

THE HIGH COURT

[2011/5843P]

BETWEEN

IRISH BANK RESOLUTION CORPORATION LIMITED & ORS

PLAINTIFFS

AND

SEAN QUINN & ORS

DEFENDANTS

JUDGMENT of Mr. Justice Fullam delivered the 3rd day of March, 2015

1. This application concerns a novel but inevitable discovery issue in this jurisdiction involving electronic documentary evidence.

2. In essence the plaintiffs are seeking the court's approval for the use of a process known as Technology Assisted Review ("TAR") which combines predictive coding (a technology which produces a relevance score for documents using algorithms) with human expertise.

3. The plaintiffs claim that TAR will save time and be more cost effective compared to the traditional manual/linear method of discovery.

Background**Proceedings**

4. IBRC Limited is the successor to the former Anglo Irish Bank which is in liquidation. The defendants comprise of two groups, the personal defendants and the Senat defendants.

5. In the proceedings the plaintiffs claim that the defendants conspired to wrongfully convert assets to the value of €455 million, the property of the former Anglo Irish Bank.

Discovery

6. On the 19th March, 2014, the commercial court (Kelly J.) ordered *inter partes* discovery but deferred setting a discovery timetable until the plaintiffs conducted a scoping exercise on the retrieval of potentially relevant electronic documentation and reported back to the court.

7. On 18th June, the plaintiffs sought the defendants' consent for the use of TAR in making their discovery as set out in the affidavit of their expert Dr. Damir Khadvezic.

8. On 23rd June, the plaintiffs informed the court that they wished to proceed with the predictive coding approach to discovery having regard to the very large volume of documents, the resources required and the costs involved. The application was grounded on the affidavit of the plaintiff's solicitor Ms. Karyn Harty. The court adjourned the matter to allow the plaintiffs provide a clearer estimate of the time required. That estimate depended on whether the defendants consented to the carrying out of predictive coding discovery.

9. Initially the defendants' legal representatives and the sixth defendant, Ms. Aoife Quinn, engaged with the process. A joint meeting was held on 7th July at which the plaintiffs' expert attended and provided explanations in relation to the search terms used, issues of transparency, cut off, and the possibility of documents being overlooked. The meeting was followed by letter dated 9th July from Ms. Harty addressing all the issues raised at the meeting and enclosing the search terms (tailored for each specific data-set) on a confidential basis.

10. Ultimately, the defendants' position was one of opposition, basically on the ground that TAR did not comply with O.31 r.12 of the RSC.

11. On the 9th September the court (Kelly J.) directed a hearing to determine the following issues:

(1) Whether technology assisted review and predictive coding complied with the rules of the Superior Courts for discharging a party's discovery obligation; and

(2) If so, to determine the most appropriate methodologies to be used.

Evidence

12. The plaintiff's initial **scoping exercise** involving a key word search yielded 1.7 million potentially relevant documents. By September, following removal of duplicates (deduplication) and documents in other languages, that number had reduced to 680,809 documents suitable for predictive coding. Mr Crowley would expect that less than 10% of the 680,809 documents would need to be reviewed if predictive coding is employed. Ms. Harty estimated that a traditional linear review, using a team of 10 experienced reviewers, would take 9 months at a cost of €2 m leaving supervision and technology costs aside, whereas, the use of predictive coding would enable the plaintiffs to make discovery within a much shorter timeframe and at substantially lower cost. Mr Crowley suggests "at a fraction of the cost".

Dr. Mee, the defendants' expert suggests that a linear review using 10 reviewers could be completed in 113 days at a cost of €220,000.

13. Counsel advised the court that the search terms provided by Ms. Harty were wider than they appeared, as they incorporated a "Boolean search" flexibility which captured documents containing not only the specific words of the search terms but also any words connected to them.

14. The defendant's solicitors inquired about having a written protocol along the lines proposed in the US case of *Da Silva Moore* (11 CIV.1279(ALC)(AJP) which was the first instance of a US federal court accepting the use of predictive coding in discovery. The defendants suggested that the key-word search without the defendants' input left the TAR process "vulnerable to any oversight". Ms Harty indicated that while the defendants' request for input into the keyword search was "very late", she would be prepared to have a dialogue on it and that she had no objection in principle to a written protocol but said that the *Da Silva Moore* protocol was inappropriate for this jurisdiction.

15. On 30th July, Ms. Harty provided a draft "Protocol for TAR" for discussion to Messrs Arthur McLean, the personal defendant's solicitors, explaining that once the Training Set (T1) and the Control Set (C1) were completed the next training phase would not commence until either agreement had been reached between the parties or the Court approved a protocol for the use of TAR and predictive coding.

