

5-21-2019

Electronic Evidence and Technology-Assisted Review

Brian Almquist

MacEwan University School of Business, almquistb@macewan.ca

Follow this and additional works at: <https://aisel.aisnet.org/mwais2019>

Recommended Citation

Almquist, Brian, "Electronic Evidence and Technology-Assisted Review" (2019). *MWAIS 2019 Proceedings*. 25.
<https://aisel.aisnet.org/mwais2019/25>

This material is brought to you by the Midwest (MWAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MWAIS 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Electronic Evidence and Technology-Assisted Review

Brian Almquist

MacEwan University School of Business

almquistb@macewan.ca

ABSTRACT

We consider the technological developments relevant to procedures characterized as “discovery” in legal proceedings over the last two decades. As storage technology has improved, the idea of a “document” as a piece of discrete discoverable information is becoming increasingly outdated. The courts, while adapting to the new technologies, have also had to reckon with increasing costs of document production, and more expensively, review of information for relevance and privilege or work-product protections. Technology-Assisted Review, using machine learning, is rapidly being employed to handle large document productions. This brings eDiscovery right in line with some of the most interesting current developments in information technology.

Keywords

Electronically stored information, electronic discovery, technology-assisted review, information governance

INTRODUCTION

News watchers of the day might find that current discussions of the American political environment are rife with threatened and real lawsuits and investigations. A particularly ominous aspect of these mechanisms is the potential for “discovery,” which might compel an organization to provide sensitive information to either investigators or the court. Most organizations are aware that their electronically stored information (ESI), in the form of traditional business records and regular communications, are subject to regulation and discovery, subpoenas, or even search warrants. (Volonino, 2003) Companies must now plan to retain electronic records to remain in compliance with laws and exchange regulations.

The process of discovery looms large in popular culture. Popular legal dramas on television frequently feature the sequence where a noxious defendant, forced to hand over their files, provide overwhelming in unsorted boxes to understaffed teams of plucky young plaintiff’s attorneys. In the subsequent montage, the “smoking bullet” is found by the exhausted protagonist and justice is delivered in short order. (Apted, 1991). The reality of ESI is that this scene is obsolete. As businesses moved to electronic storage of documents for convenience and cost-reasons, the laws and regulations for retention and delivery of legally pertinent business documents have adjusted to reflect the potential for rapid search and document duplication.

E-DISCOVERY

Since 1970, the Federal Rules of Civil Procedure made ESI discoverable in legal proceedings. Additionally, the Federal Rules of Discovery mandates that organizations involved in legal proceedings have a duty to preserve relevant documents and records as they pertain to a case and to deliver requested records in a “readable format by a specified date.” (Volonino, 2003; Fed. R. Civ. P., 1970). These documents are not limited to documents of business transactions but can include communications including e-mail and instant messages. Electronic evidence is any ESI “that can be used as evidence in a legal action.” (Volonino, 2003). This material is subject to forensic searches for evidence that still exists or had been (inappropriately) deleted. In turn, deletion of relevant records can lead to charges of criminal destruction of evidence.

Before a legal proceeding reaches trial, each side has the right to “discover” as much as possible about their opponent’s case. As such, discovery requests are made for any information that is relevant to the facts that lead to litigation. Since parties are required to provide this information in a readable format by a specific date, industry best-practice ESI management strategies can prevent unnecessarily giving the courts the impression that an organization is attempting to obstruct justice. In turn, poorly managed ESI can lead to expensive processes involving the location, restoration, sorting and review of long-forgotten documents. Given the size and storage technology used, along with the particular circumstances of the discovery request, the time and effort involved in producing relevant evidence for a legal proceeding can reach millions of dollars. (Volonino, 2003). Companies with a comprehensive and “systemic” policies and procedures for review, retention and destruction of

documents produced as part of business operations may find themselves better positioned to respond effectively and at less expense when facing legal obligations – a situation that may be inevitable depending on the organization’s industry sector.

