



Neutral Citation Number: [2016] EWHC 256 (Ch)

Case No: HC-2014-000038

IN THE HIGH COURT OF JUSTICE
CHANCERY DIVISION

Royal Courts of Justice
Strand, London, WC2A 2LL

Date: 16/02/2016

Before :

MASTER MATTHEWS

Between :

- (1) Pyrrho Investments Limited
- (2) MWB Business Exchange Limited
- and -

Claimants

- (1) MWB Property Limited
- (2) Rick Aspland-Robinson
- (3) Keval Pankhania
- (4) Richard Balfour-Lynn
- (5) Jagtar Singh

Defendants

James Knott (instructed by **Reynolds Porter Chamberlain LLP**) for the **Claimants**
Clive Freedman QC (instructed by **Simons Muirhead & Burton**) for the **Second Defendant**
Andrew George QC (instructed by **Taylor Wessing LLP**) for the **Fourth Defendant**

The other parties were represented by their solicitors, but did not address the court

Hearing date: 2 February 2016

Approved Judgment

I direct that pursuant to CPR PD 39A para 6.1 no official shorthand note shall be taken of this Judgment and that copies of this version as handed down may be treated as authentic.

.....

MASTER MATTHEWS

Master Matthews :

Introduction

1. On 2 February 2016 I made an order in this claim on an application concerning the parties' obligations regarding electronic disclosure ('e-disclosure'). That order recited the court's approval of the use in the present case of what has here been called 'predictive coding' in the disclosure process. In the circumstances, the parties were quite right to seek the court's approval. Because of the novelty, in this jurisdiction at least, of such use, I said that I would give my reasons later for that approval. These are those reasons. I straight away record my gratitude to counsel and solicitors involved for their assistance.
2. The claim form was originally issued on 20th March 2013. The First Claimant was a significant shareholder in the Second Claimant. However, the First Claimant sued as assignee of the Second Claimant, in respect of payments that had been allegedly made by the Second Claimant as a result of the breach of fiduciary duty by the Second to Fifth Defendants as directors of the Second Claimant. The Second Claimant is joined in case of any issues about the assignment. The value of this part of the claim is said to run into the tens of millions of pounds. The Claimants say that some of the payments enured for the benefit of the First Defendant, which is alleged to be liable to account for them.
3. In addition to the claims about payments made through breach of fiduciary duty of the Second to Fifth Defendants, there is a specific claim in respect of a dividend that was declared by the Second Claimant on 11th June 2009 in the sum of approximately £9 million. The Second Claimant was formerly listed on AIM. Its ultimate parent company was listed on the main stock exchange, but went into administration in November 2012.
4. The claim was amended in 2014 so as to include a second group of claims. These complained that the Second to Fifth Defendants caused the Second Claimant to enter into transactions with companies in which they themselves were secretly interested, and thereby extracted some £28.5million from the Second Claimant over a period of five years. The claim form has since been re-amended on one further occasion. The trial in this matter is now fixed for June 2017.

Disclosure

5. Disclosure, and in particular e-disclosure, can be a problem in any case. It is a particular problem in this case. It is common ground that the bulk of relevant documents are likely to be in the control of the Second Claimant. The Second Claimant controls back-up tapes on which data from email accounts used by the Second to Fifth Defendants are stored. To give an idea of the scale of the exercise, the total number of electronic files restored from the back-up tapes of the Second Claimant was originally more than 17.6 million. This has since been reduced to some 3.1 million by a process of electronic de-duplication. But it is still a large and costly number to search.
6. Disclosure is governed by CPR Part 31 and its Practice Directions. I set out some of these provisions later. It is also governed by the directions given by the Court in the

particular case. In this case, paragraphs 10-15 of the consent order of 7 August 2015 made the following directions concerning disclosure:

“Disclosure

10. The parties shall by 4.30pm on Friday, 26 February 2016 give standard disclosure by simultaneous exchange of documents by lists and categories. Any requests for inspection or copies of disclosed documents shall be made within 14 days of service of the lists and shall be responded to within 7 days of receipt of the request.

