Chapter 21

AN EXTENDED MODEL FOR E-DISCOVERY OPERATIONS

David Billard

Abstract Most models created for electronic discovery (e-discovery) in legal proceedings tend to ignore the technical aspects mainly because they assume that only traditional digital forensic tasks are involved. However, this assumption is incorrect. The time frames for conducting e-discovery procedures are very restricted, and investigations are carried out in real time with strict non-disclosure dispositions and changing demands as the cases unfold. This paper presents an augmented model and architecture for e-discovery designed to cope with the technological complexities in real-world scenarios. It also discusses how e-discovery operations should be handled to ensure cooperation between digital forensic professionals and legal teams while guaranteeing that non-disclosure agreements and information confidentiality are preserved.

Keywords: Electronic discovery, technical aspects, non-disclosure

1. Introduction

Electronic discovery (e-discovery) refers to any process in which electronic data is sought, located, secured and searched with the intent of using it as evidence in civil or criminal legal proceedings [3]. The most popular e-discovery model is the Electronic Discovery Reference Model (EDRM) [2], which is presented in Figure 1.

EDRM expresses the phases of e-discovery from the point of view of an attorney. The six e-discovery phases, which are very similar to those proposed by McKemmish [4] for digital forensic investigations, are summarized as:

- **Information Management:** This phase is not necessarily part of e-discovery. Rather, it is a pre-processing step that should be performed by an entity in case litigation should occur.

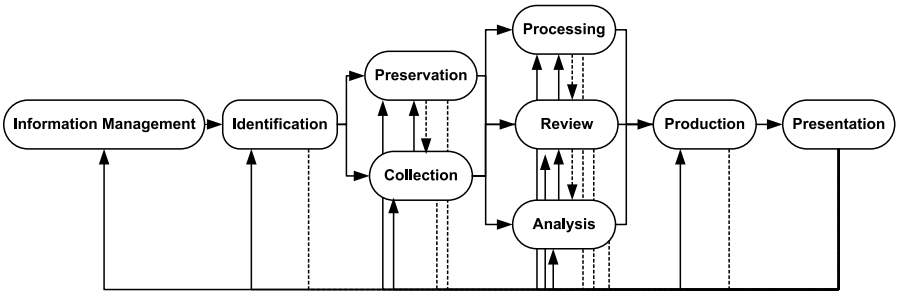

Figure 1. Electronic Discovery Reference Model.

- **Identification:** This phase involves the location of the potential evidence containers.
- **Preservation and Collection:** This phase deals with the preservation and collection of the potential evidence containers, during which exact copies of the evidence are made in a forensically-sound manner. Note that only the evidence deemed to be relevant to the case is considered.
- **Processing, Review and Analysis:** This phase involves the processing of evidence by digital forensic experts. The results of the processing are reviewed by legal teams and analyzed by the concerned parties.
- **Production:** This phase involves the production of privileged material in a human-readable format.
- **Presentation:** This phase involves the presentation of the evidence in the legal framework of the case.

Unfortunately, EDRM is too simplistic to capture the complexities of e-discovery in the real world. For example, the Processing, Review and Analysis Phase often involves several (possibly hundreds of) individuals with different qualifications and roles operating under different legal and contractual frameworks. Moreover, the information that is ultimately produced (i.e., discovered) along with its byproducts are often designated to be released to certain individuals and not others.

EDRM is also unnecessarily complicated by the presence of many return paths (arrows in Figure 1). Some of the return paths do not need to exist and may even introduce flaws in e-discovery procedures. Consequently, we propose to augment EDRM and, at the same time, eliminate certain return paths.

E-discovery processes are tightly controlled by procedures and court orders and usually operate in restricted and regulated time frames. For example, Rule 26(a) of the U.S. Federal Rules of Civil Procedure allows for an initial disclosure before an actual discovery request is made; Rule 16(b) imposes a scheduling order; and Rule 26(f) requires the parties to confer at least 21 days before a scheduling order is due. Fortunately, even if each e-discovery case is unique, it is possible to capitalize on certain invariants.

