Conceptualizing Data Security Threats and Countermeasures in the E-Discovery Process with Misuse Cases

Janine Spears
College of Computing and Digital Media, DePaul University, Chicago, IL, United States., jspears@cdm.depaul.edu

Follow this and additional works at: http://aisel.aisnet.org/amcis2012

Recommended Citation
http://aisel.aisnet.org/amcis2012/proceedings/ISSecurity/17

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Conceptualizing Data Security Threats and Countermeasures in the E-Discovery Process with Misuse Cases

Janine L. Spears
DePaul University
jspears@cdm.depaul.edu

ABSTRACT

During a lawsuit, an organization is required to discover and produce relevant electronic data. In many cases, relevant data includes confidential data, such as personal information or trade secrets. During the course of a lawsuit, the discovered data may exchange many hands. This study analyzes data security threats and corresponding countermeasures within the e-Discovery process by constructing a misuse case diagram. The analysis revealed seven data security threats, the agents who may carry out such threats, and twelve countermeasures. Of the twelve countermeasures identified, two require advanced planning and investment, while the remaining ten are inexpensive procedural controls. Thus, organizations can significantly improve data security during e-Discovery at relatively low cost. Misuse case diagrams used for visual conceptualization of information security can be used as a means to brainstorm and communicate security risk and controls with stakeholders of an e-Discovery process.

Keywords
Information security, E-Discovery, legal hold, litigation hold, misuse case diagram, conceptual modeling
INTRODUCTION

At any given time, the average U.S. business is involved in an estimated three hundred lawsuits (ISACA 2011). Moreover, organizational records that are needed as evidence in lawsuits are typically in electronic format. With the proliferation of communication technologies (e.g., email, instant messaging, text messaging) and mobile devices, identifying, preserving, and collecting data relevant to a lawsuit can be an enormous task. Adding to this challenge, the U.S. Supreme Court updated the Federal Rules of Civil Procedure in 2006 to include a section on electronic discovery requiring that, for lawsuits filed in federal court, an enterprise must be able to produce electronically stored information (ESI) as evidence within a practical timeframe. Many states have copied these rules for e-Discovery.

Electronic discovery, hereafter referred to as e-Discovery, is the process of identifying, preserving, collecting, preparing, reviewing, and producing electronically stored information (“ESI”) in the context of the legal process (The Sedona Conference 2010b). Though e-Discovery is most often associated with civil litigation, e-Discovery is also relevant to criminal litigation and regulatory compliance. For court admissibility, discovered data must be available and maintain integrity. In addition, based on the nature of the legal matter, discovered data may include confidential or otherwise valuable information. Therefore, data security risks need to be managed throughout the e-Discovery process. The present paper makes a contribution by identifying and mapping threats and countermeasures to business objectives within the e-Discovery process.

The remainder of the paper is organized as follows. The next section provides an overview of the E-Discovery processes. Next, an analysis of data security threats within the e-Discovery process is conducted, followed by suggestions for countermeasures. Finally, a discussion section includes the study’s contribution, limitations, and conclusions.

THE E-DISCOVERY PROCESS

The e-Discovery process begins when an organization reasonably expects legal action, at which point data stewards and record custodians are identified. Stewards are those individuals who work with the parties or data involved in the legal matter, while record custodians are individuals responsible for the physical storage and protection of records throughout their retention period (The Sedona Conference 2010b). Stewards and custodians help identify relevant electronically stored information (ESI), and are also required to preserve relevant ESI until the legal case has been resolved. A discoverable document includes any designated documents or ESI, including writings, drawings, graphs, charts, photographs, sound recordings, and images stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonable usable form (Federal Rules of Civil Procedure 2010:58).

