

# Understanding ESI Technology and Workflows

Mark Walker

2011



# Understanding ESI Technology and Workflows

By Mark Walker, Director of Consulting  
Forensics Consulting Solutions, LLC

## Introduction

Most, if not all of you have, I assume, heard the hue and cry. I recently “Googled” the term “E-discovery is too expensive” to see what would come up. One article I found claimed that leading attorneys were “stymied” by the “apparently uncontrollable, even irrational costs of E-discovery.”<sup>1</sup>

Recently the question came up on the litsupport listserv, with most of the vendors, of course, questioning the reasoning behind the questions, and others resorting to sarcasm. The American College of Trial Lawyers and the Institute for the Advancement of the American Legal System at the University of Denver recently released a survey which basically stated that “the civil justice system in the United States is so bogged down in a “morass” of E-discovery that it is often too expensive for litigants to take their cases to trial.”<sup>2</sup> One author, a judge, has gone so far to say that “[e]lectronic discovery is killing litigation.”<sup>3</sup>



From the legal side, there are such proclamations as we’ve already cited. From “E-discovery side, we see a different perspective. George Paul assigns responsibility for the situation to the technological revolution.<sup>4</sup> Ralph Losey, an attorney and blogger who specializes in electronic discovery, blames the lawyers.<sup>5</sup> *At least, he blames them a little.*

## A New Set of Principles

From another angle, what we as E-discovery professionals must realize is that we are in the midst of great transition, brought on by the advent of powerful yet inexpensive computers and global connectivity. Data is everywhere. Evidence is data. Combine these facts with our tradition of open discovery, traditionally managed by the parties themselves, and you find yourself in this “morass.” The issue here is that another set of skills is needed to help manage the process of identifying the relevant data, preserving it, and getting it to the other side or into court. The fact is that E-discovery is now a multi-disciplinary endeavor. A successful project needs an attorney who is able to utilize help from others, one or more technical types, and input from other stakeholders, such as employees or people who worked on the project that is now at issue. Mostly what E-discovery requires is a professional approach to managing the project. So the principles we see here are not new to the world at large, they’re just novel to the legal world.

<sup>1</sup> Martin Felsky, *Is E-discovery too Expensive?*, SLAW.CA, Feb.9, 2009 (<http://www.slaw.ca/2009/02/09/is-E-discovery-too-expensive/>).

<sup>2</sup> *Litigation Too Costly, E-Discovery a ‘Morass,’ Trial Lawyers Say*; By Martha Neil; *ABA Journal Law News Now* (September 8, 2008) [http://www.abajournal.com/weekly/litigation\\_too\\_costly\\_e\\_discovery\\_a\\_morass\\_trial\\_lawyers\\_say](http://www.abajournal.com/weekly/litigation_too_costly_e_discovery_a_morass_trial_lawyers_say)

<sup>3</sup> Royal Furgeson, CIVIL JURY TRIALS R.I.P.? CAN IT ACTUALLY HAPPEN IN AMERICA?, *40 St. Mary’s L. J.* 795, 818 (2009).

<sup>4</sup> *The Discovery Revolution – Amendments to The Federal Rules of Civil Procedure*; by George L. Paul and Bruce H. Nearon; (December 2005, American Bar Association)

<sup>5</sup> *The Law Firm Apprenticeship Tradition And Why Most Lawyers Are Still Untrained in e-Discovery*; by Ralph Losey (Blog: eDiscovery Team; January 9, 2011)

## Designing and Managing E-discovery Projects

The primary thing that must change with respect to E-discovery projects is the approach. There is data that is scattered all over Hades, in different forms and formats, and in various types of repositories. Some of it is duplicative; some of it is not reasonably accessible. The entire need is driven by the legal context and the legal rules. These change from project to project. The first change is to embrace the philosophy of taking control by proactively managing the project, rather than simply being reactive. By applying proper principles of design and management, you should be able to:

- Add value to the project
- Increase efficiencies
- Select appropriate technologies
- Select appropriate methodologies
- Get the best operational result for the money

## Evolving thoughts about E-discovery Project Management

The electronic discovery industry, and much of the legal world, have now come to the conclusion that proper project management principles must be applied to, at the very least, the more complex projects. Steven Bennett of Jones Day recently penned an article that compared EDD Projects with large, complex construction projects. He did this because as an attorney he has worked a lot on construction-related cases. Because construction requires people from many different professions, and because without strict management a construction project can run out of control, he has come to realize that in many ways, the need to manage EDD projects similarly exists. Bret Burney, a technology consultant, has discussed “the emerging field” of EDD Project Management.<sup>6</sup>

Ultimately, one part of the equation in “fixing” E-discovery is to engage in better, more professional management of electronic discovery projects.