16. The court had the following affidavit evidence;

- (1) Ms. Karyn Harty on 19th June, which referred to the expert evidence of Dr. Khadvezic in his affidavit of 18th June.
- (2) Ms. Aoife Quinn on behalf of the personal defendants dated 1st August.
- (3) Ms. Vivien Mee, an IT forensic expert, on behalf of the Senat defendants dated 12th August.
- (4) Ms. Harty's replying affidavit of 2nd September which referred to the affidavits of two experts of the same date, namely a supplemental affidavit from Dr. Khadvezic and an affidavit from Mr. Conor Crowley, a US attorney specialising in electronic discovery.

General Description TAR Methodology Using Predictive Coding

17. (1) Technology Assisted Review is a methodology for identifying relevant electronic documents using a combination of technology based on predictive coding, and expert human input. The TAR methodology narrows the universe of a party's electronic documents to discoverable relevant documents in two stages.

- 1 the universe of data is narrowed to data that is potentially responsive to a discovery request relevant ; and
- 2 the potentially responsive data is narrowed down to what is in fact responsive by means of manual review.

18. In his judgment in *Da Silva Moore* delivered on February 24th, 2012, Judge Peck explained the methodology as follows at page3:

By computer-assisted coding, I mean tools (different vendors use different names) that use sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer.

Unlike manual review, where the review is done by the most junior staff, computer -assisted coding involves a senior partner (or small team) who review and code a "seed set" of documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer's coding. (Or the computer codes some documents and asks the senior reviewer for feedback.)

When the system's predictions and the reviewer's coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents. Typically the senior lawyer (or team) needs to review only a few thousand documents to train the computer.

Some systems produce a simple yes/no as to relevance, while others give a relevance score (say, 0 to 100 basis) that counsel can use to prioritise review. For example, a score above 50 may produce 97% of the relevant documents but constitutes only 20% of the entire document set.

Counsel may decide , after sampling and quality control tests, that documents with a score of below 50 are so highly likely to be irrelevant that no further human review is necessary. Counsel can also decide the cost-benefit of manual review of the documents with scores of 15-50.

Judge Peck pointed out "that every person who uses email uses predictive coding, even if they do not realise it. The "spam filter" is an example of predictive coding,"

Predictive coding is based on a modelling framework called Predictive Analytics which encompasses a variety of techniques from statistics, data mining, and game theory that analyse current and historical facts to make predictions about future events.

19. The plaintiffs' expert evidence was that TAR using predictive coding is accurate and cost effective for data sets of 500,000 documents but is most cost effective for data sets in excess of 1 million.

20. The exercise involves a number of stages;

- (1) The preliminary identification of potentially relevant documents by the application of search words to all electronic documents held by a party.
- (2) Training the computer

The expert group selects a first seed-set/training-set of documents, usually 25, from the data set. The expert interacts with the

system by asking a yes/no question of the document against a series of controlled samples. (The plaintiff's proposed question in this case is: *Is this document relevant to any discovery category?*) The documents in the training set will include privileged documents. The system builds a knowledge model as it learns from the expert and presents further samples for review.

21. Normally around 25-50 iterations (repetitions) are sufficient to build the model to the point where it can predict what the expert will choose as responsive in the sample being reviewed.

22. Approximately 1,000 documents will be used in the training sets in this case to get to the stable point.

23. A further "control-set" is used to test the prediction model.

24. Once the system is found to accurately predict relevance over a series of consecutive samples the model is considered statistically stable and can be applied to the rest of the collection.

25. (3) The model is then applied to the data set and grades the electronic documents into bands of potential relevance. These bands are at 10 percentage point intervals.

26. (4) Cut Off/ Threshold

The model identifies a cut-off point above which it predicts these bands will contain truly relevant documents. These are then manually reviewed and any false positives are eliminated.

27. (5) Remainder

The remaining documents below the cut-off threshold are predicted to contain mostly non-relevant documents. However, a percentage of the remainder documents are also reviewed to ascertain the presence of relevant documents (false negatives). That percentage is determined by the model using random sampling.

This exercise will yield some relevant documents which will be added to the truly relevant documents above the cut off.

28. (6) Effectiveness

The effectiveness of the model is judged on the basis of three tests which allow the model to be transparent, verifiable and tracked throughout the process.

(i) Recall (completeness) measures the percentage of responsive documents that have been identified.

