

**LA REVUE DES DOCUMENTS ÉLECTRONIQUES ET
LA CAUSE
*AIR CANADA V. WESTJET AIRLINES LTD***

M^e Marie-Andrée Vermette
WeirFoulds LLP

Incisive Media

Troisième édition

**ADMINISTRATION DE LA PREUVE ÉLECTRONIQUE
E-discovery**

Les 1^{er} et 2 octobre 2008
Hôtel InterContinental
Montréal

THE MANUAL REVIEW OF ELECTRONIC DOCUMENTS

**Marie-Andrée Vermette
WeirFoulds LLP**

TABLE OF CONTENTS

DESCRIPTION	PAGE
PART I PRELIMINARY STEP: DATA CULLING	2
PART II NECESSARY STEP: MANUAL REVIEW	10
PART III MAKING THE MANUAL REVIEW PROCESS EASIER.....	20
PART IV CONCLUSION	23

In the discovery process, the main purpose of document review is to identify the relevant documents to be produced and the privileged documents to be withheld.¹ However, document review also plays a critical role in the broader context of the litigation: it is the stage where lawyers gain a greater understanding of the issues in the case and where legal strategies begin to develop.²

Document review has traditionally meant paper review. Such a review has now become impractical in many cases due to the dramatic increase in the volume of potentially relevant documents in the digital world compared to the paper world. In addition to eliminating the burden associated with printing all electronic documents for a paper review, an electronic review (i.e. the review of documents in electronic format) offers reviewers a greater ability to deal with significant volumes of electronic documents.³ The use of electronic tools at the review stage can increase the efficiency of the review and assist in reducing the costs associated with document review. This is an important consideration given that the real cost of electronic discovery is generally not in the collection of the data, but in the human cost of reviewing the information.⁴

Given the volume of electronic data involved in certain cases, some have argued in favour of eliminating the step of document review in the e-discovery process. According to this view, after a party has performed data culling, such as the application of search terms, and done some quality control checks with respect to the culling

¹ This paper discusses the review of electronic documents in the context of cases where the parties have the obligation to list or produce all the relevant documents that are in their possession, control or power, which is the case in the Federal Court and in all Canadian provinces except Québec.

² Electronic Discovery Reference Model, "Review Node", available at www.edrm.net/wiki/index.php/Review_Node.

³ Electronic Discovery Reference Model, "Review – Evolution of Document Review", available at www.edrm.net/wiki/index.php/Review_-_Evolution_of_Document_Review.

⁴ M.R. Arkfeld, *Arkfeld on Electronic Discovery and Evidence*, 2nd ed. (Phoenix: Law Partner Publishing, 2008) at p. 5-53.

and review: in large data collections, culling rates often exceed 85%.⁶ However, the strategies adopted must be reasonable and defensible in order to achieve the desired results and to withstand the scrutiny of the Court (and the opposing party). Some culling techniques – such as de-duplication - are widely used and accepted, but attention must still be paid to the manner in which they are applied and to their ramifications. The reasonableness of other culling techniques – such as the application of search terms – will depend on the specific culling criteria applied in each case and whether they are adequate for the purpose of identifying the relevant documents within the overall collection of documents.

The issue of data culling is addressed in Principle 10 of the *Guidelines for the Discovery of Electronic Documents in Ontario* ("**Ontario Guidelines**") which provides the following:

A party may satisfy its obligation to produce relevant electronic documents in good faith by using electronic tools and processes, such as data sampling, searching, or the use of selection criteria, to identify the documents that are most likely to contain relevant data or information.

The Commentary that accompanies Principle 10 sheds some light on its meaning. It states:

Particularly where searches for relevant electronic documents must be undertaken on large computer systems, containing vast amounts of information, including materials that are likely to be irrelevant, it may be impractical or prohibitively expensive to review all that information for relevance and privilege. In such circumstances, it is reasonable for parties to use electronic techniques to search within electronic document sources, in collecting the materials that will be subject to a detailed review for relevance and privilege. **The objective should be to identify a subset or**

⁶ M. Mack and M. Deniston, *A Process of Illumination: The Practical Guide to Electronic Discovery* (Portland, Oregon: Discovery Center for Excellence, 2004) at p. 70; Electronic Discovery Reference Model, "Processing – Scoping Electronic Discovery Projects", available at www.edrm.net/wiki/index.php/Processing_-_Scoping_Electronic_Discovery_Projects.