## Variations in Project Lifecycles

Projects are not all the same. A particular project, depending on the legal context, may require different tasks than prior projects. There is, however, a set of collective tasks, illustrated by the EDRM<sup>7</sup> that might be considered standard. Even using this as a basis however, the project looks different depending on one’s point of view. The company view of a successful E-discovery project might be the lowest possible cost, where a technical professional would view operational success as finding all relevant documents and eliminating all non-relevant documents. An attorney’s view might depend upon the content discovered.

Ultimately, the client’s best interest is the appropriate point of view. The client is served best by using defensible protocols in response to the court’s expectations.

Project lifecycle is basically the same regardless of the actual scope of the project, and it will contain these elements:

- Pre-project planning phase
- Design phase
- Execution phase
- Review phase

---

<sup>6</sup> <http://www.criminallawlibraryblog.com/Burney-EDManagement-TF%2009-01-09.pdf>

<sup>7</sup> <http://edrm.net>

## Pre-Project Planning & Infrastructure

One big factor in initial phases of an electronic discovery project is the infrastructure and early litigation methodologies already adopted by the client company. If the company uses an email “vault” system which simplifies and de-dupes e-mail messages, or a document management system that provides de-duplication of files, then post-collection processes will change. Sophisticated technology solutions within an organization can be costly, however. In an economy where corporations are hard pressed to make and sell widgets, much less fund a fleet of consultants and a massive IT undertaking, this is a big pill to swallow for most in the business world.

The vast majority of corporate America views records management and associated technology as more of a compliance issue because litigation just does not consume enough of the corporate budget to get finance attention. When it does, especially for those big guys that get sued a lot, there is ample technology and consultants to implement those solutions. While the legal departments and IT professionals would really prefer to get rid of all that unnecessary information, the business people who make and sell the widgets think their RIM programs work just fine. They can generally find what they need, when they need it. That is not to say that every company – large or small, litigious or not – couldn’t benefit from some proactive data organization, the reality is that in most cases those programs just don’t get properly funded. So, we are left with solving the real problem – the cost of sending documents out to the lawyers and investigators.

“People have an average of 30,000 e-mails per year per person,” says Atlanta-based SunTrust Banks deputy general counsel Brian Edwards. “Over the last five years, that has meant as many as 1.5 million documents for a single matter. Throwing \$150-\$300-per-hour law firm associates at the mess, for privilege and responsiveness review, is too expensive. “Without a tool that would let you do it faster ... you could get 50-100 document decisions per hour per person,” Edwards says. That’s 3,000 documents for one person’s 40-hour week: “Then do the math for 900,000.”<sup>8</sup>

If I’m in charge of the client’s role, and the client has sufficient resources (i.e., in-house counsel, an in-house E-discovery team, etc...), my preference would be for the client to be able to handle it’ own data to such a degree that it can ensure a “smooth hand-off” to outside counsel by providing:



- Documentation on the data sources;
- Data maps;
- Documentation on user’s personal systems;
- Assistance from in-house resources.
- Continued Maintenance and Monitoring

However, as previously noted most corporations don’t have the budget, or the technology for that smooth handoff. As a result, careful planning and implementation of a process that leverages internal corporate technology and human resources coupled with external resources as needed is critical. What external resources are needed, if any, is going to depend upon the organization. What is clear as noted by SunTrust’s Brian Edwards is that the present course by most in the business is not dealing with the out of control data sizes and the resulting cost.

---

<sup>8</sup> See *How to Keep ESI at Bay in E-Discovery* by Erik Sherman, Corporate Counsel, November 29, 2010 (LTN Technology News at Law.com)

## Project Design

Designing a project is basically the act of building a project plan. It's similar to crafting a blueprint or creating a plan for software development. It works like managing any project with various, complex parts. The steps for creating a project plan are as follows:

- Review current systems already in place
- Gather requirements
- Design a solution
- Test the solution
- Make Adjustments
- Definition of the Goal or Objective of the Project

You begin the project design by doing your best to set limitations. This may be done using a scoping document. A scoping document looks at the matter involved, the number of custodians and other data sources, the variety of data sources. The calculation of expected data volume and examination of data file types involved are also elements. You can take the estimated volume of data through an entire workflow to calculate costs. At this time, discussion of the form of production should also begin, if it hasn't already. Again, because the scope of a matter is a variable, it may involve all or merely a portion of the EDRM.