Of course, “relevance” is often a source of contentious discussion between parties. Because the process of discovery can be expensive, discussions of “relevance”, “responsiveness” and “privilege” can be subject to negotiations between both parties to a suit. In many cases, the results of these negotiations can be formatted in a structured manner that resembles the Boolean search term queries used for many research databases, a format that many lawyers will have familiarized themselves with through their training searching for appropriate precedents in case law databases. (Baron, Lewis, & Oard, 2006; Blair & Maron, 1985) As search technology has improved, strategies for managing document collections have adjusted, and the question of “relevance” has moved beyond the presence of Boolean search terms in a document and more towards the paper-trail idea of “who communicated what to whom, when, and, to the extent possible, why.” (Ashley & Bridewell, 2010; Conrad, 2010) Just as documents that are relevant must be found and turned over to a requesting party, subsets of documents are excluded from production to the requesting party. Similar techniques to those used to identify relevant documents in a collection may also be used to identify (and thus withhold from production) those documents that represent work product or privileged communications.

BUSINESS CONSIDERATIONS

As mentioned, even with well-maintained systems, review is expensive. For poorly maintained systems, costs can include recovery and sorting. A thorough search can cost millions of dollars: the search for relevant e-mails in the Monica Lewinsky scandal were estimated at \$11.7 million. (Streza, 2003; Volonino, 2003) Businesses are responsible for maintaining their digital repositories. If they have consistent and justifiable policies for retention (and destruction) of records, these can be used to address concerns about obligations to retain and produce ESI. Maintaining the records in easily searchable formats is important for speed of processing. One problem for organizations has been reduced over time. Physical media backups were once subject to discovery leading to expensive reconstruction, especially as they were typically organized by the operating system and not for human access. As the courts have dealt with these issues for decades, and as the costs of recovering inaccessible records increase, so does the likelihood that the requesting party will be required to bear some portion of those costs, especially as the likelihood of finding relevant evidence decreases. To that extent, it is now expected that production requests against disaster recovery systems are to only be issued when absolutely necessary. (The Sedona Conference, 2018)

SOME HISTORY

When the Federal Rules of Civil Procedure were amended in 1970, it intended to expand the definition of “documents” that were subject to discovery in federal cases to include “electronic data compilations,” noting the possibility that producing electronically stored information from certain systems might require generating human-readable printouts. By 2006, it became clear that the term “document” was inadequate to describe the growing varieties of ESI that could contain information relevant to discovery. For instance, data maintained in databases might differ significantly in structure from “fixed expression on paper.” The amended rules also attempt to require that produced information cannot be manipulated in format to obscure relevant information. For instance, if the data is searchable on the respondent’s system, then that feature must also be available to the requesting party.

The Sedona Conference is an organization that seeks to “move the law forward in a reasoned and just way”, primarily through education and creation of “consensus commentaries” (The Sedona Conference, n.d.) concerning a variety of technology related legal issues. Their oldest working group and series of commentaries focus on eDiscovery concerns. *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production* was first produced in 2004, revised in 2007 to reflect the recent amendments to the Federal Rules of Civil Procedure, and most recently updated in 2017 to reflect evolving technologies and practices related to the management and production of ESI. The Conference supplements the publication of these principles, and the accompanying commentaries, with glossaries, guidelines for managing information and records, and detailed commentaries on eDiscovery-related topics.

The Sedona Conference Working Group on Electronic Document Retention (often referred to as WG1), operates effectively as a permanent task force. Besides regular updates to the Sedona Principles, they have concerned themselves with technological improvements that potentially have an impact on eDiscovery-related best practices. For instance, the industry-standard Boolean-style searches for relevant documents within collections of ESI both miss many responsive documents while still generating large sets of documents that need to be reviewed for relevance and privilege. (Blair & Maron, 1985) Working with the National Institute of Standards and Technology, the Sedona Conference supported five years of specialized tracks at the annual Text REtrieval Conference (TREC) with the initial goal of increasing the number of relevant documents in search results without increasing the size of the result set. As the conference progressed, more sophisticated problems

emerged, including the possibility of using new assessments of relevance to inform, or train, systems for improved queries. By the fifth year, in 2010, the efforts at TREC were focused on improving predictive assessments of relevance and incorporating privilege as an aspect of manual review.