(ii) Precision (exactness) measures the percentage of truly responsive relevant documents within the identified set.

A high recall rate guarantees that the review finds most of the relevant documents while a high precision rate guarantees that the documents found do not contain many irrelevant documents.

(iii) The f - Measure

The f-measure is the harmonic mean of recall and precision and is a measure of the effectiveness of the information retrieval search. It strikes a balance between the two and is used as a target to achieve a high level of recall while at the same time minimising the amount of manual review required. In order to achieve a high f1 score, a search effort must achieve both high recall and high precision. The harmonic mean, unlike the more common arithmetic mean (average), falls closer to the lower of the two quantities. As a summary measure a harmonic mean may be preferable to an arithmetic mean because a high harmonic mean depends on both high recall and high precision, whereas a high arithmetic mean can be achieved with high recall at the expense of low precision, or high precision at the expense of low recall.

29. There is no standard for recall, precision or f-measure rates. Dr. Khadvezic's view based on the recommendations of Symantec (the providers of the Clearwell system) was that the minimum f-measure should be 80% to guarantee a high amount of truly responsive (relevant) documents while giving a good balance to cost of manual review.

30. Dr. Khadvezic cited a study by Maura R Grossman and Gordon V Cormack entitled "*Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*" (Richmond Journal of Law and Technology). As the title indicates the study concluded that technology assisted processes can yield superior results as measured by recall, precision as well as the f-measure. The Grossman/Cormack study is cited in many of the US cases dealing with the issue.

The Plaintiffs Proposed Protocol

31. The proposed protocol requires court approval (section.11). The plaintiff proposes a protocol of ten stages.

32. In the Preliminaries Section (section 1) the protocol identifies how it will deal with the transparency issue arising from the process.

(a) At 1.2 the plaintiffs undertake to notify and consult with the defendants and will seek to reach agreement with them in relation to any necessary derogation from the protocol. In the event that agreement is not possible there is generally liberty to apply to court.

(b) The defendants will nominate a Barrister ("the defendant's nominated counsel") for the purposes of the disclosure procedure set out in section 6.

As there are two sets of defendants in this case the protocol will need to be amended in this regard.

33. A keyword scoping exercise (section 2) (resulting in the identification of 680,809 potentially relevant e-documents) has been completed and the first stage of training the computer (Section 7) has commenced.

34. The defendants have been furnished with the key words used in the collection of potential discovery material (section 3).

35. The elements of TAR are set out in section 4.

36. The details of how predictive coding, (section 5), will operate have been explained to the defendants.

37. Section 6 proposes a disclosure procedure for documents used in training the computer. All documents used in training the computer will be assigned an electronic folder (the PC Folder). These documents will have been coded either relevant or not-relevant. These subsets will include privileged and confidential documents.

38. Relevant confidential documents which are not privileged will be discovered in the normal way. Relevant privileged documents will be ultimately included in the Second Part of the First Schedule of Ms. Harty's affidavit of discovery.

39. The "not- relevant" subset will include documents which are not relevant (simpliciter), privileged, confidential and commercially sensitive. The plaintiff will list relevant, not-relevant and confidential not-relevant documents in a PC Schedule. The balance, i.e. the privileged documents and the commercially sensitive documents are described as the "X" documents. Under the protocol, the X documents will be excluded from disclosure procedure.

40. The purpose of the section 6 procedure is to provide a mechanism for the defendants to challenge the documents coded as not relevant by the plaintiffs while at the same time protecting the commercially sensitive not relevant documents from disclosure. As documents coded not-relevant will include confidential documents, inspection will be restricted to the defendants' nominated Barristers. This avoids putting the defendants' solicitors in a position of conflict.

41. Essentially, the proposed mechanism involves notifying the defendants of the total number of "X" documents in the PC folder, and second, the defendants nominated counsel being given a copy of the PC schedule on his/her undertaking not to make copies of it or to transmit it to the defendants.

42. The defendants' nominated counsel have seven days to challenge the plaintiffs' coding of any document in the PC Schedule as "not relevant" and if there continues to be disagreement after the seven days as to the status of the document, the protocol provides for application to the court.

43. After the completion of the disclosure process, the predictive coding process will continue with the training phase as set out in section 7. The plaintiffs' expert will advise when the system has reached the appropriate stabilisation point following which the system will rank the documents in the data set for likelihood of relevance. The plaintiff's expert will identify the threshold following which, the plaintiff's solicitor will advise the defendants of same. If the defendants wish to dispute the threshold they may apply to court for a determination of an appropriate threshold.

