E-DISCOVERY AND KEYWORDS: NOT SO KEY AFTER ALL

FACE 2 FACE

A CONFERENCE FOR LITIGATION SUPPORT

CBA - NS

FRIDAY, DECEMBER 7, 2012

HALIFAX, NOVA SCOTIA

HALIFAX MARRIOTT HARBOURFRONT

PRESENTATION BY: DANIELA BASSAN

PAPER BY: DANIELA BASSAN AND IAN BRENEMAN[1]

---

1 ***Daniela Bassan, Partner, Stewart McKelvey; and Ian Breneman, Associate, Stewart McKelvey***.

E-Discovery and Keywords: Not So Key After All

TABLE OF CONTENTS

E-discovery and Keywords: Not So Key After All

INTRODUCTION

To date, there is limited Canadian case law on acceptable e-discovery methodologies, whether by traditional keywords or more advanced computer-assisted review.[2] While keyword searches have been condoned and accepted by courts as part of an appropriate e-discovery process, it is clear that keyword searching, alone, is insufficient.[3]

Courts have emphasized that while costs and proportionality are key drivers in the e-discovery process, there is also an overriding duty to preserve and produce relevant information. The Canadian judiciary has also emphasized the importance of collaboration and cooperation between counsel throughout the e-discovery process. This is in accordance with the Sedona Principles including the Cooperation Proclamation.[4]

Recent case law from the United States has discussed the shortcomings of traditional keyword searches, and endorsed computer-assisted review as a more acceptable way to undertake e-discovery (particularly in large data cases).

The purpose of this paper is to provide a brief introduction to these topics by highlighting a series of cases in the Canadian and US context.

CANADIAN CASE LAW

The leading case on e-discovery procedure in Canada is ***Air Canada v. Westjet Airlines Limited***, 2006 CanLii 14966 (ONSC) ["*Air Canada*"]. In that case, the plaintiff brought a motion to confirm that any privileged documents inadvertently produced during e-discovery would not constitute a waiver of privilege and that production would not constitute an admission of relevance. The motion was dismissed. In the course of its decision, the Ontario Superior Court

---

[2] At a recent online event sponsored by The Sedona Conference, ADMA EDRM Advisory Board, the concept of computer-assisted review was explained as follows (from a webinar entitled: "Statistics – The Truth About Computer Assisted Review" June 24, 2012):

> Also known as "predictive coding," computer assisted review is technology that helps complete review in eDiscovery and other document-intensive matters. The majority of documents are coded by computers, not humans. Humans still play a critical role in computer-assisted review by coding a small sample of documents and helping the computer "learn" about the case.
>
> The difference between computer assisted review and a "traditional" review is that computers perform the majority of the document review work. Rather than training a human team of document reviewers, the computer is trained to perform the review. This training is accomplished by coding a sample set of documents; the computer then analyzes those decisions and applies that same logic with great accuracy to the entire document collection therefore cutting time and costs.
>
> Advanced search and analytics can increase accuracy, reduce over collection and improve efficiency. These techniques include conceptual search, automated clustering & contextual analysis.

[3] Studies have shown that keyword searches find only approximately 20% of relevant data: *see Paul, George L. and J. R. Baron, "Information Inflation: Can the Legal System Adapt?" (2007) 13 Richmond Journal of Law and Technology 10 at pages 22-24:*

http://law.richmond.edu/jolt/v13i3/article10.pdf.

[4] See https://thesedonaconference.org/cooperation-proclamation. Regarding the new "strategic cooperation paradigm" for lawyers in e-discovery, see "*Information Inflation*", *supra*, at paras. 26-35.

of Justice noted that a "diligent search" remains as the overall obligation of the parties, but also recognized that different categories of documents will require different levels of review (para. 17).

Air Canada proposed to use a set of keyword searches and review less than 5% of the results for effectiveness. In expressing its disapproval of that proposal, the Court held as follows:

> *I will say that I would not consider a process that relied almost entirely on electronic searches and a less than 5% manual review (or a 40% manual review as counsel for Air Canada said at the hearing had now been done) to be satisfactory for this purpose. Rather, such a process appears to be much closer to what is discussed in commentary 10.d. to Sedona Principle #10, as "clawback" or "quick peek" production – a practice that is characterized as "ill-advised" in that same commentary (para. 20).*

In ***GRI Simulations Inc. v. Oceaneering International Inc.***, 2010 NLTD 85, the defendants applied for relief from email document production, effectively to amend an earlier consent order setting out which of the defendants' custodians' emails would be subject to disclosure. In the alternative, the defendants sought to have the plaintiff pay for the costs of the e-discovery.