The paper presents an augmented model for e-discovery. It identifies the various actors involved in e-discovery and their roles, and proposes an augmentation to EDRM designed to cope with the technological complexities in real-world scenarios. Finally, it discusses how e-discovery processes should be handled to ensure cooperation between digital forensic professionals and legal teams.

2. E-Discovery Actors

Several individuals and teams of individuals are involved in e-discovery operations. These actors do not have the same levels of knowledge about the case and are bound by various contracts and non-disclosure agreements. We distinguish four actors that operate with respect to these non-disclosure agreements.

- **Digital Forensic Team:** This team is responsible for extracting and collecting potential evidence from all types of devices: hard drives, cell phones, backup tapes, GPS devices, etc. The potential evidence containers are carved, decrypted, de-duplicated, indexed, searched and made user-readable using advanced forensic tools and dedicated software.

- **Research Teams:** These teams usually comprise attorneys and legal assistants who have access to the potential evidence. The role of the research teams is to pre-sort information pertaining to the case as "privileged" (i.e., information pertinent to the resolution of the case), "confidential" (i.e., private information about the individuals whose devices were searched), and "irrelevant." In general, there are at least two research teams, one for each party involved in an e-discovery case. The research teams focus on the meaning of the documents and, therefore, may not incorporate technical personnel.

- **Party Counsels:** These actors are the attorneys who are in charge of procedures on behalf of their respective clients (parties). They have the ultimate say about relevant information pertaining

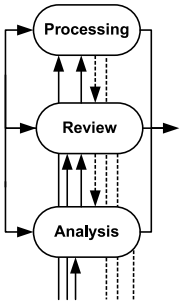

Figure 2. Processing, Review and Analysis Phase.

to the case. They guide the research teams' activities and may ask for additional investigations to be conducted by the digital forensic team.

- **Chief of Forensic Operations:** This individual is in charge of dispatching evidence to the research teams; maintaining the technological means to ensure non-disclosure; securing privileged information along with the scientific processes that support its findings; and liaising with the digital forensic team, research teams and party counsels to ensure that the processes are carried out correctly. Needless to say, serving as a chief of forensic operations is a most demanding task, with intense pressure and a close relationship to the case core.

The interactions between these four actors are complex and vary considerably throughout the e-discovery process. This complexity is captured using the e-discovery framework described in the next section.

3. E-Discovery Framework

The framework described in this section is intended to fully support e-discovery processes in the real world. The first three EDRM phases are relatively traditional. They do not involve research teams and, from the technical point of view, can be handled adequately by trained digital forensic professionals using state-of-the-art tools. However, the fourth EDRM phase, Processing, Review and Analysis, is complex and requires special consideration.

3.1 Processing, Review and Analysis Phase

The Processing, Review and Analysis Phase is illustrated in Figure 2. This phase of e-discovery is the most complex and costly. In the

following sections, we discuss the processing, review and analysis steps in detail and associate them with the various actors involved in e-discovery operations.

Processing: The processing step involves several tasks.

- **Document Carving:** Carving is used to retrieve documents, images, audio, video and, above all, email. Several forensic tools are available to accomplish this step; they can retrieve deleted documents as well as documents embedded in emails, compressed files and archives.

- **Decryption:** Some organizations use cryptography to secure their data. This is usually a good policy and is recommended in the normal course of business. Unfortunately, it complicates the work of the digital forensic team because decryption keys and passphrases may be missing.

- **De-Duplication:** When dealing with a corporate email system or document repository, it is often the case that the same document is found multiple times. By "same document" we mean a document with the same contents and metadata. For example, a document emailed by an executive to company employees may be present in the sender's mailbox and in the mailboxes of the other recipients, and even more mailboxes if the original mail was forwarded to others. While the fact that the document was sent might be important to the case, it is unnecessary (and a waste of resources) to preserve every copy of the email during e-discovery. Consequently, the digital forensic team would use de-duplication tools to identify duplicate files and retain only one copy of the file.

- **Search Indexing:** In general, e-discovery operations rely heavily on keyword searches. Consequently, it is important to index data and to use powerful search engines.