44. The plaintiffs will then conduct a linear review of the Review Set for relevance by category and for privilege.

45. At section 9, the plaintiffs undertake to employ all reasonable quality control measures for the purpose of testing the coding of documents falling below the threshold. The quality control checks include Discrepancy Analysis and a Remainder Test using random sampling.

46. Section 10 provides that following the linear review, the plaintiffs will make discovery on oath in accordance with O. 31, r. 12 of the RSC 1986 (as amended) including listing documents by category, save that to the extent that any specific electronic document falls within more than one category it will only be listed once.

At 10.2 the Protocol states:

"It will remain the obligation of the Plaintiffs to assess the relevance of documents to the categories of discovery and to satisfy themselves that all reasonable steps have been taken to identify relevant documents and to make discovery in accordance with O. 31, r. 12 of the RSC".

Paragraph 10.3 of the Protocol states:

"The plaintiffs will, when making discovery, produce an expert certificate confirming that the TAR was statistically valid and providing a detailed basis for drawing that conclusion."

Objections of Defendants.

47. 1. **TAR will not capture all relevant documents and therefore is not compatible with the obligations of a party making discovery, which is the objective target of 100 per cent of relevant documents.**

Mr. Crowley, says that predictive coding is the most efficient and effective method of identifying relevant information in that jurisdiction.

2. **TAR is not suitable for data sets of less than 1 million documents.**

Mr. Crowley says that TAR is wholly appropriate and suitable for data sets of less than 100,000 documents. Dr. Khadvezic says that the accuracy of the review is not compromised by using TAR for a smaller data set than 1 million documents.

3. **The Court has not been told what the f-measure is going to be.**

Dr. Khadvezic's view is that the minimum f-measure should be 80%.

4. **The Training Sets are not specific to the categories of discovery, and the sets might not contain any relevant documents.**

The plaintiffs propose a broad question namely "Is this document relevant to any discovery category?" The system will learn what is and what is not relevant. Dr. Khadvezic says the defendants' objection is unfounded because training the computer is a supervised approach. The first training set is created judgmentally, but subsequent training sets involve selecting documents which are most difficult to predict rather than being chosen at random. Furthermore, Mr. Crowley states that attempting to intentionally teach the computer wrongly will not work. He cites a study - *The Impact of Incorrect Training Sets and Rolling Collections on Technology Assisted-Review* which concluded that "the impact of wrong training documents was smaller than expected; inserting up to 25% wrong documents resulted in only 3-5% less classification quality". Furthermore, the control set is created by the system and has a 95% confidence measure.

5. The concept of a cut- off point for responsive documents will lead to the omission of relevant documents, and the quality control checks for retrieving relevant documents below the cut-off are not explained.

Dr. Khadvezic says that predictive coding technology is only one aspect of technology assisted review which involves "human input at every level and considerable quality control checks". The cut- off point is not arbitrary but is determined by the system to "offer a best balance between precision and recall". Furthermore, documents below the cut- off will not be ignored entirely as "rigorous quality controls and sampling will be conducted to ensure that the system has not incorrectly missed relevant documents."

He says that "all review methods can result in the omission of relevant documents." Predictive coding results in the omission of fewer relevant documents than other methods and the results achieved can be quantified, which is not the case with manual review, in particular where the review team is comprised of junior reviewers.

Mr. Crowley says that the defendant's objection is based on a false premise of the process not capturing all relevant documents. Predictive coding should identify more relevant documents than manual review.

6. There will be no savings in cost and time.

Dr. Mee's objection was based on her view that ten reviewers could complete a linear review within 113 days at a cost of €220,000. She further states that using predictive coding there would still have to be a manual review of in excess of 20% of the data set.

Ms. Harty and Dr. Khadvezic say that Dr. Mee's estimate of 113 days is based on each reviewer checking 600 documents per day and in their experience, that is wholly unrealistic. Ms. Harty suggests that using Dr. Mee's figures "the reviewers would spend no more than 60 seconds on each document even if they worked a 10 hour day. Some of the documents are of considerable length and in her view this could not amount to a quality of review adequate to comply with the obligations involved in discovery of this nature and she would not be satisfied professionally to stand over a review conducted on that basis".

Mr. Crowley says that the TAR process would involve a linear review of less than 10% of the data set which could be concluded at a fraction of the cost of manual review.

