IN THE SUPREME COURT OF VICTORIA
AT MELBOURNE
COMMERCIAL COURT
TECHNOLOGY ENGINEERING AND CONSTRUCTION LIST

S ECI 2014 000497

McCONNELL DOWELL CONSTRUCTORS (AUST) PTY LTD   Plaintiff
(ACN: 002 929 017)

v

SANTAM LTD (REGISTRATION NUMBER 1918/001680/06)   Defendant
AND OTHERS (ACCORDING TO THE SCHEDULE)

JUDGE:	VICKERY J
WHERE HELD:	Melbourne
DATE OF HEARING:	25 November 2016
DATE OF JUDGMENT:	2 December 2016
CASE MAY BE CITED AS:	McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd & Ors (No 1)
MEDIUM NEUTRAL CITATION:	[2016] VSC 734

PRACTICE & PROCEDURE – Appointment of Special Referee – Questions as to discovery of documents and inspection of documents in a large document case referred to Special Referee – Reference conducted as a facilitation process rather than an adversarial process – Adoption of Special Referees' Report –*Supreme Court (General Civil Procedure) Rules 2015* (Vic) O 50.

PRACTICE & PROCEDURE – Discovery in a large document case– Use of predictive coding technology (TAR or Technology Assisted Review) – Orders for TAR made in accordance with the recommendations of the Special Referee following a facilitation process conducted with the parties in a reference conducted under O 50 *Supreme Court (General Civil Procedure) Rules 2015* (Vic).

APPEARANCES:	Counsel	Solicitors
For the Plaintiff		Baker & McKenzie
For the First Defendant		Holding Redlich
For the Second and Third Defendants		Clyde & Co

HIS HONOUR:

Introduction

1   In 2014 the Plaintiff ('McConnell Dowell') issued proceedings against the First Defendant ('Santam'), the Second Defendant QBE Underwriting Ltd as managing agent for QBE Syndicate 386 and QBE Syndicate 1886 ('QBE') and the Third Defendant Liberty Mutual Insurance Company ('Liberty'). The proceedings were issued in the Technology Engineering and Construction List.

2   This proceeding involves a large claim involving tens of millions if not hundreds of millions of dollars arising from the design and construction of a natural gas pipeline in Queensland.

3   The Construction Contract at the centre of this case and an associated arbitration have generated approximately 4 million electronic or PDF documents. These have been scanned and are available in a searchable format. McConnell Dowell has reduced the number of documents said to be relevant to the present proceeding from approximately 4 million to approximately 1,400,000. The use of predictive coding technology (TAR or Technology Assisted Review) is likely to dramatically reduce this volume and contain it further within reasonable and manageable bounds.

4   The central issue for the Court is this: given the enormous number of documents generated by the proceeding, how discovery should be managed, consistently with the principles of proportionality and s 9 of the *Civil Procedure Act 2010* (Vic.) (the 'CPA') which requires the Court to further the Overarching Purpose of the CPA when making orders or giving directions. Section 7 of the CPA provides as to the Overarching Purpose:

> **Overarching purpose**
>
> (1)   The overarching purpose of this Act and the rules of court in relation to civil proceedings is to facilitate the just, efficient, timely and cost-effective resolution of the real issues in dispute.
>
> (2)   Without limiting how the overarching purpose is achieved, it may be achieved by—
>
>   (a)   the determination of the proceeding by the court;

(b)   agreement between the parties;

(c)   any appropriate dispute resolution process—

(i)   agreed to by the parties; or

(ii)   ordered by the court.

5   The very large number of documents involved in the proceeding calls for special management. The cost of traditional discovery processes in a case such as this dictates that they are not appropriate. The prospect of McConnell Dowell's solicitors conducting the manual review of 4 million documents for relevance in a cost effective manner is unrealistic. Even with a reduction to 1,400,000 documents, say following use of de-duplication technology,[1] for example, a junior solicitor taking one minute to review and catalogue each of the 1,400,000 documents for relevance in an initial review exercise would take over 23,000 hours, equating to over 583 working weeks. This time does not include the time or cost of a senior solicitor to conduct a review of the process to ensure that only relevant documents are discovered. Following the manual review and provision of the reviewed documents by the providing party to the other parties, relevant documents would then be required to be inspected, reviewed and assessed by the other parties, giving rise to further rounds of delay, and potentially massive cost and expense.