The most important limitation relates to the lack of standardized terms used in conversations and documents. Because of this reality, it is unlikely that keyword searches retrieve all the documents that are relevant to the search terms employed. Keyword searching simply matches the word without any regard to the meaning of the search term. As a result, the formulation of a query can be difficult if the keywords have numerous synonyms and can be described in numerous different ways. Even if the "right" keywords are used, queries often locate ambiguous uses of the keywords and retrieve "hits" of the words that are not relevant. In addition, a keyword search will not retrieve documents containing a keyword if the keyword is misspelled in the query or in the documents.¹¹

Two of the concepts used to discuss the ability of keyword searching to retrieve the information that is relevant are recall and precision. The recall rate is the number of relevant documents retrieved compared to the total number of relevant documents in the data set. Thus, if an information retrieval technique achieves 70% recall, this means that 70% of all relevant documents were actually found and 30% of all relevant documents were not found. The precision rate, for its part, only relates to the retrieved documents, and represents the number of relevant documents retrieved compared to the total number of documents retrieved. Thus, if a system has a precision

intervention and administration to build. The specific technology used for concept searching produces an index of the data that maps the language use, word patterns, concepts and ideas of a document. This then allows the user to search documents for like ideas or similar concepts without having to match an exact keyword or phrase. See Electronic Discovery Reference Model, "Review - Emerging Technologies", available at www.edrm.net/wiki/index.php/Review_-_Emerging_Technologies.

¹¹ M.R. Arkfeld, *Arkfeld on Electronic Discovery and Evidence*, 2nd ed. (Phoenix: Law Partner Publishing, 2008) at pp. 5-24 – 5-25, 5-26.

Search features typically support either recall or precision. For example, the use of proximity connectors increases precision but decreases recall, and wildcard searches increase recall but decrease precision.¹⁵

Manually-created Boolean queries have generally been found to perform poorly with respect to both recall and precision in a number of studies. The two key limitations of this type of queries is that they require negotiation of a very steep trade-off between recall and precision, and they leave the user in ignorance as to the actual performance of the system. These limitations also apply to other types of query-based methods, such as natural language systems and the "concept searches" built on top of them, which have also been found to have poor results in terms of recall and precision. According to some studies, query-based methods can be expected to retrieve no more than 50%, and in all probability much less than 50%, of the documents they were employed to find, and will do so at the cost of including a very large amount of irrelevant documents in the "hits".¹⁶

Further, the fact that recall and precision are inversely related poses a dilemma to lawyers. On the one hand, recall is critical in litigation because parties have an obligation to produce all relevant documents, and a more complete set of relevant documents reduces the risk of overlooking information that could be critical to the case. On the other hand, precision is also very important because a low precision rate means that a large number of irrelevant documents will have to be reviewed, resulting in

Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery" (2005) 6 Sedona Conf. J. 237.

¹⁵ M.R. Arkfeld, *Arkfeld on Electronic Discovery and Evidence*, *ibid.* at p. 5-28.

¹⁶ B. Hedin, "Searching in all the Wrong Places: The Effectiveness of Search Tools in E-Discovery", January 2007, available at <http://www.h5technologies.com/pdf/searchtools.pdf>.

inform all the e-discovery guidelines that have been developed so far, i.e. the principle of reasonableness and the principle of proportionality.²⁰

The fact that the use of query-based searches is now inescapable, however, does not mean that lawyers should ignore their inherent limitations. Without a proper understanding of the limitations of search and filtering technologies, lawyers will be lulled into believing that they have completed a comprehensive search for relevant electronic documents and will fail to take appropriate steps that could locate additional relevant materials. This could result in an incomplete record that makes it difficult to prove certain aspects of the case, or could provide the basis for motions to compel production and allegations of spoliation. In contrast, lawyers who have a good knowledge and awareness of the limitations of query-based searches will be able to develop a better search methodology. Further, they will be in a better position to attack their opponents' methodology and to ensure that such methodology is adequate and reasonable in the circumstances.