Electronic Disclosure

11. The parties shall seek to agree, if possible, the scope of the reasonable search to be carried out by each party for electronic documents. Further:
- i. the Defendants shall respond to RPC in relation to electronic disclosure (in particular RPC's letters dated 25 February 2015 and 6 March 2015) by no later than 4.30pm on Friday 13 November 2015;
 - ii. the Claimants shall by no later than 4.30pm on Friday 20 November 2015 provide the Defendants with a reply to the responses received pursuant to paragraph 11.i above.
12. In the event that the parties have not reached agreement in relation to the scope of the reasonable search for electronic documents to be carried out by any party by 4.30pm on Friday 27 November 2015 any outstanding issues shall be considered and determined at the Disclosure CMC referred to below.

Disclosure CMC

13. The Parties shall within 7 days of the date of this order seek to fix a further CMC (the “Disclosure CMC”) for the first mutually convenient date after 27 November 2015 with a time estimate of 2.5 hours.
14. The purpose of the Disclosure CMC shall be to determine:
- i. any issues remaining in relation to electronic disclosure, in accordance with paragraphs 11 and 12 above; and
 - ii. (as between the Claimants and the Third Defendant) the appropriate mechanism to be adopted in relation to documents currently held in quarantine by Stroz Friedberg which have been the subject of correspondence between the Third Defendant and the Second Claimant.
15. Any issues referred to in paragraph 14.i shall be dealt with first in time at the Disclosure CMC and the First, Second, Fourth and Fifth Defendants need not attend the remaining part of the Disclosure CMC (which shall be given over to dealing with the issue referred to in paragraph 14.ii). In the event that there are no issues arising under paragraph 14.i for determination at the Disclosure

CMC the First, Second, Fourth and Fifth Defendants need not attend the Disclosure CMC.”

(The Disclosure CMC contemplated by those directions was in part held on 2 February 2016 and the order referred to in paragraph 1 above was made. The hearing contemplated by paragraph 14.ii has yet to be held.)

7. Under the Civil Procedure Rules, where the obligation of a party to Part 7 proceedings is to give standard disclosure (see CPR rule 31.6), that party is obliged to make a search for disclosable documents other than those on which the party relies. “Disclosable documents” means those in the party’s control and falling within certain categories. For this purpose, “document” clearly includes a computer file. The search obligation is set out in rule 31.7:

“(1) When giving standard disclosure, a party is required to make a reasonable search for documents falling within rule 31.6(b) or (c).

(2) The factors relevant in deciding the reasonableness of a search include the following –

(a) the number of documents involved;

(b) the nature and complexity of the proceedings;

(c) the ease and expense of retrieval of any particular document; and

(d) the significance of any document which is likely to be located during the search.

(3) Where a party has not searched for a category or class of document on the grounds that to do so would be unreasonable, he must state this in his disclosure statement and identify the category or class of document.”

8. The rules in Part 31 are supplemented by two Practice Directions. One of them Practice Direction B, deals with e-disclosure. A number of its provisions deal with the “reasonable search” under rule 31.7. Two in particular are relevant here:

“20 The extent of the reasonable search required by rule 31.7 for the purposes of standard disclosure is affected by the existence of Electronic Documents. The extent of the search which must be made will depend on the circumstances of the case including, in particular, the factors referred to in rule 31.7(2). The parties should bear in mind that the overriding objective includes dealing with the case in ways which are proportionate.

21 The factors that may be relevant in deciding the reasonableness of a search for Electronic Documents include (but are not limited to) the following –

(1) the number of documents involved;

(2) the nature and complexity of the proceedings;

(3) the ease and expense of retrieval of any particular document. This includes:

(a) the accessibility of Electronic Documents including e-mail communications on computer systems, servers, back-up systems and other electronic devices or media that may contain such documents taking into account alterations or developments in hardware or software systems used by the disclosing party and/or available to enable access to such documents;

- (b) the location of relevant Electronic Documents, data, computer systems, servers, back-up systems and other electronic devices or media that may contain such documents;
 - (c) the likelihood of locating relevant data;
 - (d) the cost of recovering any Electronic Documents;
 - (e) the cost of disclosing and providing inspection of any relevant Electronic Documents; and
 - (f) the likelihood that Electronic Documents will be materially altered in the course of recovery, disclosure or inspection;
- (4) the availability of documents or contents of documents from other sources; and
- (5) the significance of any document which is likely to be located during the search.”

9. These rules and other provisions demonstrate that what matters fundamentally in the disclosure process is the scope and quality of the *search*, rather than the *listing* and *production for inspection* of the relevant documents found (though these are important too). If the search is defective it will not be corrected by what happens at the stages of listing or production.