6   This process puts all parties at risk of bearing these costs in the event of losing the case and suffering an adverse costs order. Thus, employing a traditional manual discovery process can work to place the cost-benefit of conducting litigation in a large document case at serious risk.

---

[1]   De-duplication technology is a recognised method of using technology to eliminate duplicate documents from an initial discovery pool of documents. The parties may be directed to use data de-duplication of emails and other documents by use of such methods as hash based data de-duplication using a hashing algorithm such as Secure Hash Algorithm 1 or 256 (SHA-1 or SHA -256) or Message-Digest Algorithm 5 (MD5) or by use of such other suitable technology to identify and, where appropriate, remove duplicates from their exchanged document collections (together called "De-duplication Technology"). This may be based on the approach agreed during a discovery conference. De-Duplication Technology was used, for example by the High Court of Justice (UK) in *Pyrro Investments Ltd v MWB Business Exchange and Ors* [2016] EWHC 256 (Ch) where at [5] it is noted, an initial 17.6 million potentially discoverable documents were reduced to 3.1 million by de-duplication technology.

7   In this case, traditional manual discovery of the Plaintiff's documents is not likely to be either cost effective or proportionate.

8   For these reasons, on 8 September 2016, in accordance with s 7(2)(c)(ii) of the CPA, I made an Order for the appointment of a Special Referee pursuant to Order 50 of the *Supreme Court (General Civil Procedure) Rules 2015* (Vic) (the 'Rules') to conduct a reference for the purposes of answering questions as to the appropriate management of discovery in the proceeding and to deliver a report to the Court on those questions.

9   Mr Anthony Nolan QC was appointed as the Special Referee.

10   Prior to this step being undertaken, the parties focussed their efforts on carrying out 'traditional' discovery processes. The parties were unable to agree upon an alternative procedure to achieve the goals of the Overarching Purpose prescribed by the CPA.

11   The reference was conducted with the cooperation of the parties as a facilitation process rather than an adversarial process.

12   A positive outcome was that the parties were able to agree on various protocols and procedures for discovery facilitated by the Special Referee.

13   The reference has served to bring to the attention of the parties and their legal advisors the use of predictive coding technology (TAR or Technology Assisted Review) and provide an opportunity for them to gain an understanding of and assess the process. It has also served to provide an opportunity for the parties to undertake a 'due diligence' exercise to estimate the cost benefit of using the technology. This has assisted the parties to make the commercial decision to invest in the process.

14   Implementation of these protocols and procedures will, more than likely, be an ongoing process which may call for further case management. By way of examples it is likely that implementation of the predictive coding protocol ('TAR' or 'Technology

Assisted Review')[2] may take several months, and the determination of privilege issues, unless resolved by agreement, may require further management.

Background

15   In or about November 2009, McConnell Dowell entered into a joint venture with Consolidated Contracting Company Australia Pty Ltd. A construction contract was then entered into by the joint venture with QCLNG Pipeline Pty Ltd to design and build a natural gas pipeline in Queensland for $730 million ('Construction Contract').

16   McConnell Dowell took out various policies of insurance with Santam, QBE and Liberty. Each has denied liability under these policies. The grounds for the denial of the liability would appear to be a central issue in this proceeding.

17   Difficulties arose in the design and construction of the pipeline and in particular the welding of joints. This resulted in an arbitration between the parties to the Construction Contract conducted in 2012.

Predictive coding (or TAR)

18   The use of computer technology to assist parties manage discovery has been recognised and endorsed in other jurisdictions.

19   *Pyrrho Investments Limited v MWB Property Limited*[3] was a case decided in the High Court of the United Kingdom. Initially the case involved some 17.6 million documents. This number was reduced to approximately 3.1 million by the use of de-duplication technology.[4] The English Civil Procedure Rules require a party to make 'a reasonable search for documents' which must comply with the obligation for the discovery to be 'proportionate'.[5] Master Matthews described the process of predictive coding[6] and endorsed its use. He described the cost involved as 'proportionate'.[7]

---

[2]   Predictive coding ('TAR' or 'Technology Assisted Review') is described briefly below.
[3]   [2016] EWHC 256 (Ch) ('*Pyrrho*').
[4]   Ibid [5].
[5]   Ibid [8].
[6]   Ibid [17]-[24].