Figure 1 depicts the e-Discovery process and follows the E-Discovery Research Model (EDRM 2005)\(^1\) with two variations explained below. In the beginning of e-Discovery, relevant data must be identified. Once an initial understanding is gained on the sources of relevant data, a legal hold is issued to preserve the relevant ESI. According to the Sedona Conference (2010b)\(^2\), a legal hold, also referred to as litigation hold, is “a communication issued as a result of current or reasonably anticipated litigation, audit, government investigation, or other such matter that suspends the normal disposition or processing of records.” That is, whenever litigation or investigation is reasonably anticipated, threatened, or pending against an organization, that organization has a duty to undertake reasonable and good faith actions to preserve relevant and discoverable information as evidence (The Sedona Conference 2010a). For example, during a legal hold while relevant ESI is to be preserved, data backup tapes containing relevant ESI should be pulled out of rotation to prevent relevant ESI from being over-written. Similarly, during a legal hold, data retention schedules for relevant ESI are frozen so that data does not get destroyed. Relevant ESI is then collected from identified sources. Collected data are synthesized to exclude non-relevant data and then converted, as necessary, to formats conducive to legal review and analysis; this stage is referred to as processing in the EDRM (EDRM 2005). After legal analysis, ESI is produced (i.e., delivered) to outside parties, such as external counsel or the opposing party. The produced ESI is then presented in court. Upon resolution of the legal matter, the legal hold is released; that is, normal disposition and processing of records can be resumed. For the purposes of this study, the e-Discovery process in Figure 1 is depicted as a linear workflow, while the EDRM is said to be non-linear and iterative.

---

1 The E-Discovery Reference Model was developed by a working group with the same name. The model is presented and described at http://www.edrm.net and is widely referenced in industry when discussing E-Discovery.

2 Sedona Conference is a non-profit, non-partisan research and educational institute that has published guidance on E-Discovery that has served as guidance in court cases.
Figure 1 illustrates the stages of an e-Discovery process.

Though “e-Discovery” is inherently focused on electronic information, both paper and electronic data are included in this study and hereafter referred to simply as “data”. Given the close relationship between paper and electronic documents in a legal case (e.g., paper documents can be scanned and presented as evidence), data security is considered to apply to both forms for the purpose of this study.

ANALYSIS

Conceptual models provide a means for an analyst to gain a greater understanding of the problem domain (Burton-Jones and Meso 2006; Wand and Weber 2002). As such, conceptual modeling was used in this study to identify security threats within the e-Discovery process. By visually depicting components within the model, the researcher (i.e., analyst) gains a greater understanding of the domain by seeing how model components relate to each other and where analytical gaps exist. A misuse case diagram was the analytical technique chosen for the present study for two reasons. First, a research stream exists on how a misuse case diagram may be used to integrate security considerations into IS functional modeling (Alexander 2003; Hope et al. 2004; Matulevicius et al. 2008; Sindre and Opdahl 2000; Sindre and Opdahl 2008). Secondly, just as use case diagrams are particularly useful when analyzing a business process where stakeholder (e.g., user) interaction is of primary interest, misuse case diagrams are particularly useful when analyzing data security threats posed by stakeholders. In the context of e-Discovery, various stakeholders are involved in the process. Therefore, modeling their interaction with the system (or process) facilitates the development of threat scenarios.

A use case diagram depicts how stakeholders (called “actors”) interact with a system, and the desired system functionality from the stakeholder’s perspective. Conversely, a misuse case diagram provides a means to model undesirable system events that threaten successful completion of the system functions that were modeled as use cases (Sindre and Opdahl 2008). If a business function carried out within a system is viewed as an organizational objective and represented as a use case, then a misuse case depicts a threat to an organizational objective. That is, while use cases illustrate desirable system functionality, misuse cases illustrate undesirable events that could occur and disrupt the desirable system functionality. In a misuse diagram, threats are modeled as misuse cases, threat agents as “mis-actors”, mitigating controls that counter the specified threats as use cases, and the associations between these components (Sindre and Opdahl 2000; 2008). In the present study, a misuse diagram was constructed as a means to analyze data security threats within the e-Discovery process.