- **Presentation:** The purpose of the processing step is to create content for the research teams. The content must be delivered in a format that enables the research teams to sort and label the documents quickly and efficiently. Documents should be pre-categorized with respect to their potential value to the case. For instance, documents written in French might be relevant and those in German less relevant, or emails with attached spreadsheets should be examined first.

- **Comparison Chart/Timeline Preparation:** File content is not the only information of interest in e-discovery operations. For example, it might be important to know if a person engaged in certain stock market transactions before or after receiving an email. In such an instance, the digital forensic team has to create charts and timelines from the available files and their metadata, and from information pertaining to the files (e.g., dates). Investigative tools are available to facilitate the production of charts and timelines.

- **Repository Creation:** Increasing amounts of information are maintained in databases as part of a software suite or in databases built specifically for organizations. It is of extreme importance to be able to connect to and access information from these databases without disrupting normal business operations. The collected information could be stored in a specially-designed repository and searched using business intelligence tools. These tools sort through the data, present information in a condensed form and offer analytic services.

The seven tasks described above are performed by the digital forensic team with input from a party counsel. The party counsel would provide the categories of documents and the keywords to be used in searches. Typically, there is close cooperation between the forensic team and the party counsel in developing the keyword list. For example, the party counsel might provide a list of nicknames or a list of phone numbers to be used in searches.

Review: The review step involves the examination of the documents produced during the processing step. The documents to be examined are given to the research teams by the party counsels (via the digital forensic team) as the case goes along. The party counsels orchestrate the information flow during the review step.

It is important to prevent information leaks during the review step. The individuals participating in the review step can be in the presence of very sensitive documents, including documents that are not pertinent to the case. Therefore, the research teams must work in strictly-controlled physical locations with no telephone service or Internet connectivity. Computers used by the research teams should have their USB, firewire, wireless and CD/DVD functionality disabled. Also, all internal network communications should be encrypted.

The research teams sort the documents based on the filters provided by the party counsels. As a result, the documents are categorized as

"privileged" (relevant to the case), "confidential" (private nature), or "irrelevant" (not processed any further).

Analysis: The party counsels analyze the privileged documents with respect to the case objectives and applicable laws. The analysis could lead to additional processing as well as the inclusion of new evidence containers. For example, the analysis of the documents might shed light on the behavior of a new person in the case and his cell phone becomes a new potential evidence container. The party counsels may also request the chief of forensic operations to conduct new searches. The chief of forensic operations quantifies the duration of the searches, oversees their progress and ensures that the schedule is maintained.

3.2 Additional Steps

The additional steps include the documentation of forensic processes and the cleaning of digital media.

Forensic Process Documentation: Every forensic task performed during the Processing, Review and Analysis Phase should be documented in detail. Documentation may not be considered as a step as such, but it should be done continuously.

Digital Media Cleaning: All digital media should be cleaned at the end of the Processing, Review and Analysis Phase. This includes every computer used by the digital forensic team, research teams and chief of forensic operations that could have any data relating to the case. Two common cleaning techniques are to erase all data using a DoD-certified method or to physically destroy all the data containers. The first method is very time consuming – several hours may be required to wipe a single hard drive. The second method, which we believe is more appropriate, is to destroy the data containers using hammers, drilling tools and possibly fire and acid.

3.3 Modified E-Discovery Reference Model

We propose a modified EDRM incorporating both simplification and augmentation. The EDRM workflow is simplified by incorporating fewer feedback loops:

- The party counsels might identify a new potential evidence container. In this case, the potential evidence container must be treated as new evidence and should be preserved.