20   The TAR process is far more sophisticated than a word search facility. In effect, the software enables a computer to be 'trained' to recognise concepts in the electronic documents fed into the system which are relevant to the issues in the proceeding. Master Mathews in *Pyrrho* described the TAR process as follows:[8]

> … the predictive coding process runs more or less like this. First of all, the parties will settle a predictive coding protocol, setting out the process in more detail, including definition of the data set, sample size, batches, control set, reviewers, confidence level and margin of error. Then criteria (perhaps agreed, perhaps unilateral) must be decided upon for inclusion of documents in the process. Those criteria will include who had the documents ("custodians") and the date range, but perhaps also whether the documents contained any of the keywords chosen. Certain types of documents, not having any or any sufficient text, will be excluded (they will have to be considered manually). The resulting documents are 'cleaned up', by removing repeated content (*eg* email headers or disclaimers) and words that will not be indexed (*eg* because not useful in assessing relevance).

> Then a representative sample of the 'included' documents is used to 'train' the software. In the present case, Mr Spencer suggests that it will comprise 1600-1800 documents (a size set by the size and variety of the entire document set). A person who would otherwise be making the decisions as to relevance for the whole document set (*ie* a lawyer involved in the litigation) considers and makes a decision for each of the documents in the sample, and each such document is categorised accordingly. It is essential that the criteria for relevance be consistently applied at this stage. So the best practice would be for a single, senior lawyer who has mastered the issues in the case to consider the whole sample. Where documents would for some reason not be good examples, they should be deselected so that the software does not use them to learn from. The software analyses all of the documents for common concepts and language used. Based on the training that the software has received, it then reviews and categorises each individual document in the whole document set as either relevant or not.

> The results of this categorisation exercise are then validated through a number of quality assurance exercises. These are based on statistical sampling. The sampling size will be fixed in advance depending on what confidence level and what margin of error are desired. The higher the level of confidence, and the lower the margin of error, the greater the sample must be, the longer it will take and the more it will cost.

> …

> The samples selected are (blind) reviewed by a human for relevance. The software creates a report of software decisions overturned by humans. The overturns are themselves reviewed by a senior reviewer. Where the human decision is adjudged correct, it is fed back into the system for further learning.

---

7   Ibid [33]-[34].
8   Ibid [19]-[24].

(It analyses the correctly overturned documents just as the originals were analysed.) Where not correct, the document is removed from the overturns. Where the relevance of the original document was incorrectly assessed at the first stage, that is changed and all the documents depending on it will have to be re-assessed.

The process of sampling is repeated as many times as required to bring the overturns to a level within agreed tolerances, and so as to achieve a stability pattern. This is usually not less than 3, making 4 rounds in total. …The trend of overturns should be lower from round to round. Ultimately there will be a final overturn report within the agreed tolerance, so that the expense of further rounds of review will not be justified by the reduced chance of finding further errors, and the list of relevant documents can be produced.

Although the number of documents that have to be manually reviewed in a predictive coding process may be high in absolute numbers, it will be only a small proportion of the total that need to be reviewed in the present case. Thus - whatever the cost per document of *manual* review - provided that the exercise is large enough to absorb the up-front costs of engaging a suitable technology partner, the costs overall of a predictive coding review should be considerably lower. It will be seen that, because the software has to be trained for every case, each use of the predictive coding process is bespoke for that case.

21   As to accuracy of the TAR process, reference is made to *Irish Bank Resolution Corporation Ltd & Ors v Quinn & Ors*.[9]

22   In *Irish Bank,* the High Court of Ireland dealt with another case involving large scale discovery. The initial keyword search resulted in 1.7 million documents of potential relevance and after de-duplication that number reduced to 680,809 documents.[10] The plaintiff sought orders from the court that the defendants make discovery with the assistance of predictive coding. Fullam J described the process in detail.[11] His Honour set out the requirements of discovery under the Irish Rules and stated:[12]

The making of an order of discovery is premised on the documents being relevant and necessary for the fair disposal of the cause or matter or for saving costs. While there is no specific reference in rule 12 to the concept of proportionality, the courts increasingly refer to it as a relevant factor in assessing whether the necessity requirement has been satisfied on the facts of a particular case.

---

[9]   [2015] IEHC 175 ('*Irish Bank*').
[10]   Ibid [12].
[11]   Ibid [19-30].
[12]   Ibid [48].