The defendants rely on *Progressive Casualty Insurance Company v Delaney* 2014 WL 2112927(D.Nev.) The court rejected the plaintiff's motion seeking leave for the use of predictive coding on facts, which the defendants say, are similar to the present case. The plaintiffs agreed to a linear protocol, began manually reviewing 565,000 documents, failed to comply with deadlines to produce documents, gave undertakings to the court in respect of outstanding obligations, consulted Mr Crowley and determined that utilising predictive technology would be more effective and efficient. They selected the Equivio Relevance Program and began utilising predictive coding techniques without the defendant's agreement to amend the existing protocol and without seeking leave of the court to amend the ESI order.

The Law

48. Order 31 Rule 12 General Requirements

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.

49. In *Framus Ltd v. CRH Plc* [2004] 2 IR 20, Murray J. said:-

"Thus it follows that in making an order for discovery, the court must bear in mind not merely the relevance to the issues in the proceedings and necessity for the fair disposition of the case or cost saving but must also maintain a sense of proportion between that which is asked for and that which is required".

In *Bula (In Receivership) v. Crowley*, Murphy J. in an extempore judgment in the High Court said:

"Discovery is a procedure which is left to the integrity of the parties themselves. The party who fails to make an adequate discovery is precluded from relying upon that document. The deponent who swears the affidavit has the final word on what is relevant and it is difficult if not impossible for the court to go behind that."

Absence of specific rule for use of technology assisted discovery

50. While O. 31, r. 12 was amended in 2009 to include "all electronically stored information" in the definition of documents, there is no rule providing for the adoption by a party of technology assisted review using predictive coding. Neither is there a specific rule requiring linear manual review.

51. Since the *Da Silva Moore* decision in April 2012, legal practitioners and courts in the US and elsewhere have recognised the value of technology assisted review in appropriate cases. Rules of court take their cue from the Sedona Conference Principles and the Sedona Conference Cooperation Proclamation.

52. In the foreword to Good Practice Guide to Electronic Discovery in Ireland published in April 2013, Mr. Justice Clarke of the Supreme Court, said:

"In at least some jurisdictions, significant changes in the Rules of Court and other procedural laws have been adopted in an attempt to control the disclosure process and to prevent the very real risk that the costs associated with complying with disclosure obligations (which can in many cases reach 40% to 50% of the total costs of litigation) do not become a barrier on access to justice. In Canada, the Sedona Principles have been developed to provide best practice guidance in the field of discovery and disclosure. Those principles are, in my view, essential reading for anyone who has any interest in ensuring that discovery remains an important tool for establishing the truth while at the same time ensuring that the cost and complexity of discovery does not, itself become a barrier to the truth being established."

Obviously a basic requirement is that the parties should come together at the outset, agree a protocol and discovery and proceed on the basis of consent.

53. In this case the personal defendants have not consented, although, it is to be noted that in the main proceedings, in which they

are plaintiffs, they are consenting to TAR.

54. Therefore, in this case the court must also have regard to the following general principles.

Reasonable approach

55. As regards the extent of the duty of the party making discovery, Budd J. said in *Atlantic Shellfish Ltd. v. Cork County Council* [2006] IEHC 215, a judgment which was delivered prior to the amendment of Order 31:-

"An order for discovery under the Superior Court Rules carries with it the duty to search archives of records and files diligently for material documents including computer records... a party is required to make a reasonable search for documents falling within the scope of the order."

56. In *Thema International Fund Plc v. HSBC Institutional Trust Services (Ireland)* [2011] IEHC 496, Clarke J. said in the High Court at paragraph 2.10

"It is important to recall that the obligation on a party making discovery is to disclose, in so far as it may be reasonably possible, all documents which come within the categories agreed or directed by the court. However, the courts have always accepted that there is some risk, particularly in large discovery, that there will be an innocent failure to disclose documents which may be relevant. Clearly, where documents emerge which should have been, but were not disclosed, the court needs to assess the reason for the failure to disclose. It seems to me that where a party adopts a reasonable approach to the search of a large universe of documents by means of key words and the like, then it is unlikely that that party would suffer any adverse consequences if it were to transpire that, notwithstanding its best efforts, some document fell through the net. It should of course, be noted that the assumption in that last statement is that the party acted reasonably and used its best efforts. A clever use of key words which may raise a suspicion that same were deliberately designed to minimise the risk of damaging documents being selected might, of course, leave the court reaching entirely different conclusions."

At paragraph 2.11, Clarke J noted that in the case before him it was not possible for the parties to reach agreement on key words, but that agreement would

"give an added degree of comfort to parties so that any failure to throw up a relevant document by means of the use of agreed key words would be much more likely to be viewed as an unfortunate but unavoidable accident rather than a deliberate act."

