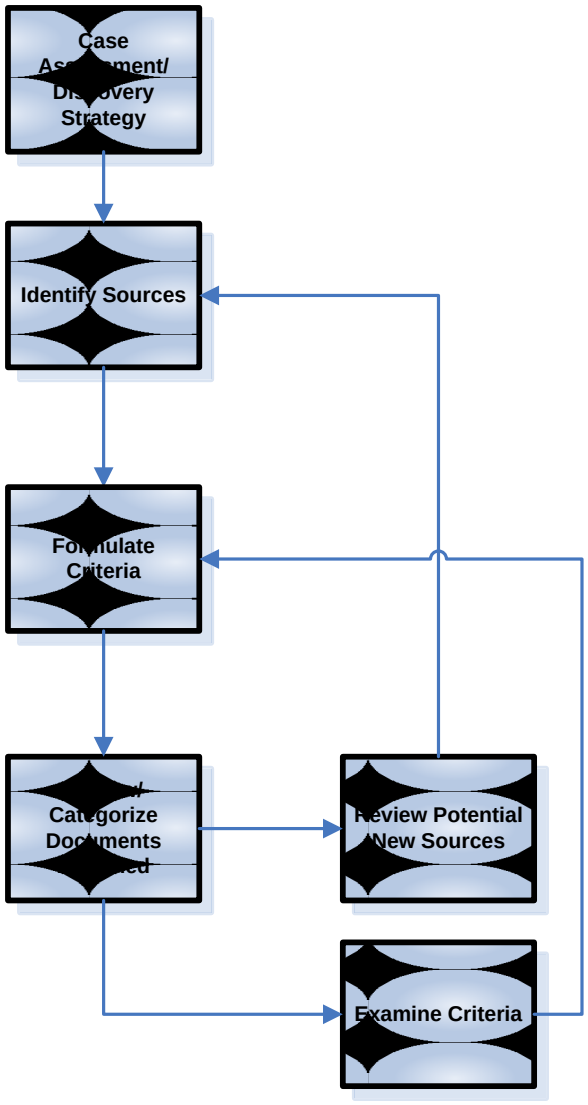


Figure 10: Planning Model Including Feedback Loops

The key idea illustrated above is that the process needs to be cyclical in nature rather than the linear version described earlier. When making decisions on the search protocol, it is recommended that discovery team leaders:

1. Make reasonable initial assumptions;
2. Test those assumptions against results after a sufficiently significant sample of documents have been reviewed and "coded";
3. Make the appropriate modifications called for by the analyses above; and

> 4. Execute a new iteration of the protocol, incorporating the modifications made above.

In this model, attorneys look at the responsive data set for clues for refining their initial assumptions. By generating a list of terms contained by relevant documents, discovery team members may find essential terms that were not initially included or find that some terms have been completely ineffective. An examination of e-mail correspondence that is coded as responsive may reveal previous custodians that should be disclosed to the opposing counsel or investigated for inclusion in the review. Analysis of the storage location of electronic files may be helpful in determining what repositories might need further investigation. Key term analysis might indicate helpful key word combinations that could be used in sampling processes of network repositories or shared drives. Documentation of all steps taken to test initial assumptions of and implement any modifications to the search protocol should be vigorously undertaken and carefully guarded to repel concerns over the process.

Challenges exist in implementing a cyclical workflow. First, the pricing models of many litigation database applications are volume-based. This encourages culling operations prior to loading in order to reduce volume stored and cost. This often creates the situation where data is left residing in a number of different repositories. At best, this means implementing a rather inefficient workflow for modifying the search protocol, and at worst, it may make searching some of the data impossible without further technical manipulation and additional costs.

Effective design of a process needing multiple adjustments requires that those modifications be relatively easy to make. Unfortunately, there are no "Discovery in a box"[73] solutions available, so one must deal with a workflow encompassing several technical components,[74] some of which may not be open to customization. They also may not integrate with each other, forcing the need for additional data manipulation or programming. The cyclical process described here is most efficient, however, if accomplished in a setting with maximum integration. If a search filter is applied at the collection, or worse, at the preservation phase[75] of the project, than modifications emanating from the

---

[73] At least, the author's research has discovered none.

[74] In other words, one application may be used for preservation and collection of the data, another for culling and de-duplication, another for attorney review, and yet another for processing associated with producing the data to the other side. Attorneys might even utilize another application for post-production evidence management to determine which documents are to be used at depositions and trial. Not every one of these applications will charge for volume, but some of them will. Therefore, parties' often emphasize minimizing the volume that is passed through those types of programs. An interactive or cyclical process will often cause an increase in volume through one or more of these programs, leading to higher costs, and discouraging use of these types of workflows.

[75] The author strongly recommends against using a search filter for determining preservation requirements. In a process that assumes modifications will be necessary for initial assumptions,

review portion of the project become more onerous. In theory, if the entire data universe were already loaded into the review database, then the modification to the search protocol would be much easier to make, because the modifications only need to be propagated through a single database, rather than through multiple datasets residing in separate repositories. Of course, this is an improbable scenario, especially for companies with substantial IT enterprises, due to costs and other considerations.

Next, many of the tools currently used in the E-Discovery space do not possess analytical utilities to accomplish those tasks listed above. The good news is that in many cases, ways can be created to work around the limitations of these applications.[76] Once developed, the utilities would be useful for subsequent matters.