Conceptually Modeling Misuse in the E-Discovery Process

Prior to constructing the misuse case diagram, a normal use case diagram was constructed (Sindre and Opdahl 2000) by referencing the activities in Figure 1 and identifying the relevant actors for each use case. Threats were then identified by considering what could go wrong (i.e., threaten) the successful completion of the normal use cases (Peterson and Steven 2006) and by consulting industry frameworks on e-Discovery (EDRM 2005; The Sedona Conference 2010a). These threats were modeled as misuse cases. An association was drawn between the misuse case and use case whose successful completion was threatened; the arrow is from the misuse case to the use case and labeled as “threatens” (Sindre and Opdahl 2008), as shown in Figure 2. Next, the mis-actors that may actually carry out the misuse case were identified. Both misuse cases and mis-actors are shaded in gray in Figure 2 to increase readability. Countermeasures to each threat were identified and modeled as use cases, shaded in yellow in Figure 2 to increase readability. Finally, countermeasure use cases were associated with misuse cases, with an arrow drawn from the countermeasure to the misuse case and labeled as either preventive or detective (Sindre and Opdahl 2000). In the next sections, the content in Figure 2 is described.
Legend: (1) Gray indicates threats and threat agents, modeled as misuse cases and mis-actors, respectively.
(2) Yellow indicates data security countermeasures modeled as use cases.
(3) White indicates business process objectives and stakeholders, modeled as use cases and actors, respectively.

Figure 2. A Misuse Case Diagram for the E-Discovery Business Process
Threat Agents as Mis-actors

The actors in an e-Discovery process often include data stewards and custodians, collectively referred to as “key players” (The Sedona Conference 2010a), who are familiar with and/or have possession of data relevant to the particular legal case. Also involved is inside legal counsel and internal IT personnel who help with data collection (Heikkila 2008). Various types of external service providers may be involved in the e-Discovery process, such as outside legal counsel, the legal team’s document service providers, external backup tape vendors, digital forensic experts, or other e-Discovery vendors. Ideally, the role of IS security will also be involved. Though the IS security role may be among IT personnel, a distinction is made in Figure 2 between these two roles since the work performed is different. From the perspective of Figure 2, the IT personnel work on preserving, collecting, and converting relevant ESI, while the IS security role is focused on implementing controls that manage data security risk. The roles of e-Discovery project manager, compliance manager, and records manager may also be part of an e-Discovery team (Heikkila 2008); for the purpose of this study, those roles may also be considered as key players in that they aid with the identification and preservation of relevant ESI.

Data security threats may be carried out by threat agents (e.g., roles, systems) that are internal or external to an organization, and their adverse actions may be intentional or accidental. Consequently the threat agents, represented as mis-actors in Figure 2, include the same roles that handle the relevant ESI. For some threats (i.e., misuse cases) in Figure 2, a particular mis-actor is singled out and linked to that particular misuse case because that mis-actor represents a higher impact or higher likelihood of carrying out the given misuse case. In other misuse cases, any person handling the data who was identified above as an actor may adversely impact data security during data handling. To increase model readability, this latter mis-actor is generically referred to as data handler in Figure 2. For example, any legitimate actor handling the data may breach data integrity or may compromise the authenticity of the data. Legacy systems also pose a threat to data security in E-Discovery in that needed data may not be reasonably accessible due to technological obsolescence. Finally, external hackers (or more correctly, crackers with malicious intent) pose a threat to circumvent security measures.

Threats as Misuse Cases

A threat is essentially something that could go wrong and would result in an undesirable outcome. From a data security perspective, threats within the e-Discovery process may arise from either accidental or intentional behavior. Natural disasters such as hurricanes, and manmade disasters such as fires, also pose threats to data security. However, for simplicity purposes, natural and manmade disasters are not addressed in this study. Next, a description of each misuse case in Figure 2 is provided.