Again, effective design requires collaboration from several actors. These actors will have input which will help determine some of the aspects that will be incorporated in the final project plan. Some of these aspects include:

- Linear review vs. Non-linear technologies (clustering)
- Tagging scheme development (taxonomy)
- "Searching/Culling Protocols"

## Basic Precepts

Some basic precepts should exist with respect to designing a project. If collection is included in the project scope, don't "cull" prior to preserving. This is a little counter-intuitive, because the very notion of identification of data means eliminating some data as a potential source. The use of search filters on data that is in daily use, or is accessible to users (e.g., a user's laptop) who might delete some of that data, is not recommended. Eliminating data at the source through whatever means carries considerable risk. Doing so provides, essentially, one bite at the apple. Collect broadly and process, review and produce narrowly plan is the better approach. The greatest portion of the cost of data is not preservation, but in the processing, review and production of that information. The reasons underlying each major decision, such as who is a custodian, what Boolean terms are used to filter, what method is used for collection, should be well documented. This provides the client and the team with a "foundation" upon which rests the defensibility of their decisions. This documentation should list the initial assumptions made (such as search terms, custodian identities, etc...), and the reasons why those assumptions were taken. The document should further discuss testing of the initial assumptions, and adjustments taken after reviewing testing results.

## Design Deliverables

The project plan should encourage active management of the project. Design deliverables can include the following elements:

- A Project Mission Statement
- A Statement of Work (might be combined with above)
- Workflows and their narratives
- Risk and Pricing Point Assessments
- What measurements are going to be used to monitor performance and determine success?
  - Cost Estimates
  - Benchmarks
  - Timelines



These elements should be carefully considered and completed with an eye toward measuring a project's progress against expectations. If the project expands or events require a change of scope, then the documentation should be amended appropriately. Additional documents describing the change of scope should also be generated.

## The Importance of ESI Technologies

Decisions determining the technologies that will be employed to collect, process and review ESI are critical, as they influence cost and risk. The decisions to use certain technologies should be justified. This justification can be laid out in the Mission Statement or Statement of Work. When putting together a project that will consist of more than one platform, it is important to understand how one application affects the data, and how this affects applications residing downstream in the workflow. Another factor to consider is how much integration or manual data manipulation will be executed outside the applications themselves.

## Considerations for Choosing Technology

Consider the whole picture (if possible, given the scope of the matter) with respect to costs. Higher technology costs might be offset by lower costs elsewhere. This is readily apparent when you look at the total costs (including attorney review costs) of the project. You also want to have good empirical data from past projects so you don't get caught up in the Hypothesis-Speculation Game of "this is how we thought the costs would line up". Map requirements to the technology platform to make sure those requirements are properly addressed. Fill in the "gaps," by explaining how requirements not addressed by the applications in the design will be resolved.

## Project Execution

The Project Execution phase consists of active management of the project. Resources are allocated towards execution, communication and reporting protocols are established, and the project progress is managed in accordance with expectations contained within the Project Design documents. When expectations prove unrealistic, or scope changes, the Project Manager alerts the appropriate stakeholders and provides input on scope change management or re-adjustments of milestones.

Project execution is when the team takes the ball and runs with the game plan designed earlier. As the Execution phase begins, the design is tested monitored to see if the expectations are met. Should initial assumptions prove faulty, adjustments are made to the project plan. Should the scope of the project change, scope change management protocols are initiated and followed.

## Pitfalls

We have seen, time and time again, in various opinions, what happens when electronic discovery projects are not managed carefully. One cautionary tale comes from the case of *Covad Communs. Co. v. Revonet, Inc.*, 2009 U.S. Dist. LEXIS 75325 (D.D.C. Aug. 25, 2009).



Previous decisions in this contentious case required, among other things, that the producing party submit itself to a third party forensics expert for an examination of its search protocol. In the instant decision, producing party was defending the fact that, after having produced e-mail in hard copy format, and then being ordered to re-produce it in native format, it only produced a native sub-set of the original imaged production. The producing party, in effect, blamed the processing platform for its difficulties in reconciling the two production sets. The court responded:

*While the nature of the discrepancy is not explained, I have to assume that fewer e-mails have been produced in native format than were produced on paper. Revonet explains that the platform that Revonet originally used to search for documents was only capable of exporting documents to an HTML format. Thus, Revonet had to use a different platform to obtain .pst files and therefore could not re-run the original search exactly. This may explain the discrepancy. Neither party provides information about the magnitude of the discrepancy, however, largely because Revonet claims that it would be too burdensome for it to cross-reference the electronic documents against the hard copies to determine how many and which e-mails are missing.*

*While I appreciate that it would be difficult for Revonet to go back through its papers to determine whether all of the documents contained therein have since been produced and that Revonet's present counsel did not supervise or conduct the August, 2008 search for e-mails, I also appreciate that it is a burden of Revonet's own making. Covad should not be penalized by Revonet's failure to maintain its discovery materials in some sort of organized fashion or keep some record of its own actions in this lawsuit. Wyeth v. Impax Labs., Inc., 248 F.R.D. 169, 171 (D. Del. 2006) ("[P]roducing party must preserve the integrity of the electronic documents it produces. Failure to do so will not support a contention that production of documents in native format is overly burdensome.").*

In other words, a burden of one's own device is not a basis for asserting undue burden. Another maxim: choose your platform wisely.