What the limitations associated with keyword searches emphasize, though, is the importance of carefully developing one's list of search terms, of adopting an iterative process (as discussed further below), and of implementing quality assurance mechanisms to test the searching strategies adopted and to measure their performance. All of these steps usually require, as discussed further below, that a manual review be conducted of the documents identified through the application of search terms. Subjecting the entire mailboxes of key witnesses to a message-by-message manual review is another "mechanism" or process that is recommended in order to address the

²⁰ See, e.g. Principles 2 and 5 of the *Sedona Canada Principles*; Principles 2, 5 and 10 of the *Ontario Guidelines*; and Principles 2, 5 and 11 of the *Sedona Principles*.

Failure to employ an iterative process can lead to an incomplete set of evidence which could, of course, jeopardize your case.²²

In short, cases evolve. In many instances, the issues that were thought to be important at the beginning of the litigation are different from those deemed to be important later on. The identification of new issues through the review and analysis of the collected data set frequently reveals the need to collect additional documents or to modify one's list of search terms and filtering criteria because of newly discovered relevant custodians, topics, time frames, or keywords.²³

The *Sedona Principles* recognize the importance of following an iterative process, and state that agreements between parties on the search methods to be used, including search terms and concepts, "should take account of the iterative nature of the discovery process and allow for refinement as the parties' understanding of the relevant issues develop."²⁴

A necessary ingredient of an iterative process is a manual review performed by human reviewers. The mere application of search terms or advanced searching techniques such as concept searching will not, without a manual review of the results, reveal the need to collect additional documents, or the need to apply different search terms because of the particular jargon used by the key players.²⁵ As stated in *The*

²² Electronic Discovery Reference Model, "Analysis – Pitfalls to Avoid", available at www.edrm.net/wiki/index.php/Analysis_-_Pitfalls_to_Avoid.

²³ Electronic Discovery Reference Model, "Analysis – Updating Assessment as Issues Evolve", available at www.edrm.net/wiki/index.php/Analysis_-_Updating_Assessment_as_Issues_Evolve; Electronic Discovery Reference Model, "Analysis- Focusing Collection", available at www.edrm.net/wiki/index.php/Analysis_-_Focusing_Collection; Electronic Discovery Reference Model, "Processing – Searching", available at www.edrm.net/wiki/index.php/Processing_-_Searching; G.L. Paul & J.R. Baron, "Information Inflation: Can the Legal System Adapt?" (2007) 13 Rich. J.L. & Tech. 10 at paras. 50-55.

²⁴ *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production*, 2nd ed., June 2007 (Sedona, Arizona: The Sedona Conference, 2007) at p. 57, available at www.thesedonaconference.org.

²⁵ M.R. Arkfeld, *Arkfeld on Electronic Discovery and Evidence*, 2nd ed. (Phoenix: Law Partner Publishing, 2008) at p. 6-26.

happen in the absence of a manual review -- may be just as damaging to the client's interests and the litigation process as incomplete production.²⁸

Principle 11 of the *Sedona Principles* is to the same effect as Principle 10 of the *Ontario Guidelines* in that Comment 11.a states that search methodologies allow the identification of potentially responsive information, which must then be reviewed to satisfy the party's search obligations.²⁹

In contrast to the *Ontario Guidelines* and the *Sedona Principles*, the *Sedona Canada Principles* contain extensive remarks that deal specifically with the issue of review of electronic documents, and discuss the merits of an automated review compared to a human review. Notably, after stating that the volume of electronic information may still be overwhelming after the application of various electronic tools and processes to reduce such volume, Comment 7.c states that "[r]esearch in the information science field has demonstrated that automated review is statistically more reliable than human review of large data collections for the purpose of identifying relevant electronically stored information", and refers to an American article in support of this statement.³⁰ It is questionable, however, whether this statement is accurate, especially if the expression "automated review" is meant to exclude any human review after an electronic relevancy assessment is performed through the application of filtering criteria and search techniques. The American article referred to does not support such a position. In fact, the authors of the article expressly recognize the importance of

²⁸ *Ibid.* at p. 6.

²⁹ *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production*, 2nd ed., June 2007 (Sedona, Arizona: The Sedona Conference, 2007) at p. 57, available at www.thosedonaconference.org.