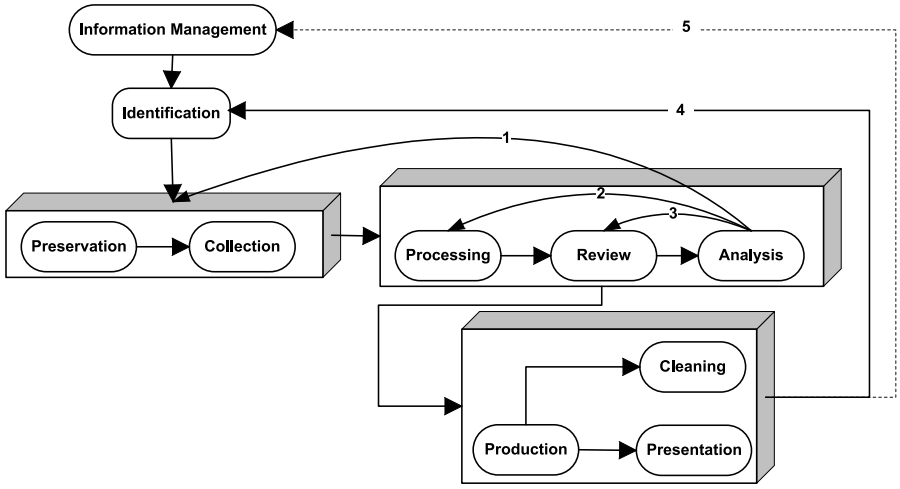

Figure 3. Modified EDRM.

- The party counsels may alter the keyword lists or search filters. In this case, the new information is sent to the digital forensic team for processing.

- The party counsels might modify their reviewing criteria. In this case, the research teams are informed about the change.

- Additional data may be sought by the judge or by the parties after the privileged information is produced and presented. In this case, a new identification phase is initiated.

- The lessons learned during the entire process are integrated in corporate information management systems in the event of additional e-discovery demands.

We also introduce a cleaning step to the model in order to erase all the data on the devices used in e-discovery.

Figure 3 presents our modified EDRM schema. The schema is simpler, but more accurately reflects the complexity of the e-discovery process. The boxes represent atomic, closed steps corresponding to distinct units of work. (Note that "production," "cleaning" and "presentation" fall in the same unit of work.) Some additional return paths can be drawn; however, we believe they constitute exceptions and are, therefore, not included.

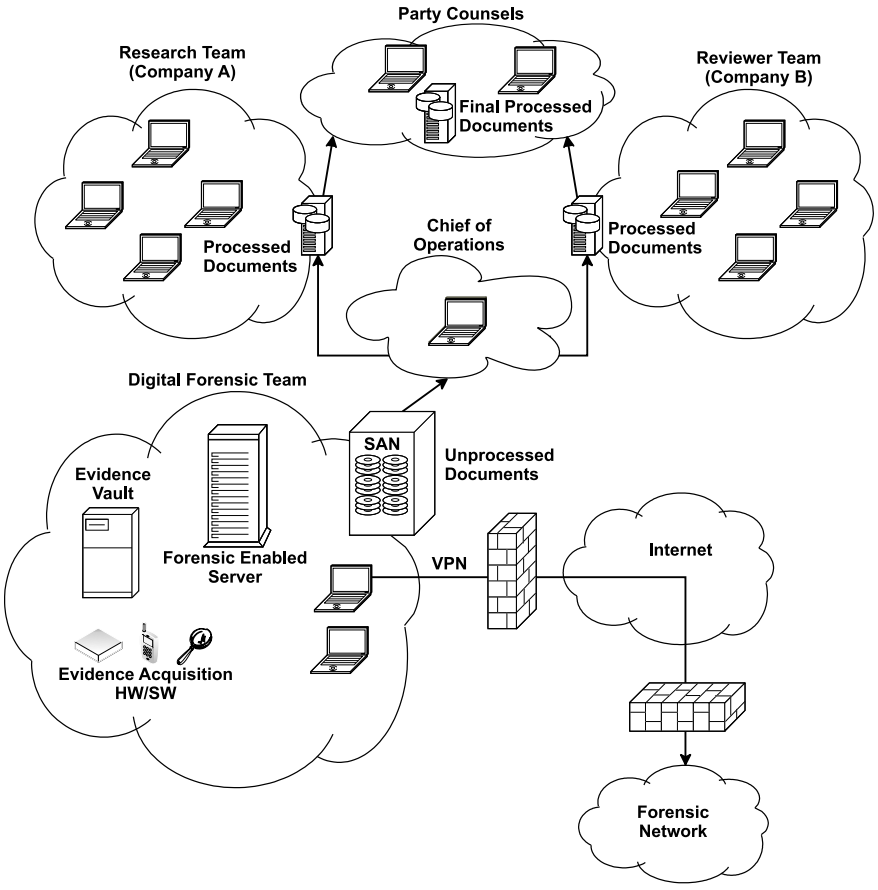

Figure 4. Technical architecture for e-discovery.