A second fable can be found in the case of the missing attachments. In *PSEG Power N.Y., Inc. v. Alberici Constructors, Inc.*, 2007 U.S. Dist. LEXIS 66767 (N.D.N.Y Sept. 7, 2007), the court took one party to task for deficiencies in its management of the electronic discovery aspect of the case.

The opinion in opens with the magistrate expressing his understanding of the difficulties posed in this case by electronic discovery:

*For nearly six months, the parties and the Court have been grappling with an electronic discovery monstrosity with the hope that it could be corralled and definitively resolved, thereby obviating the need for motion practice. Alas, attempts to resolve the issue in lieu of briefs fell woefully beyond the parties' grasp and, as the last straw, they have set the matter at our feet for appropriate resolution.*

PSEG responded to Alberici's first document request by producing, over a period of six months, 211,000 pages of hard copy, which Alberici then converted to electronic format (TIF images), and loaded into a litigation database at its own expense. PSEG also produced a disc of e-mails. Alberici had also produced a large amount of document to PSEG. At this point, all is straightforward. PSEG and Alberici each produced their own documents, at their own expense. Each paid for post-production conversion of the other party's documents as well. Here's where the tale goes astray. Judge Treece describes the problem:

*In January 2007, it became evident to Alberici that PSEG had produced emails without the attachments which were referenced as being a part of the emails. Apparently a technical glitch occurred whereby numerous emails were "divorced" from their attachments caused by limitations in the downloading software. The separation of the emails from the attachments happened at the interface between the different software used by PSEG and the vendor when reducing the documents in a form that could be reviewed by counsel. It appears that the "vendor's software was not compatible with the HTML format in which PSEG had provided its documents and that this incompatibility had resulted in the parent child link between the emails and attachments being broken."*

We've lost the connection between our e-mail messages and their attachments. This means that in responding to discovery requests by Alberici, PSEG failed to produce relevant material. The good news is that the required data exists, and this isn't a case of evidence destruction. The bad news is that quite a bit of technical expertise will now have to be put to use in order to either 1) re-process the e-mails in a manner that preserves the connection to attachments; or 2) use the existing evidence to (in the court's words) "re-marry" the two.

Judge Treece tells us what the parties did upon discovering the problem:

*Throughout this ordeal, the raw data was not lost. All 750 gigabytes of unfiltered data remained intact in its original format. Realizing that the underlying data still existed, the next proposal included PSEG sharing with Alberici's vendor a sample of the metadata for analysis. However, the dearth of metadata related to the emails and attachments rendered this proposal fruitless. In the interim, the parties' vendors explored other ways to reverse engineer the available data and "re-marry" the attachments to their emails. This exploration was for naught inasmuch as the data necessary to complete this task was destroyed during PSEG's collection and formatting of the emails.*

So, although the "raw data" is still intact, the data resulting from downstream processing is in such a state so as to render the re-connecting of parent email to child attachment impossible. At this stage, I'd like to point out something very important. Changes to data have occurred in the EDD processing (or loading, or ingestion, or whatever you want to call it) phase, rendering that data unuseable for the task at hand. However, there is a reservoir of "raw data" available because of (what one supposes is) correctly collected data. This illustrates the vital importance of a defensible collection process. The opinion here gives no description of how data was collected, because the collection isn't being challenged. Nevertheless, without that defensible process in place, there would be no acceptable "raw data" reservoir as backup, and the issue might not be late production of attachments in an appropriate form, but destruction of data, a much more serious affair.

After the parties' attempts to collaborate on solving the problem of the missing attachments failed to produce an accord, the magistrate got involved:

*[T]he issues are several-fold: (1) is Alberici entitled to receive the emails with the related attachments together as opposed to their current state of separation, lacking coordinated identification with each other; (2) although PSEG has provided these emails and attachments in hard copy albeit not "married," is PSEG obligated to provide these documents in their original format; and (3) if re-production is required, which party bears the cost of this production?*

Alberici felt as if the e-mails and attachments should be connected in some easily identifiable way. PSEG argued that Alberici “is impermissibly seeking a “perfect” or “ideal” production, regardless of expense or benefit.” Furthermore, “[s]uch a re-do effort would be duplicative and entirely unnecessary in its view.”