TECHNOLOGY-ASSISTED REVIEW

In the 2017 update to the Sedona Principles, the authors note that the preceding decade had seen the “emergence of search, retrieval, and review methods for discovery, now known as technology assisted review.” The goals of the TREC-activities found themselves in processes and products designed to reduce the burdens and costs of the discovery process. In its broadly applied form, Technology Assisted Review (TAR) involves using software-based classifiers to sort documents into categories “based on input from expert reviewers.” Of interest for the discovery process is identifying records that are responsive to a production request (relevant to the subject of the legal proceeding and must be turned over) or privileged (excluded from discovery for a variety of legal reasons).

The model for TAR in legal proceedings also follows a path partially defined by the technology. Parameters, including goals and protocols for the process must be set. Human reviewers, usually trained lawyers, are involved, and they need to be educated on the system, the subject of the proceedings, and other concerns that will affect the classification of the documents. The process enters a cycle where a sample of documents from the set to be considered are manually classified by the reviewers, the system predicts results for the remaining documents, the predictions are sampled and tested, and the tested results are evaluated – often using measures derived from information retrieval research – to identify improvements to be made in classification (minimally by adding the sampled and tested results to the training set). The cycle is repeated until the tested results reach the goal conditions for the process. While it might be assumed that human-only review might be a more costly but more accurate method, Grossman and Cormack conducted a study that indicated that TAR can also be used to get improved results in both recall and precision-based measures. (Grossman & Cormack, 2011) As a consequence, vendors have put forth many commercially available TAR solutions to assist organizations preparing for potential discovery.

CONCLUSION

For IS professionals, the implications of all this is significant. In 2002, early guidance was to encourage organizations to develop reasonable and defensible retention and destruction policies for data, stressing that all forms of data handled by an organization needed to be covered by record management policies. Additionally, the human component is addressed by a need for training, compliance procedures, and enforcement mechanisms for these policies. (Volonino, 2003) This has evolved into full-scale data and information governance policies. Generally, these policies to recognize the needs of the stakeholders within an organization and their relationship to the organization’s ESI. From business users, to an organization’s counsel, and their IT department, each assesses information and data within the organization in terms of value (to the operation of the business), duty (to hold and potentially produce for discovery), and risk. All stakeholders must coordinate to ensure that obsolete and non-useful data is appropriately destroyed while useful data remains accessible for the business user. (EDRM (edrm.net), 2012) With documented policies, the organization is best prepared from an argumentation perspective to confront any requests for discoverable information.

REFERENCES

1. Apted, M. (1991) Class Action, United States: 20th Century Fox Film Corp.
2. Ashley, K., & Bridewell, W. (2010) Emerging AI & law approaches to automating analysis and retrieval of electronically stored information in discovery proceedings, *Artificial Intelligence and Law*, 18, 4, 311–320.
3. Baron, J. R., Lewis, D. D., & Oard, D. W. (2006) TREC 2006 legal track overview, in E. M. Voorhees & L. P. Buckland (Eds.) *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, 79–98.
4. Blair, D. C., & Maron, M. E. (1985) An evaluation of retrieval effectiveness for a full-text document-retrieval system, *Communications of the ACM*, 28, 3, 289–299.
5. Conrad, J. (2010) E-Discovery revisited: the need for artificial intelligence beyond information retrieval, *Artificial Intelligence and Law*, 18, 4, 321–345.

6. EDRM (edrm.net). (2012) Information Governance Reference Model. Retrieved February 24, 2019, from <https://www.edrm.net/frameworks-and-standards/information-governance-reference-model/>
7. Grossman, M. R., & Cormack, G. V. (2011) Technology-assisted review in e-discovery can be more effective and more efficient than exhaustive manual review, *Richmond Journal of Law and Technology*, 17, 3.
8. Streza, R. (2003) Discovery unplugged: Should internal e-mails be privileged confidential communications? *Defense Counsel Journal*, 70, 1, 36.
9. The Sedona Conference. (n.d.) About. Retrieved February 24, 2019, from <https://thesedonaconference.org/>
10. The Sedona Conference. (2018) The Sedona principles, third edition: Best practices recommendations & principles for addressing electronic document production, *The Sedona Conference*, 19, 1.
11. Volonino, L. (2003) Electronic evidence and computer forensics, *Communications of the Association for Information Systems*, 12, 457–468.