Failure to timely locate relevant data. Litigation cases (e.g., Zubulake v UBS Warburg 2004) have illustrated exorbitant costs that organizations can incur from not being able to readily identify relevant ESI. This threat is modeled as a data security threat in that the data are not available when needed. Conversely, a threat also exists that an organization will provide too much data to the opposing party. Though this threat was not explicitly modeled as a data security threat, countermeasures in Figure 2 (discussed later) for data unavailability may also counter the threat of providing opposing parties too much data.

Delete relevant data. Unauthorized data destruction is a data security threat that could conceivably result from either accidental or intentional mis-actor behavior. For example, accidental data destruction could occur in cases where an organizational member thought the relevant tape backups were taken out of rotation, only to later discover additional tape backups were relevant that had not been taken out of backup tape rotation, so ultimately were over-ridden, destroying relevant data. Malicious data destruction could occur, for example, in cases where an employee who has a significant loss or gain at stake in the legal case may have a motive to destroy relevant data by shredding or deleting relevant documents. This is an example of spoliation and illustrates malicious intent in that the employee’s actions are intended to unduly influence the outcome of the legal case.

Failure to retrieve legacy data. An example of this threat is legacy software or hardware not being available to access relevant ESI that could potentially support the organization’s position in the legal case.

Breach to data integrity. For court admissibility, data must have integrity (Chisholm 2010). Integrity is a condition existing when data is unchanged from its source and has not been accidentally or maliciously modified, altered, or destroyed (NSTISSC 2000).

Compromise to data authenticity. For court admissibility, data must demonstrate authenticity, meaning that there is some proof a document comes from the person, organization, or other legal entity claiming to be its author or authorizing authority (ARMA 2009).
Keep data beyond usefulness. Legal cases may include confidential data such as trade secrets and personal information (Heikkila 2008). If that confidential data are not destroyed upon completion of the legal case, the threat of unauthorized exposure remains, and protection of this confidential data is at the mercy of the external service provider until such data are destroyed. Therefore, security risks to confidential data must be managed throughout the data lifecycle, including data destruction (e.g., Prosch 2009). As long as data exists, a threat to its confidentiality exists, even if the data is no longer in use. This threat is particularly noteworthy in cases where confidential data has been provided to external service providers during the e-Discovery process.

Hack outside counsel’s web portal. As standard practice, an outside legal counsel places on its web portal a client organization’s ESI that was collected for a legal case (Heikkila 2008). In doing so, the ESI is electronically accessible to the legal team working on the case. Security weaknesses with an external service provider’s environment may result in a breach to confidential, integrity, or availability of the client organization’s ESI. The breach could be performed by an external hacker unfamiliar with the particular legal case. Alternatively, poor security controls could enable an employee of an external service provider to copy relevant ESI and give or sell the ESI to someone on the opposing side.

Countermeasures as Use Cases

This section contains a description of each countermeasure in Figure 2.

Maintain organizational records inventory and an asset inventory. Implicit in this countermeasure is that a record and an asset inventory need to be created prior to the need for e-Discovery in order to counter the risk of failing to identify data in a timely matter. In addition, given that records and computing assets evolve over time, inventories need to be routinely updated in order to be of value to e-Discovery.

Maintain central(ized) document management. A centralized document management system facilitates the search (i.e., identification) of relevant data and later retrieval. Conversely, a less desirable alternative is to have documents, including email archives, stored on user workstations and other distributed storage locations.

Maintain inventory of collected data. An inventory should be maintained of what data were collected for the e-Discovery case. Such an inventory enables an organization to establish tighter control, including the ability to detect if data later come up missing from theft or unauthorized destruction. For example, if documents were inventoried and assigned a sequential number, any documents later missing from that numerical sequence would be detectable.

Log system access. By maintaining logs of who has accessed databases, network drives, or other computing resources containing the relevant data, an organization may be able to detect any deviant behavior from someone engaged in unauthorized data modification or destruction.