At paragraph 2.12 he said:

"The whole point of narrowing the universe (of documents) by means of key word searches is to reduce the number of documents that require direct personal review. If the key words are too wide, then the selection process will not do that job. If the key words are too narrow (or, perhaps, deliberately or inappropriately skewed), then same is likely to enhance the risk of false negatives. Some reasonable balance has to be achieved between these two ends. Provided that a party acts bona fide, and that the approach to the use of search tools is along the lines which I have described, it does not seem to me that a party should face criticism or adverse consequences if it should transpire that, despite those best efforts some document slipped through the net. In addition, I should note, in that context, that if truly key documents were to slip through the net that might, of itself, lead to real questions as to whether anyone could reasonably have believed that the methodology was right in the first place."

57. Obviously in this case the process has advanced beyond the initial key word search stage and the issue is whether the plaintiffs may proceed with TAR incorporating predictive coding.

58. In this regard the words of Fennelly J. in *Ryanair Plc v. Aer Rianta CPT* [2003] 4 IR 264 are apt:

"The change made to O.31, r.12 in 1999 exemplifies, however, growing concerns of the dangers of costly and protracted litigation and, in particular, the burdens on parties and courts arising from excessive resort to automatic blanket discovery. The public interest in the proper administration of justice is not confined to the relentless search for perfect truth. The just and proper conduct of litigation also encompasses the objectives of expedition and economy."

Inherent Power of the Courts to Adapt in Absence of Specific Rule

59. It seems to me that the approach the court should take is that flagged by Mr. Justice Geoghegan in *Dome Telecom Ltd. V. Eircom Ltd* [2008] 2 IR 726 which again was a case prior to the amendment of Order 31. At paragraph 12 he stated:

"I would reject any idea that the right to discovery of documents should be exclusively based on an interpretation (literal or otherwise) of the relevant rule of court..."

...In modern times courts are not necessarily hidebound by interpretation of a particular rule of court. More general principles of ensuring fair procedures and efficient case management are frequently overriding considerations. The rules of court are important and adherence to them is important but if an obvious problem of fair procedures or efficient case management arises in proceedings, the court, if there is no rule in existence precisely covering the situation, has an inherent power to fashion its own procedure and even if there was a rule applicable, the court is not necessarily hide bound by it. It is common knowledge that a vast amount of stored information in the business world which formerly would have been in a documentary form in the traditional sense is now computerised. As a matter of fairness and common sense the courts must adapt themselves to this situation and fashion appropriate analogous orders of discovery."

Background to Da Silva Moore Decision

The *Da Silva Moore* decision was based on the Sedona Principles and Proclamation, various research papers, in particular two articles published in 2011, the first by Grossman and Cormack entitled *"Technology Assisted Review in EDiscovery Can Be More Effective and More Efficient than Exhaustive Manual Review"* and the second by Magistrate Judge Andrew Peck entitled *"Search Forward: Will manual document review and keyword searches be replaced by computer assisted review?"*

Judge Peck who was the judge in *Da Silva Moore* said at page 17 of his judgment:

The court recognises 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 is not a case of machine replacing humans: it is the process used and the interaction of man and machine that the court needs to examine.

The objective of review in e-discovery is to identify as many relevant documents as possible, while reviewing as few non-relevant documents as possible. Recall is the fraction of relevant documents identified during a review: precision is the fraction of identified documents that are relevant. Thus, recall is a measure of completeness, while precision is a measure of accuracy or correctness. The goal is for the review method to result in higher recall and higher precision than another method, at a cost proportionate to the value of the case.

At p 18 he said: *Moreover, while some lawyers still consider manual review to be the gold standard, that is a myth as statistics clearly show that computerized searches are at least as accurate, if not more so than manual review. The judge was referring to an empirical assessment by the Electronic Discovery Institute "to answer the question of whether there was a benefit to engaging in a traditional human review or whether computer systems could be relied on to produce comparable results and concluded that on every measure, the performance of the two computer systems was at least as accurate (measured against the original review) as that of human re-review".*

MSL agreed to "turn over" the seed set, excluding privileged documents, to the plaintiffs. Judge Peck said at page 23 - *"While not all experienced ESI counsel believe it necessary to be as transparent as MSL was willing to be, such transparency allows the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called "black box" of the technology.*

The decision in *Progressive* can be distinguished on the facts. Judge Leen did not reject the plaintiff's application because predictive coding didn't meet the requirements of the relevant discovery rule. On the contrary, she said at p8-

"Had the parties worked with their e-discovery consultants and agreed at the onset of this case to a predictive coding based ESI protocol, the court would not hesitate to approve a transparent mutually agreed upon ESI protocol".