23   Fullam J agreed that the process of technology assisted review was appropriate and reliable, observing:

> 66.   The evidence establishes, that in discovery of large data sets, technology assisted review using predictive coding is at least as accurate as, and, probably more accurate than, the manual or linear method in identifying relevant documents. Furthermore, the plaintiff's expert, Mr. Crowley exhibits a number of studies which have examined the effectiveness of a purely manual review of documents compared to using TAR and predictive coding. One such study, by Grossman and Cormack, highlighted that manual review results in less relevant documents being identified. The level of recall in this study was found to range between 20% and 83%. A further study, as part of the 2009 Text Retrieval Conference, found the average recall and precision to be 59.3% and 31.7% respectively using manual review, compared to 76.7% and 84.7% when using TAR. What is clear, and accepted by Mr. Crowley, is that no method of identification is guaranteed to return all relevant documents.

> 67.   If one were to assume that TAR will only be equally as effective, but no more effective, than a manual review ,the fact remains that using TAR will still allow for a more expeditious and economical discovery process …

> 69.   Pursuant to the legal authorities which I have cited supra, and with particular reference to the albeit limited Irish jurisprudence on the topic, I am satisfied that, provided the process has sufficient transparency, Technology Assisted Review using predictive coding discharges a party's discovery obligations under Order 31, rule.12.

24   In another UK case, *David Brown v BCA Trading*,[13] Registrar Jones adopted and applied the decision of Master Matthews and acknowledged that the protocols concerning technology assisted review used in the Technology and Construction Court were appropriate.[14] Registrar Jones ordered that discovery in the proceeding could proceed by way of predictive coding.[15]

25   In *Rio Tinto v Vale*,[16] Judge Peck of the United States District Court (Southern District of New York) discussed the use of Computer Assisted Review in American Courts and highlighted that it was now accepted as *'black letter law'* in the United States.[17]

---

[13]   [2016] EWHC 1464 (Ch).
[14]   Ibid [8].
[15]   Ibid [11].
[16]   14 Civ. 3042 (RMP)(AJP) (2 March 2015).
[17]   *Rio Tinto v* Vale 14 Civ 3042 (RMP(AJP) (2 March 2015) 2.

26   In Australia, there are at least three major providers of litigation support that now offer predictive coding software and supporting services to assist parties in the discovery process.

27   In the TEC List of the Supreme Court of Victoria, on 1 September 2016 a Standard Operating Procedure (TEC SOP 5 [TAR]) was issued as an interim measure[18] to provide guidance to parties and legal representatives for the review of documents by using TAR (Technology Assisted Review or Predictive Coding).[19]

28   This development was in response to the unique challenges presented to the TEC List in the management of construction and engineering cases of the scale of the present. Large document production is particularly prevalent in the field of construction and engineering projects, where technology is used to generate and disseminate project information to a large number of participants. With the commonplace use of ever more powerful computers in large construction and engineering projects, collection and storage of electronic data is growing at an exponential rate. The end result is that an enormous volume of information is now created, exchanged and stored electronically, as 'electronically stored information' or ESI. This phenomenon calls for special management in the event of litigation or arbitration in a construction or engineering case.

29   However, the pressures caused to the litigation process described are not confined to construction and engineering cases.

30   TEC SOP 5 [TAR] has now been replicated in a new Practice Note of the Court projected to be issued on 1 January 2017, known as the 'Technology in Civil Litigation Practice Note SC Gen 5'.[20] This Practice Note is designed for general application

---

[18]   TEC SOP 5 [TAR] is an interim measure and will apply for use in the TEC List pending inclusion of a TAR procedure in a Technology in a Civil Litigation Practice Note to be issued.

[19]   Technology Assisted Review (TAR or Predictive Coding or Computer Assisted Review) are described by the High Court of Justice (UK) in *Pyrro Investments Ltd v MWB Business Exchange and Ors* [2016] EWHC 256 (Ch) at [17]-[24], with findings as to the potential effectiveness and utility of TAR described at [31].