Archive essential records when there is a change in technology. If an organization upgrades its technology (e.g., transaction system; email system; tape backup system, etc.) and the data retention period for the affected data has not expired, the organization should maintain the ability to access data from the legacy system, even if the data are no longer in daily use. One means to do this would be to export the data to flat files (e.g., text or XML) and then archive the data. The key point is to have a strategy in place so that data from legacy systems remain available for as long as the data is retained.

Consult e-Discovery integrity experts. Examples of such experts include a digital forensics specialist or a consultant specializing in document processing for e-Discovery. The purpose of including such a specialist is to ensure that data maintains its integrity for court admissibility.

Compute hash value. A best practice method for being able to prove data integrity is to compute a hash value for said data. Hashing is defined as the process of taking an amount of data, such as a file or bit stream image of a hard drive, and applying a mathematical algorithm to generate a numerical identifier (the hash value) unique to that data (Chisholm 2010). For example, suppose a hash value is computed on a data file, that file is later changed, and a hash value is computed again. The two hash values will not match, indicating that the contents of the file have changed, and therefore, integrity is in question. Digital forensics experts advise that the initial hash value should be recorded into the chain of custody (defined in next countermeasure) and verified at each stage of data access or transfer (Chisholm 2010). As such, hashing and chain of custody are two countermeasures that should be used together to protect both data integrity and authenticity.

Produce a chain of custody. A key method for meeting the requirement of data authenticity for court admissibility is to maintain a chain a custody, defined as the chronological documentation and/or paper trail showing the seizure, custody, control, transfer, analysis, and disposition of evidence, physical or electronic (EDRM 2005). The purpose is to be able to
prove who has had access to the data as the data are transferred from person to person throughout the e-Discovery process. A chain of custody should be established at the point of data collection (Chisholm 2010; EDRM 2005).

**Cancel legal hold.** A legal hold goes into effect from the time an organization has a reasonable expectation that litigation is forthcoming until the time the legal case has been settled. Once the legal hold is canceled, normal data retention schedules are again applicable. Given that a litigation case can take multiple years to resolve, it is possible that lifting the legal hold gets overlooked. Alternatively, it is possible that all of the appropriate parties, both internal and external to the organization are not notified that the legal hold has been lifted. Consequently, data is maintained beyond its usefulness, and in cases where confidential data is involved, the data is at unnecessary risk of exposure. Thus, explicit plans should be in place for whom and when to notify that the legal hold has been lifted, so that the data can be destroyed if no longer needed. This is particularly true for external service providers with access to confidential data related to the legal case.

**Ensure secure deletion of delivered data.** Once the legal matter has been concluded, any third parties with access to an organization’s data should be required to destroy said data. Heikkila (2008) recommends using an independent security firm to validate the data’s safe destruction.

**Assess vendor security controls.** Any vendor hosting a web portal for legal counsel that has access to a client organization’s confidential data should be required to demonstrate that security controls that the client organization deems important are in place and functioning. Such assurance could be provided through certifications, detailed self-administered questionnaires on security, external audits of security controls, etc. If the vendor has strong security controls, a breach is less likely to occur.

**DISCUSSION**

Conceptual analysis yielded seven threats to data security within the e-Discovery process and twelve corresponding countermeasures. Legitimate participants in the e-Discovery process, both internal and external to an organization, present threats to the security of relevant data. These threats are attributed to both accidental and intentional behavior. Of the twelve countermeasures revealed in the analysis, only one was a technical control, while the other eleven were operational or procedural in nature. All three countermeasures aimed at facilitating data identification require an organization to have them in place prior to an e-Discovery effort. For example, records and asset inventories, along with centralized document management systems, need to be established in advance so that an organization can more easily and cost effectively locate relevant data. This suggests that organizations must plan in advance in order to reduce the threat of an inability or costly data identification for e-Discovery.

Of the twelve countermeasures identified, only two may be considered somewhat costly (centralized document system and use of third party e-Discovery experts), while the remaining ten were procedural in nature, and thus relatively inexpensive. This finding suggests that organizations can significantly improve data security during e-Discovery at relatively low cost. Basic procedures, such as computing hash values and ensuring secure deletion of delivered ESI, can protect data integrity for court admissibility and confidentiality (e.g., trade secrets or personal information) at the conclusion of the legal case.