The judge noted at para 10 that *"the cases which have approved technology assisted review of ESI have required an unprecedented degree of transparency and cooperation among counsel in the review and production of ESI responsive to discovery requests."*

The Judge noted that discovery in the case had been contentious, the parties spent months meeting and conferring. *Progressive* had a team of 8 contract attorneys reviewing the documents only to abandon the manual review option it selected because it was taking too long and was not cost effective.

At par 11 she said:

"Progressive is unwilling to engage in the type of cooperation and transparency that its own e-discovery consultant has so comprehensively and persuasively explained is needed for a predictive coding protocol to be accepted by the court or opposing counsel as a reasonable method to search for and produce responsive ESI. Progressive is also unwilling to apply the predictive coding method it selected to the universe of ESI collected. The method described does not comply with all of Equivio's recommended best practices. The court agrees with the FDIC-R that approving Progressive's predictive coding proposal or for that matter FDIC-R's competing predictive coding protocol, will only result in more disputes."

At pp15-16, *Dynamo Holdings v CIR 143 TC No 9* (September 17, 2014) Judge Buch cited *Progressive* (par 8) as an authority for the statement that: *"predictive coding has proved to be an accurate way to comply with a discovery request for ESI and that studies show it is more accurate than human review or keyword searches."*

Commercially Sensitive Not- Relevant Documents

60. A distinction between documents which are considered "confidential documents" and "commercially sensitive documents" is made by the Plaintiffs. In effect, commercially sensitive documents are linked to the Plaintiff's previous business as a bank. It follows that the plaintiff will hold, amongst the documents which are to be searched to make discovery, files which relate to other clients who are not a party to these proceedings.

61. Section 1 of the Anglo Irish Bank Corporation Act 2009 provides a definition of "commercially sensitive information" as follows:-

"Commercially sensitive information" means information the disclosure of which could reasonably be expected to—

(a) materially prejudice the commercial or industrial interests of a person or of a group or class of persons, or

(b) prejudice the competitive position of a person in the conduct of the person's business, profession or occupation;"

62. The Freedom of Information Act 2014 gives the following definition:-

"36. (1) Subject to subsection (2), a head shall refuse to grant an FOI request if the record concerned contains—

(a) trade secrets of a person other than the requester concerned,

(b) financial, commercial, scientific or technical or other information whose disclosure could reasonably be expected to result in a material financial loss or gain to the person to whom the information relates, or could prejudice the competitive position of that person in the conduct of his or her profession or business or otherwise in his or her occupation, or

(c) information whose disclosure could prejudice the conduct or outcome of contractual or other negotiations of the person to whom the information relates."

63. The Defendants submit that neither of these definitions applies in this case as the 2009 Act applies to an assessor preparing a report, and the 2014 Act was enacted on the 14th October, 2014 after the time of the Protocol.

64. The Protocol submitted provides that the X documents, i.e. which are (a) relevant but privileged and (b) in the subset of not-relevant documents, either privileged or commercially sensitive, will not be disclosed for the purposes of reviewing the accuracy of the coding process. The Defendants raise a concern about transparency in respect of these documents. The Plaintiffs, in this regard, submit that relevant documents, over which privilege is claimed, will ultimately be discovered in the normal way on affidavit.

Discussion

1. Compliance with Order 31, rule 12

65. The Rules of the Superior Courts do not require that a manual review be carried out in the discovery process. Rather, as is stated by Fennelly J. in *Ryanair Plc v. Aer Rianta CPT*, the RSC seek to uphold the administration of justice in a manner which is equitable and "encompasses the objectives of expedition and economy". As noted by Geoghegan J. in *Dome Telecom Ltd. V. Eircom Ltd.*, the principles of efficiency and case management are integral considerations. The learned judge also clearly indicated that "if there is no rule in existence precisely covering the situation the court has an inherent power to fashion its own procedure and, even if there was a rule applicable, the court is not necessarily hidebound by it."

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.