Finally, there are no statistical benchmarks in the case law that will help guide attorneys in their decision making. For example, there is nothing that states what statistical thresholds for sampled data will safely allow a litigant to ignore a repository, or what figure mandates collection and review. Like all decisions on process, those made in this area should be reasonable, the rationales listed in documentation, and should not be merely a "default" decision or action of mere reflex.

CONCLUSION

> "Ah how shameless—the way these mortals blame the gods. From us alone, they say, come all their miseries, yes, but they themselves, with their own reckless ways, compound their pains beyond their proper share."[77]

Zeus, King of Olympus, makes the above lament on the lack of accountability. Evidently those pleading for relief were often suffering consequences of their own making. Similar sentiment is echoed in opinions by judges on e-discovery.[78] Courts have made it clear that they will often hold

---

preserving data based on the very criterion that may need adjusting risks losing data. Because it is to be expected that the initial assumptions are not completely correct, the loss will likely be deemed foreseeable, and sanctions may be forthcoming.

[76] This statement comes from the author's experience. On many larger projects, some customization has been required that resulted in the development of either a new utility or process.

[77] THE ODYSSEY, *supra* note 1, at 1.36 – 40.

[78] *See, e.g.*, In re September 11th Liab. Ins. Coverage Cases, 243 F.R.D. 114, 125 (S.D.N.Y. 2007) ("Discovery is run largely by attorneys, and a court and the judicial process depend upon honesty and fair dealing among attorneys."); R & R Sails, Inc. v. Ins. Co. of the State of Pa., 251 F.R.D. 520, 525 (S.D. Cal. 2008) (noting that FED. R. CIV. P. R. 26(g) "requires that every discovery response be signed by an attorney and the signature 'certifies that to the best of the person's knowledge, information, and belief formed after a reasonable inquiry' that the response is complete and correct").

attorneys responsible for the conduct of their clients' document productions.[79] Furthermore, during all phases of litigation attorneys are bound by their state bar's rules for competence and diligence.[80]

Attorneys should act proactively to minimize the risks to their clients. Communicate to the other side, and secure agreements to as many difficult issues as possible. Get the help of one or more experts or attorneys to help navigate the route through the discovery phase. Create a plan that is right for the circumstances. Build a workflow and process based on reasonable assumptions that can be defended. Using a cyclical, rather than linear model, test and, when advisable, modify elements of the search protocol to ensure its integrity.

In his epic poem, Dante encounters Odysseus in the eighth circle of Hell.[81] He is walking wrapped in flame, punished for his deception that helped win the war against Troy. By taking the steps outlined above, attorneys might avoid the judicial and business equivalents to Dante's Inferno: sanctions, loss of one or more clients, and damage to reputations.

---

[79] *See* Qualcomm Inc. v. Broadcom Corp., No. 05cv1958-B(BLM), 2008 U.S. Dist. Lexis 911, at *71-72 (S.D. Cal. Jan. 7, 2008) (fining Qualcomm $8.5 million and referring several attorneys to the State Bar of California for possible disciplinary action). *But see* Qualcomm Inc. v. Broadcom Corp., No. 05CV1958-RMB (BLM), 2008 U.S. Dist. Lexis 16897 (S.D. Cal. Mar. 5, 2008) Although the ruling from January was overturned in order to allow counsel to assert the exception to attorney-client privilege and defend themselves, the opinion stands as a stark warning of the possible consequences when the court suspects malfeasance during discovery. *Id*. *See also* Phoenix Four, Inc. v. Strategic Res. Corp., No. 05 Civ. 4837(HB), 2006 U.S. Dist. Lexis 32211, at *19-20 (S.D.N.Y. May 23, 2006) (fining counsel, who was "grossly negligent," and client, who was at least negligent for failing to timely produce information residing on the client's network, though it was not known to exist. Although the case was ultimately dismissed for lack of subject matter jurisdiction, the fines remained in place, though slightly reduced).

[80] *See* MODEL RULES OF PROF'L CONDUCT R. 1.1 (2009) (requiring "competent representation," which the rule explains means that the attorney possess "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation"); MODEL RULES OF PROF'L CONDUCT R. 1.3 (2009) (mandating that an attorney "act with reasonable diligence and promptness in representing a client"); *see also* MODEL CODE OF PROF'L RESPONSIBILITY DR 6-101 (A)(1) (1980) (requiring an attorney to affiliate himself with a competent lawyer should a matter be beyond his own competence); MODEL CODE OF PROF'L RESPONSIBILITY DR 6-101(A)(2) (1980) (directing that attorneys' preparation be "adequate in the circumstances"); MODEL CODE OF PROF'L RESPONSIBILITY DR 6-101(A)(3)(1980) (forbidding the "[n]eglect of a legal matter" and reiterating the need for diligence by an attorney during activities associated with representation of his client).

[81] DANTE ALIGHIERI, THE DIVINE COMEDY: INFERNO Canto XXVI, lines 52-57 (William Longfellow trans., George Routledge & Sons 1867) (c. 1310-1314)
("Who is within that fire, which comes so cleft
At top, it seems uprising from the pyre
Where was Eteocles with his brother placed.
He answered me: Within there are tormented
Ulysses and Diomed, and thus together
They unto vengeance run as unto wrath.").