[20]   Technology in Civil Litigation Practice Note SC Gen 5, paragraphs 8.7–8.9.

within the Commercial Court and the Court generally. The new Practice Note, when it commences, will provide in relation to TAR:

> 8.7   In larger cases, ordinarily technology assisted review will ordinarily be an accepted method of conducting a reasonable search in accordance with the Rules of Court. It will often be an effective method of conducting discovery where there are a large number of electronic documents to be searched and the costs of manually searching the documents may not be reasonable and proportionate. In such cases, the Court may order discovery by technology assisted review, whether or not it is consented to by the parties.
>
> 8.8   As part of cooperating in the conduct of the proceeding, and to avoid later disputation, parties may be directed by the Court to confer with a view to agreeing on a protocol for technology assisted review as part of the discovery process and inclusion in a discovery plan.
>
> 8.9   The protocol may include:
>
> - The appointment of a Joint Operator; or
>
> - The appointment by each party of a Party Operator;
>
> - A general description of the system to be used, either by a Joint Operator or a Party Operator, in undertaking the technology assisted review process, including, but not limited to:
>
> - A continuous active learning protocol (using a constantly changing body of documents which are used to train the technology assisted review algorithm);
>
> - A simple active learning protocol (using statistical samples, including control sets, or random samples and the like);
>
> - A simple passive learning protocol (using other recognised statistical methods); and
>
> - Any other appropriate system.
>
> - A general description of the method to be used, including where relevant:
>
> - An outline of the steps to be undertaken as part of the protocol;
>
> - Where statistical measures are adopted for quality assurance purposes, details of the statistical measures;
>
> - Proposed members of the review team;
>
> - The management of non-text based documents;
>
> - The treatment of foreign language documents;
>
> - Any procedures proposed for high level culling and the elimination of repeated or duplicate content;

- The method for determining the scale of relevance;

- Any necessary manual review of the results produced by the system;

- Any document or groups of documents a party proposes to exclude from the process; and

- Any other matters relevant to the method adopted.

- Arrangements for the clawback of privileged or confidential information which may have been inadvertently exchanged or disclosed as part of the technology assisted review process; and

- Provision for the exchange of relevant documents that have not been disclosed as part of the technology assisted review process.

31   The use of technology assisted review (TAR) is not confined to cases where all parties to the litigation directly participate in the process. A single party may employ the technology in conducting a review of its own documents. For example, on 7 November 2016, an order was made in the Federal Court of Australia, in *Money Max Pty Ltd v QBE Insurance Group Ltd*,[21] ordering the Respondent, which had used technology assisted review (TAR) for the purposes of giving discovery to the Applicant, to provide a report to the Applicant describing the manner in which the Respondent had applied the technology for the purposes of giving discovery to the Applicant and the results of the application of TAR.

Opinion of the Special Referee

32   An important question put to the Special Referee for consideration in the reference was:

> Having regard to the powers of the Court under section 55 of the *Civil Procedure Act 2010* (Vic), what process or processes should be adopted in relation to:
>
> (a)   discovery by the parties of:
>
> (i)   documents in original electronic form?
>
> (ii)   PDF documents?
>
> (iii)   documents in hard copy form?

---

[21]   Federal Court of Australia (Murphy J) VID 513/2015.

(b) the inspection of documents.

33 The Special Referee reports in this regard that:

- the parties have agreed to procedures and protocols in relation to discovery of all documents;

- it is the opinion of the Special Referee that the Court should adopt the procedures and protocols agreed to by the parties. The procedures and protocols are consistent with the obligations of all parties under the CPA and the Rules;

- all parties have agreed that the use of predictive coding is appropriate for all of these documents and have cooperated in establishing a protocol for this to occur;

- in relation to the discovery which has been and will be made by the Plaintiff, all documents have been scanned and are in a searchable electronic form. The Special Referee has not been advised about any dispute concerning the authenticity of any document. Therefore discovery of relevant documents can be undertaken using the electronic form without the need to produce the original.

- all and any outstanding questions of legal professional privilege are best reserved for determination after the predictive coding process is complete and a new list of documents in respect of which privilege is claimed has been generated.

34 Pursuant to r 50.04 of the Rules, I respectfully adopt these opinions of the Special Referee expressed in his report.

Orders

35 The Court will hear the parties on what orders should be made to reflect the opinions of the Special Referee adopted by the Court.

36   In particular, I will make orders for the use of predictive coding (TAR) in the discovery process, including orders as to the appropriate protocol for this to occur.

37   I will also make orders facilitating payment of the remuneration of the Special Referee.