3.4 Technical Architecture

We have designed an architecture that supports our modified EDRM (Figure 4). This architecture has been used in a real case involving companies in the United States and Europe.

Note, however, that every case is unique and the specific e-discovery setup may have to be altered to match the objectives and local support. Sometimes, a setup has to be reproduced. For example, in a multinational case, collected data may not be transferred legally from one country to another because of different laws [1]. Therefore, the setup has to be reproduced in each country and the chief of forensic operations at one of the sites serves as the "chief of global operations."

Our example in Figure 4 has two opposing companies that are processing potential evidence (emails, spreadsheets, mobile phone calendars,

etc.). The actors include two research teams (one for each company), a digital forensic team, two party counsels and a chief of forensic operations. The infrastructure involves five switched networks interconnected through gateways with very limited (and tightly controlled) connections. All communications within and between networks are encrypted. Computers in the research teams' networks have their USB, firewire, wireless and CD/DVD functionality disabled. Also, research team members do not have access to telephone service and the Internet.

The chief of forensic operations orchestrates all activities. He takes orders from the party counsels, organizes the documents to be sent to and received from the digital forensic team and the research teams, and operates the server that hosts the final documents. The chief of forensic operations also interacts with the digital forensic team on new requests received from the party counsels.

The non-disclosure property, which is paramount in e-discovery cases, is achieved at the network boundaries. All reasonable hardware, software and policy measures must be implemented to ensure that no data can leave the secured networks.

The overall e-discovery process can be summarized as follows:

- The potential evidence is extracted and collected by the digital forensic team. The potential evidence containers are carved, decrypted, de-duplicated and made user-readable using state-of-the-art forensic tools. The resulting data sets are stored in a storage area network (SAN) or using network attached storage (NAS) (note that this can impact data transfer rates). All the servers are managed by the digital forensic team and the original potential evidence containers are secured in a vault.

- The chief of forensic operations accesses the SAN and dispatches data sets according to the case.

- The research team members blind-filter the data and transfer the relevant filtered data to the party counsels.

- The party counsels make the final decisions pertaining to the data (e.g., evidence to be retained, personal data to be discarded and irrelevant data) and store the evidence on a distinct server. The party counsels may also ask the chief of forensic operations to perform additional searches and the research teams to analyze documents using new criteria.

- A virtual private network provides access to the digital forensic team's computer center for situations where several setups are

needed for the same e-discovery operation. However, only technical information – not case data – is transmitted via this link.

- At the end of the e-discovery operation, all the hard drives (on laptops, computers, servers) are wiped clean based on DoD standards or are physically destroyed to prevent any data from being recovered.

The e-discovery infrastructure described above is not as "bulky" as it might appear. A lightweight version can be implemented with the servers running locally at the digital forensic team's computer center. Indeed, a mobile version is also feasible.

4. Conclusions

The modified e-discovery reference model described in this paper is augmented based on our extensive experience with e-discovery cases. The model is simpler than the original EDRM and effectively captures the e-discovery workflow. The information technology architecture based on this model can support the entire e-discovery process and guarantee that non-disclosure agreements and information confidentiality are preserved. A mobile implementation has proved to work very well in real cases.

References

[1] M. Daley and K. Rashbaum (Eds.), The Sedona Conference Framework for Analysis of Cross-Border Discovery Conflicts: A Practical Guide to Navigating the Competing Currents of International Data Privacy and e-Discovery – 2008 Public Comment Version, The Sedona Conference, Sedona, Arizona, 2008.

[2] EDRM, Electronic Discovery Reference Model, St. Paul, Minnesota (www.edrm.net).

[3] R. Losey, e-Discovery Team, Orlando, Florida (www.ralphlosey.wordpress.com).

[4] R. McKemmish, What is forensic computing? *Trends and Issues in Crime and Criminal Justice*, no. 118 (www.aic.gov.au/publications/tandi/ti118.pdf), 2002.