PSEG offered its own alternative to solving the problem:

*PSEG wants Alberici to identify a concise group of attachments that are important and necessary to Alberici and then it would consider producing said attachments, however, reserving its right to assert that it may be irrelevant or non-responsive or privileged. Or, if Alberici insists on a re-production, PSEG is willing to provide them but at Alberici's expense.*

So...we've already produced the stuff, although not in an optimum state. If you want it produced yet again, you pay for it. How did that go over with the court? The opinion responded by mentioning the changes to FRCP 34(b):

(ii) if a request does not specify the form or forms for producing electronically stored information, a responding party must produce the information in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable; and (iii) a party need not produce the same electronically stored information in more than one form.

*Obviously, the commonsensical purpose of this mandate has always been to prevent massive dumping of documents, without form or direction, thereby **alleviating an incalculable burden upon the requesting party of searching for the proverbial needle in a haystack**. In this respect, notwithstanding Rule 34(b)'s amendments, (ii) and (iii), **PSEG would still have to produce business records as kept in the regular course of business or in such a manner that Alberici could readily find a necessary document or two**. It has also been a seminal rule that the responding party would not have to satisfy the requesting party's whim to have the documents produced in various forms.*

The Judge then cut to the chase:

*Clearly these 3000 emails and related attachments **were not produced in accordance with this mandate**, and, as we now know have caused considerable consternation and agony for both parties, which the revised statute was attempting to avoid. Normally, one would expect that an email and its attachment would have been kept together in the regular course of business, and the production of said documents would have followed suit. Here, the difficulty has been that there was not sufficient identifying information to match attachments with their respective email. We accept Alberici's proffer that it has spent considerable time employing different methodologies to unearth attachments to correspond with the emails it has found to be pertinent. Attempting to reunite these documents has been nothing short of a donnybrook for Alberici. It has been frustrated if not completely hamstrung in locating these documents. Compounding Alberici's angst is the disadvantage it has been placed in preparing for depositions. **In essence, the first production of emails and attachments has been ineffectual.***

The judge then takes a shot at PSEG's vendor:

*We acknowledge that discovery production is rarely perfect or ideal, yet this discovery quagmire created by PSEG's vendor falls woefully short of comporting with the spirit of Rule 34.*

Ultimately the Judge decided that PSEG must take responsibility for the consequences of its own vendor's misadventures, by denying the request for cost-shifting, and granting Alberici's motion to compel.

In one project in which I worked, we noticed missing attachments for some e-mails, but not for others. This is different than what happened with PSEG, in that the attachments that the e-mail message indicating should be present didn't exist. After quite a bit of investigation, we discovered an error in collection, which had been executed by the client's own IT staff. Remote users (with laptops) would receive e-mails logged in over the web, but wouldn't actually download the attachments unless they requested the document (by clicking on it). So if they never viewed the attachment, it didn't exist on their system, despite the e-mail's indication that the attachment should be present. As a result, the data had to be collected yet again. Additionally, some rather extensive data manipulation had to be done to get the originally processed data and the re-collected data to match. Never was very pretty, but at least the issue was discovered before production.

## **Conclusion**

By applying professional management principles and methodologies to electronic discovery, we can avoid the pitfalls that show up in the opinion...at least, we can avoid the common ones that plague those who fail to manage their projects correctly. In carefully managing our projects, we should put great emphasis on process and systems, and ensure that fewer decisions are made arbitrarily or in a discretionary manner. By attempting to map out the entire project as early as possible, we encourage our discovery team to address essential aspects of the project early, leaving less room for surprises at the end. Ultimately, we want to engage in proactive management of our projects, so that we are less reactive to events, and attempt to prevent pitfalls from ever occurring.

## **About the Author**

**Mark Walker**  
**Director of Consulting Services**

Mark Walker came to Forensics Consulting Solutions as a Senior Consultant with more than twenty years in litigation support services, including time he spent as a trial paralegal, logging more than thirty commercial trials, the corporate law department, and Director of Litigation Technology and Support in the law firm environment. Mark's responsibilities included handling and coordinating litigation support services. After joining FCS in early 2007, Mark's extensive knowledge of legal technology soon became apparent and he was quickly promoted to Director of our Consulting Services organization. In his current position, Mark interfaces with law firms and corporations to develop and manage technology solutions to meet our clients' needs.