³⁰ *The Sedona Canada Principles: Addressing Electronic Discovery*, January 2008 (Sedona, Arizona: The Sedona Conference, 2008) at p. 29, available at www.thosedonaconference.org.

documents, the purpose of which was to ensure that the electronic search was effective.³⁴

WestJet Airlines Ltd. objected to Air Canada's proposed way to proceed. Nordheimer J. summarized WestJet's position as follows:

The principle [sic] dispute between the parties is over the intention of Air Canada not to undertake a manual review of the documents to be produced. WestJet says that a manual review of all of the documents to be produced is necessary both because that is the obligation of a party under the *Rules of Civil Procedure* but also because it is the only effective way of determining whether the documents to be produced are relevant. Given the different meanings that words can have, electronic searches alone cannot be relied upon by themselves to distinguish between documents that use a word in a relevant context over documents that use the same word in an irrelevant context. WestJet also says that, given the importance of solicitor client privilege in our justice system, it is inappropriate for a party not to take all available steps to ensure that privileged documents are not produced nor is it appropriate for a court to countenance that failure by giving a blanket order of the type sought by Air Canada. Further, WestJet says that no electronic search can determine whether a document is properly labelled confidential. A manual review of each of the documents must be undertaken before Air Canada can fairly label a document as Level A or Level B. Simply put, WestJet says that what Air Canada is actually trying to do through this motion is to foist onto the defendants, Air Canada's obligations to identify and produce only relevant and non-privileged documents.³⁵

Nordheimer J. agreed with WestJet's position. After referring to Principle 10 of the *Ontario Guidelines* and the Commentary following that principle, Nordheimer J. stated as follows:

Air Canada says that its proposed manner of proceeding is consistent with Principle #10 and should be endorsed by this court. I do not agree. I accept that the first stage of Air Canada's approach was appropriate, that is, the use of electronic search terms to identify the apparently relevant documents. WestJet does not dispute this. I do not accept, however, that Air Canada's intention not to conduct a manual review of the resulting documents is validated by Principle #10 nor is it consistent with the requirements of the *Rules of Civil Procedure*.

³⁴ *Air Canada v. WestJet Airlines Ltd.* (2006), 81 O.R. (3d) 48 at para. 7 (S.C.J.).

³⁵ *Ibid.* at para. 10.

type of cases would not require a manual review. In light of Nordheimer J.'s comments with respect to privilege, the limitations inherent in the use of search terms and the requirements of the *Rules of Civil Procedure*, and in light of the fact that Air Canada was ordered to conduct a manual review despite the very high number of documents involved, it is difficult to imagine a case where Nordheimer J. would not require that a manual review be performed. Thus, the exception to the manual review requirement appears to be very narrow, but it will probably be clarified as the case law develops.

As found by Nordheimer J., Canadian discovery rules also support the requirement that a manual review be performed. In Canada, contrary to the United States, there is a general duty to produce relevant documents imposed by the different provincial rules of civil procedure.³⁷ These rules typically require the preparation and delivery of an affidavit or list of documents.³⁸ In fulfilling their discovery obligations under the rules of civil procedure, both the parties and their counsel have certain obligations. In Ontario, for example, parties must swear or solemnly affirm that: (1) they have conducted a diligent search of their records and have made appropriate enquiries of others in order to make the affidavit of documents; and (2) the affidavit discloses, to the full extent of the party's knowledge, information and belief, all documents relating to any matter in issue in the action that are or have been in the party's possession, control or power.³⁹ In order to be in a position to make these statements in their affidavits of documents, parties must conduct a manual review to ensure, among other things, that the

³⁷ Except in Québec, which has a different legal regime.

³⁸ *The Sedona Canada Principles: Addressing Electronic Discovery*, January 2008 (Sedona, Arizona: The Sedona Conference, 2008) at p. 8, available at www.thesedonaconference.org.

³⁹ Rule 30.03 and Forms 30A and 30B of the Ontario *Rules of Civil Procedure*, R.R.O. 1990, Reg. 194.

obligations that have been imposed on counsel in the United States, which are similar to counsel's obligations in Canada. In *Zubulake v. UBS Warburg LLC*, Scheindlin J. stated that counsel have a duty to monitor their clients' efforts to retain and produce relevant documents, to make certain that all sources of potentially relevant documents are identified and preserved, and to produce information responsive to the opposing party's requests.⁴³ Similarly, in *Cardenas v. Dorel Juvenile Group, Inc.*, Waxse J. stated that counsel have an obligation to review the documents received from their clients to see whether they indicate the existence of other documents not previously retrieved or produced. This obligation flows from counsel's duty to ensure that their clients discharge in good faith their duties under the discovery provisions of the *Federal Rules of Civil Procedure*, and counsel's duty to exercise some degree of oversight over their clients to ensure that they are acting competently, diligently and ethically in order to fulfill their responsibility to the court and opposing parties. Waxse J. held that counsel had not met these duties in that particular case, and found that the imposition of monetary sanctions was appropriate.⁴⁴

Conducting a manual review after the application of culling criteria is not only required, it also makes sense. Not reviewing search results before production only postpones the time at which the documents will have to be reviewed. Chances are that

Methods in E-Discovery" (2007) 8 The Sedona Conference Journal 189 at p. 209, available at www.thesedonaconference.org under "Publications".