**Research Contribution**

While there have been some pioneers who have studied this domain (e.g., Heikkila 2008; Lomas 2010), there is clearly a dearth in research on how to protect data security during the e-Discovery process. The present study makes a research contribution by examining and documenting data security threats within the e-Discovery process. Study results suggest that countermeasures that manage the threat to data identification require advanced planning and investment, while security threats within the remainder of the e-Discovery process can often be countered with inexpensive procedural controls.

A second research contribution is made in constructing a detailed misuse case model. There is a stream of literature on misuse cases that describe why and how to construct the diagrams (see literature review in Sindre and Opdahl 2008). Additional research has been needed that provides detailed examples of integrating security into use case diagrams. The present paper attempts to make such a contribution. In addition, the literature on misuse diagrams appears to be geared toward system developers. However, the present paper illustrates that researchers can also use this modeling technique as a means to analyze security threats within a problem domain.
Implications for Practice

In an industry study (Ponemon Institute 2011), training was found to be the leading preventive measure taken by organizations following a data security breach. The results from this study can be used by organizational practitioners to communicate with and train employees on data security risks in e-Discovery before a costly data breach occurs. In other words, a visual image of security threats and countermeasures is likely to facilitate communication among stakeholders of a subject that is reasonably complex, such as identifying and countering information security risk. Findings from this research may also be used as a detailed example for IS students of how to integrate security concerns in system development models. Indeed, IS scholars have found that IS security is scarcely covered in the vast majority of systems analysis and design textbooks (Biros et al. 2007; Parrish et al. 2009). Misuse case diagrams can be used to conceptualize security risks and countermeasures within a given business process.

Study Limitations

Given the constraints of modeling “everything” in a single diagram, it is likely that a significant data security threat, countermeasure, or association is not included in the model. For example, a chain of custody should be maintained from the point of data collection through presentation (Chisholm 2010). However, drawing three lines from the “chain of custody” countermeasure use case to the relevant business objective use cases would have significantly convoluted the diagram; a compromise was made to only model “providing” a chain of custody for the use cases for data production and presentation. Nonetheless, the model does contain a reasonable amount of detail in a visual form that may spur additional discussion by organizational stakeholders on security threats and countermeasures.

Similarly, space constraints in a model also result in a visually busy model containing many lines. However, while conducting the analysis the author did not find this limitation to threaten the ability to better understand the problem domain. Indeed, mental connections were made that likely would have gone unnoticed had this analysis been conducted strictly with words, as opposed to an image.

Future Research

In building on the present study, a contribution would be made by developing misuse case descriptions that provide detail behind each misuse case in the model that could serve as threat scenarios. Similar to use case descriptions, misuse case descriptions provide more actionable steps that could, in the context of data security, be used by security professionals to design and implement security controls for data within an e-Discovery process. Given that individual legal cases can span years, the investment to develop security measures, tailored to individual legal cases, would be useful. Extant literature has provided some guidance on how to construct misuse descriptions (Sindre and Opdahl 2008), though a standard has not yet emerged. Further developing a standard for misuse descriptions would be helpful future research.

Many of the countermeasures described in the present study were of a procedural (i.e., operational) nature. Future research is needed on administrative controls for data security in e-Discovery. In particular, research on organizational policies that define IT governance around the e-Discovery process would be valuable.

CONCLUSION

A misuse case model was constructed as a means to analyze data security threats and countermeasures within the e-Discovery process. Study results suggest that countermeasures that manage the threat to identifying relevant data require advanced planning and investment, while security threats within the remainder of the e-Discovery process can often be countered with inexpensive procedural controls. Misuse case diagrams used for visual conceptualization of information security can be used as a means to communicate and brainstorm security risk and controls with non-security professionals who are stakeholders in an e-Discovery process.
REFERENCES