68. As technology assisted review combines man and machine, the process must contain appropriate checks and balances which render each stage capable of independent verification. A balance must be struck between the right of the party making discovery to determine the manner in which discovery is provided and participation by the requesting party in ensuring that the methodology chosen is transparent and reliable. Ordinarily, as the rules in other jurisdictions provide, this is a matter of agreement between the parties at the outset. Agreement, as Clarke J. said in *Thema*, gives the parties "an added degree of comfort that a failure of the system to throw up a relevant document will be more likely to be viewed as unfortunate but unavoidable rather than a deliberate act".

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.

2 The Plaintiff's Proposed Protocol.

Provided a party seeking to make discovery using predictive coding acts bona fide and the proposed system is transparent, opposition or non-cooperation by the requesting party should not deter the Court from making an appropriate order.

The evidence establishes that once it became apparent to them that there was scope for considerable savings in costs and time, the plaintiff's solicitors:

- sought the defendants' consent
- furnished the defendants with their expert's explanation of the methodology on affidavit;
- produced their expert at a joint meeting on 7th July, for questioning by the defendants and their lawyers; and
- followed up with a detailed letter setting out the expert's explanations on the matters raised by the defendants at the joint meeting, and, in particular, provided a copy of the search words used in the scoping exercise.

Ms. Harty indicated a willingness to consider any search terms the defendants might suggest.

For whatever reasons, the defendants chose not to suggest additional or alternate search terms and engage in a stage of the process common to manual and technology assisted review.

It is not reasonable, therefore, for the defendants to object to the search terms adopted by the plaintiffs.

When it became clear that the defendants would not consent to the use of predictive coding in this case, the plaintiffs quite properly sought the approval of the court, in accordance with the proposed Protocol.

70. In summary, the process has involved:

- the plaintiffs requesting the defendants consent at the outset;
- the defendants being provided with a schedule of the key-words used and afforded the opportunity of suggesting their own;
- the plaintiffs indicating that the organisation and conduct of the plaintiff's discovery obligations would be carried out by a team of senior lawyers (expert reviewers) under Ms. Harty, who will have expert assistance; and
- the defendants being provided with detailed explanations of the TAR methodology and proposed protocol by the plaintiff's expert.

The Protocol further provides for:

- notification to the defendants of the threshold when the training phase has reached stabilisation point and having the opportunity to challenge it in court and

- a disclosure procedure for checking the plaintiff's coding of non-relevant documents including documents which are confidential.

71. I am satisfied that the proposed protocol will be more efficient than manual review in terms of saving costs and saving time. Mr. Crowley, in his affidavit sworn 2nd September, 2014, avers his belief that less than 10% of the 680,809 documents would need to be manually reviewed after employing predictive coding. It is clear from this evidence how the cost of the discovery process and timeframe in which it would take place would, thus, be substantially reduced.

72. The defendants' objection to the procedure proposed for verification of the plaintiff's coding of commercially sensitive, non-relevant documents must be seen in context. First, as the commercially sensitive documents are a component of the subset of non-relevant documents, the reliability of the plaintiff's coding can be tested by reference to the recall, precision, and f-measure results and in particular the quality control tests for checking documents below the threshold. Therefore, in the interests of further transparency, it is not necessary to extend the test. Second, in a manual review, a requesting party would not be entitled to rummage through all the discovering party's non relevant documents in the hope of finding relevant documents. The US cases indicate that the requesting party has no entitlement to see non-responsive documents. In my view, the plaintiffs' solution for the protection of commercially sensitive non-relevant documents accords with the "happy medium" decision of the US Tax Court in *Dynamo Holdings v CIR* in which the court followed *Da Silva Moore*. In *Dynamo*, the respondent suggested two options to the court in relation to the discovery of data held in 2 storage tapes. One that the petitioner hands over the tapes to the respondent subject to a claw-back clause relating to privilege and confidentiality. Or alternatively, that the petitioner would carry out a full linear review to exclude privileged and confidential information. The court granted the petitioner's motion granting leave for the use of predictive coding to identify relevant documents.

73. I am satisfied that the plaintiffs' definition of "commercially sensitive documents" is appropriate.

74. In my view, the plaintiffs have approached the issue of transparency in a bona fide manner.

75. I am satisfied in the circumstances of this case that, subject to one amendment, the Protocol "contains standards for measuring the reliability of the process and builds in appropriate levels of participation by the Defendants".

Amendment

(a) Both the personal defendants and the Senat defendants should have a nominated counsel for the purposes of section 6.3. The undertaking required of the nominated counsel shall not prohibit discussion of the PC Schedule with other counsel engaged by the defendants.

(b) For the purposes of section 6.5 each nominated counsel may attend the meeting accompanied by one other counsel from that defendant's team.