⁴³ 229 F.R.D. 422 at pp. 432-433 (S.D.N.Y. 2004).

⁴⁴ 2006 U.S. Dist. LEXIS 37465 at pp. 22-24 (D.Kan. 2006). For further discussion of the obligations of counsel under the *Federal Rules of Civil Procedure*, see *Qualcomm Inc. v. Broadcom Corp.*, 2008 U.S. Dist. LEXIS 911 (S.D. Cal. 2008), where six attorneys were referred to the California State Bar for failing to make a reasonable inquiry into their client's discovery search and production. According to the Court, a reasonable inquiry should have included the use of certain search terms and the search of the computers of certain custodians. The sanction order was subsequently vacated and remanded to the Court because the attorneys should not have been prevented from defending their conduct by their client's attorney-client privilege: see *Qualcomm Inc. v. Broadcom Corp.*, 2008 U.S. Dist. LEXIS 16897 (S.D. Cal. 2008).

need to be reviewed separately from the general document collection. This can be of great assistance in determining review assignments and resource allocation. Thus, groups of documents that are potentially privileged can be assigned to more senior reviewers, while groups of documents that are likely to be irrelevant can be assigned to first-year associates. Similarly, categories of documents like spam mail may potentially be ignored altogether or assigned to junior reviewers, and foreign language documents can be separated and assigned to reviewers with the appropriate expertise.⁴⁶

Another advantage of these tools is to give the reviewers a more global view of the data set:

Getting a "wide perspective" of the electronic document universe is another highly beneficial aspects of emerging technologies like linguistic patterns and conceptual searching. Lack of structure is one of the biggest challenges when dealing with electronic documents and electronic discovery. When contextual similar documents are grouped together, certain facts about the document universe may, [sic] be discerned. For instance, a grouping of like documents will often result in threads of e-mail messages sent, received, replied to and forwarded among a group of custodians. A reviewer can see the full context of the electronic exchange and make a more informed categorization decision regarding the group of e-mails as a whole, rather than categorizing each message as a stand-alone communication.⁴⁷

As suggested in the excerpt above, searching and grouping technologies allow for categorization consistency. Grouping like documents together and having them

⁴⁶ Electronic Discovery Reference Model, "Review – Emerging Technologies", *ibid.*; Electronic Discovery Reference Model, "Review – Planning the Review", available at www.edrm.net/wiki/index.php/Review_-_Planning_the_Review; R. Frederick, "Efficiency Matters – Using Topic Review Technology in Document Review" (2007) 8 Hearsay 5 at p. 6, available at <http://www2.mnbar.org/sections/new-lawyers/Winter07.pdf>; R. Nunez, "Automate Analysis During Review" (2006) 5 E-Discovery Advisor 20, available at <http://e-discoveryadvisor.com/doc/18303>; P. McLaughlin, "Manage Online Document Reviews for Electronic Discovery" (2006) 5 E-Discovery Advisor 16, available at <http://e-discoveryadvisor.com/doc/18304>; J.R. Baron, ed., "The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery" (2007) 8 The Sedona Conference Journal 189 at pp. 195, 215, available at www.thosedonaconference.org under "Publications".

⁴⁷ Electronic Discovery Reference Model, "Review – Emerging Technologies", *ibid.*

to fulfil the objectives of the discovery process and to meet the parties' obligations in that process. This is likely to be the case as long as the "goal of computational thinking to approximate the ability of human language behaviour" remains unfulfilled.⁵² However, while we are not at a point where fully automated review can replace human review, advanced technologies can be used to focus review efforts on those documents which are most likely to contain relevant information, thereby increasing efficiency, reducing costs and providing superior results.

⁵² J.R. Baron, ed., "The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery" (2007) 8 The Sedona Conference Journal 189 at p. 212, available at www.thesedonaconference.org under "Publications".