As in *Air Canada*, the Court recognized the Sedona Principles, particularly with regard to proportionality. However, it was held that the Sedona Principles do not override the rules of court with respect of the duty to disclose electronic communications. With regard to costs, the Court held that there was no reason to "depart from [the] long-standing convention" that the cost of producing documents is borne by the producing party (para. 66). The defendants' motion was dismissed.

In ***Shell Canada Limited v. Superior Plus Inc.***, 2007 ABQB 739, the defendant brought an application to compel the plaintiff to produce additional documents. The plaintiff scanned through the titles of its electronic documents manually, and then electronically filtered all of the documents using a keyword search. All documents were then manually reviewed. The defendant's motion brought into question the adequacy of the plaintiff's search terms. The Alberta Court of Queen's Bench noted the similarity of its case to that in *Air Canada*, and summarized the issue and holding in the latter case at para. 34:

> *In Air Canada there were approximately 75,000 electronic documents that were potentially relevant to the action. Air Canada's intention was to use only a variety of electronic searches to identify relevant documents without a follow-up manual search. Nordheimer J. held that this method was insufficient due to the circumstances of the case and Principle 10 of the Ontario Guidelines. What was required was an electronic search to identify potentially relevant records followed by a level of manual review.*

The Court noted that it "may not have been an issue at all if counsel had collaborated on the terms as suggested in the Ontario Guidelines" and that it was "regrettable that this was not done" (para. 36).

The Court dismissed the defendant's motion, noting that there was no expert evidence provided to suggest that the search terms were inadequate, or that the plaintiff's methodology procedure was not appropriate. Regarding the latter:

> The trifecta search approach implemented by the Respondents of manual review, electronic review and a further manual review falls within the Ontario Guidelines and the *Air Canada* reasoning (para. 41).

L'Abbé v. Allen-Vanguard, 2011 ONSC 4000, provides a relatively lengthy discussion regarding e-discovery methodology. That case involved an action regarding a $40 million escrow fund held back from the purchase price of Allen-Vanguard's acquisition of shares of Med-Eng Inc. Each party brought a motion regarding the e-discovery process. In particular, L'Abbe sought sanctions against Allen-Vanguard for breaches of court orders regarding production timelines.

Throughout the Court's decision is an emphasis on the lack of cooperation among counsel. For example, at para. 34:

> *It is not that counsel have not been talking to each other. In fact they have been in regular communication about the difficulties and concerns arising from production and delay. They have reached agreement on certain important points and they have a discovery plan. Nevertheless it has been a process marked more by motions and adversarial requests for case management orders than the kind of collaborative process now envisioned by the rules. This raises the question as to how this might have been done differently.*

The Court goes on to discuss the principles of proportionality, and the efforts that counsel must make to shift their attitudes to ensure a reasonable e-discovery process:

> *[41] ...The following objectives are enshrined in the rules:*
>
> > *General Principle - 1.04 (1) These rules shall be liberally construed to secure the just, most expeditious and least expensive determination of every civil proceeding on its merits.*
> >
> > *Proportionality - (1.1) In applying these rules, the court shall make orders and give directions that are proportionate to the importance and complexity of the issues, and to the amount involved in the proceeding.*
>
> *[42] 1.04 (1) is not new but (1.1) is. Proportionality is, in the words of Perell J. a "parsimonious principle". The thrust of the principle is that use of discovery procedures should be narrowed and focused whenever possible. Counsel should be seeking to focus their production and discovery efforts on the information necessary for resolution of the case on the merits and not on unearthing every stray piece of evidence that might have marginal relevance. While this does not relieve a party of the potential obligation to identify and preserve all relevant documentary evidence, it does create an imperative for counsel to agree on what evidence is crucial and what is not.*
>
> *[43] This kind of co-operation requires an attitudinal shift which cannot be expected to take root instantly...*

Although realizing that the appropriate e-discovery process will only take shape on a case-by-case basis, and that some latitude to counsel should be granted, the Court went on to reaffirm the production duties of each party, as follows:

> *[47] Let me however be absolutely clear. The party having the duty to produce documents has first and foremost the duty to determine what potentially relevant documents it has in its possession, power and control and where they are located. The party then has a duty to discuss the magnitude and scope of proposed preservation, review and production with the other side. Allen Vanguard may be faulted for not initially having accurately determined the scope of production and the technical requirements of recovering those documents. Armed with that information, both parties then have an obligation to try to agree on the most efficient and cost effective process of targeting the necessary production. If that information is not provided or is inaccurate then any plan is doomed to failure.*
>
> *[48] The object of discovery agreement is to avoid massive over production on the one hand and unnecessary delay and expense on the other. It is a benefit to the producing party which will otherwise be required to recover and review all potentially relevant sources of documents and then to review the documents for relevance, privilege and duplication. It is the process of review that is potentially the most time consuming and expensive process. Moreover, in the absence of a process for jointly identifying the target documents, it is then a process that must be repeated by the other party. The receiving party it obliged not only to review all of the documents produced by the producing party but must also consider whether there should be challenges to the claims of privilege in the Schedule B documents. More importantly the receiving party must then guess at what gaps may exist in the production since there is no obligation on a producing party to list in its affidavit of documents those documents that were reviewed and considered not to be relevant. This can lead to lengthy cross examination on the affidavit of documents and to various motions.*

Even in the absence of agreement, the Court noted that the parties have an obligation to engage in a transparent e-discovery process that attempts to meet objections raised by the other side in good faith (see para. 51).

The case of ***Halifax (Regional Municipality Pension Committee) v. State Street Global Advisors Ltd.***, 2011 NSSC 355 ["*State Street*"], is a lengthy decision on document discovery, generally. The plaintiff brought motions to compel the defendants to make disclosure of several categories of documents and to provide directions on that disclosure. The defendants countered with a motion seeking an order to compel the plaintiff to expand the parameters of its search terms for relevant electronic information.

Justice Duncan begins his discussion of legal principles by noting the requirement to make full disclosure of electronic information, per *Civil Procedure Rule* 14.08(1), and the "concomitant duty to find and preserve relevant information as enunciated in Rule 14.08(2)" (paras. 20-22).

The plaintiff provided its custodians with a questionnaire, which included instructions to search all e-mail files on all relevant computer drives. Among a number of other things, the defendants argued that the questionnaire did not adequately explain that the custodians needed to search for relevant emails. The Court agreed, and ordered that the appropriate custodians search their email using 38 keywords which the plaintiffs had chosen in previous rounds of e-discovery, and for email addresses with specified domain names. The method of performing those searches was left within the discretion of counsel for the plaintiff.

Although the Court relied heavily on keyword searches, the plaintiff's custodians were required to individually search their own emails. Although not explicit in the Court's decision, that process necessarily involves the individuals' memory and judgment in finding emails that may be relevant. This can be contrasted with the application of keywords indiscriminately to a massive set of data, in the absence of custodian judgment and knowledge. It also appears as though the next step in the process was a filtering of data by the plaintiff's counsel for relevance, once they received sets of potentially relevant data from the custodians.

In ***Enbridge Pipelines Inc. v. BP Canada Energy Co***., 2010 ONSC 3796, the Ontario Superior Court of Justice granted an order by consent that adopts an agreement between the parties on the discovery process. An appendix to that agreement is a technical production agreement, which provides a sample precedent for the type of agreement that can be reached between parties on an e-discovery process.[5]

UNITED STATES CASE LAW

In ***Da Silva Moore v. Publicis Groupe*** (2012), 11 Civ. 1279 (Southern District of NY, US District Court), Magistrate Judge Andrew J. Peck explicitly held that his "judicial opinion now recognizes that computer-assisted review is an acceptable way to search for relevant ESI [electronically stored information] in appropriate cases". In that case, the parties agreed to use a 95% confidence level to create a random sample of an entire email collection to determine relevant documents to train predictive coding software.[6] The Court in that case very much favoured the use of computer-assisted review, but still noted that e-discovery cannot be an entirely machine-driven exercise:

> *The decision to allow computer-assisted review in this case was relatively easy – the parties agreed to its use (although disagreed about how best to implement such review). The Court recognizes that computer-assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases. The technology exists and should be used where appropriate, but it is not a case of machine replacing humans: it is the process used and the interaction of man and machine that the courts need to examine.*

The Court goes on to discuss the shortcomings of using traditional keyword searches:

> *Because of the volume of ESI, lawyers frequently have turned to keyword searches to cull email (or other ESI) down to a more manageable volume for further manual review. Keywords have a place in production of ESI – indeed, the parties here used keyword searches (with Boolean connectors) to find documents for the expanded seed set to train the predictive coding software. In too many cases, however, the way lawyers choose keywords is the equivalent of the child's game of "Go Fish." The requesting party guesses which keywords might produce evidence to support its case without having much, if any,*

---

[5] In the recent case of *Velsoft Training Materials Inc. v. Global Courseware Inc*., 2012 NSSC 295, the Nova Scotia Court held that in the absence of such an agreement, the Civil Procedure Rules do not require that search criteria be disclosed in an affidavit disclosing documents, itself. However, the Court recognized that if a request for such information was refused, it could be relevant to costs on a subsequent motion for additional electronic production.

[6] Regarding the topic of statistical sampling techniques for the purpose of meeting discovery obligations, see "*Information Inflation*", *supra*, at paras. 47-49.

> *knowledge of the responding party's "cards" (i.e., the terminology used by the responding party's custodians). Indeed, the responding party's counsel often does not know what is in its own client's "cards." (19-20).*

The Court also notes that keyword searches are often over-inclusive, and statistically ineffective. Judge Peck sets out a number of lessons from that case at 23-25:

- It is unlikely that the court will approve an e-discovery proposal until computer-assisted review software is trained and tested;
- The staging of the discovery process, without prejudice, is a pertinent way to control costs;
- One party often has knowledge of the other party's records, and that knowledge should be used in pursuit of strategic cooperation; and
- It may be helpful to have e-discovery vendors provide evidence where the effectiveness and utility of the proposed software is in issue.

Judge Peck's decision was upheld by United States District Judge Andrew L. Carter, Jr. ((2012), 11 Civ. 1279)).

CONCLUSION

In addition to the US case law noted above, there is a growing body of literature which confirms the frailties of keyword-driven e-discovery (e.g. under/over inclusiveness, arbitrariness, inaccuracy, ambiguity, etc.). Instead of focusing on keyword searches only, the emphasis is to establish a more holistic approach to e-discovery which is informed by technology and strategic cooperation. **Schedule "A"** to this paper contains a sample listing of articles and online references on this topic. As sophisticated technologies become more prevalent and accessible, the use of artificial intelligence to review and analyze large data sets should be easier to promote in all jurisdictions. The challenge for clients, counsel, and the Courts is to keep in step with these advances, with a view to creating a culture of cooperative and efficient e-discovery for everyone.

SCHEDULE "A"

Sample references regarding Keyword Limitations and Predictive Coding:

1) "Keyword Searching Has Limitations in E-Discovery", *Inside Counsel*, Feb 2011, http://www.insidecounsel.com/2011/02/01/keyword-searching-has-limitations-in-ediscovery?page=4

2) "What is predictive coding and can it help me?", *Canadian Lawyer Magazine*, Jan 2012, http://www.canadianlawyermag.com/3988/what-is-predictive-coding-and-can-it-help-me.html

3) "Keyword searches not good enough for e-discovery, experts say", *Computerworld*, Dec 2010, http://www.computerworld.com/s/article/352604/Keyword_Searches_Disappoint

4) "In e-discovery breakthrough, judge endorses 'predictive coding'", Alison Frankel's On the Case, Thomson Reuters, Feb 2012, http://newsandinsight.thomsonreuters.com/Legal/News/2012/02_-_February/In_e-discovery_breakthrough,_judge_endorses__predictive_coding_/

5) "The Next Generation of Concept Searching: Clearwell Transparent Concept Search", Clearwell white paper, http://info.clearwellsystems.com/lp-wp-trans-concept-search.html

6) "Transparent Predictive Coding: Cost-Effective and Defensible Technology Assisted Review", Clearwell white paper, http://info.clearwellsystems.com/lp-wp-predictivecoding-clwl.html

7) "E-data costs to 'overwhelm'", *The Lawyers Weekly*, Apr 2012, http://www.lawyersweekly-digital.com/lawyersweekly/3145?pg=22#pg22

8) "Judge Grimm, Victor Stanley, And The Problem of "Black-Box" E-Discovery Search", Clearwell Systems e-discovery 2.0, Aug 2008, http://www.clearwellsystems.com/e-discovery-blog/2008/08/22/judge-grimm-victor-stanley-and-the-problem-of-%e2%80%9cblack-box%e2%80%9d-e-discovery-search/