10. However, whilst the rules contemplate the search for electronic documents, neither the CPR nor the Practice Directions deal in any detail with the question *how* the search should be conducted. In particular they do not deal with the extent to which it is permissible to conduct e-disclosure through the medium of a computer program or programs, rather than through the intervention of human beings. Nevertheless, Practice Direction B does at least say this (emphasis supplied):

“**25** It may be reasonable to search for Electronic Documents by means of Keyword Searches *or other automated methods of searching* if a full review of each and every document would be unreasonable.

26 However, it will often be insufficient to use simple Keyword Searches or other automated methods of searching alone. The injudicious use of Keyword Searches *and other automated search techniques* –

- (1) may result in failure to find important documents which ought to be disclosed, and/or
- (2) may find excessive quantities of irrelevant documents, which if disclosed would place an excessive burden in time and cost on the party to whom disclosure is given.

27 The parties should consider supplementing Keyword Searches *and other automated searches* with additional techniques such as individually reviewing certain documents or categories of documents (for example important documents generated by key personnel) and taking such other steps as may be required in order to justify the selection to the court.”

11. I add that the judges of the Technology and Construction Court support an *eDisclosure Protocol*, produced by practitioners and available on the website of the Technology and Construction Solicitors’ Association. This does contemplate the use

of predictive coding software in appropriate cases. But it is only a protocol and has no normative force.

12. Lastly on the rules, I mention that CPR rule 1.2 states that:

“The court must seek to give effect to the overriding objective when it-

- (a) exercises any power given to it by the Rules; or
 - (b) interprets any rule,
- subject to rules 76.2, 79.2, 80.2 and 82.2.”

None of the last-mentioned rules is relevant to the present case. The overriding objective is set out in rule 1.1(1), and is well-known. It is “enabling the court to deal with cases justly and at proportionate cost”. This is amplified in rule 1.1(2), and includes, so far as practicable, saving expense and dealing proportionately with the case.

13. In a case a few years ago now, *Goodale v Ministry of Justice* [2009] EWHC B41 (QB), the then Senior Master of the Queen’s Bench Division, Master Whitaker, explained the e-disclosure problem in this way:

“1. This judgment concerns a serious practical problem for the case management of disclosure which is now occurring on a regular basis. The reason is that, since certainly the beginning of this decade, increasing numbers of public bodies and private businesses, not to mention individuals, have gone over to creating, exchanging and storing their documentation and communicating with each other entirely by electronic means. The end result is that an enormous volume of information is now created, exchanged and stored only electronically. Email communication, word processed documents, spreadsheets and ever increasing numbers of other forms of electronically stored information (“ESI”) now often form the entire corpus of the documentation held by companies and individuals who become involved in litigation. So the incidence of paper disclosure is becoming less and less prevalent though in some cases it may still be critical. and the incidence of the disclosure of electronically stored information, or ESI as it is known, is becoming more and more so.

2. What is more, the volume of the ESI, even in small organisations is immense, often, as in the case of email, because of the huge quantities of documents created (including wide-scale duplication) and the fact that the documents can exist in many different forms and locations so that they are not readily accessible except at significant cost. It is also commonplace for many individuals to have more than one email account – business, personal, web-mail (for example, Yahoo, Gmail, Hotmail etc.) When ESI is available, metadata (literally data about data) associated with it can easily be unintentionally altered by the very act of collection, which in some circumstances can have a detrimental effect on the document’s evidential integrity. What is more, ESI can be moved about nationally and internationally, indiscriminately and at lightening speed.

3. What is the problem with this in litigation? Disclosure is a tripartite exercise of search, disclosure, and inspection, and the problem, when it comes to ESI is often for a party to gauge the scope of a reasonable search for ESI under CPR Rule 31.7 and PD31(2). The problem is how the parties and (if disputed) the court determines what the scope of that search of ESI should be, how it is going to be made proportionate and how it is going to be carried out correctly first time, without the court having to order it to be done again, as has occurred, for example, in **Digicel (St Lucia) Ltd and others v. Cable and Wireless Plc and Others** 2008 EWHC 2522 (Ch) in which case Morgan J ordered the defendants re-do their ESI search exercise at an additional cost of something like £2 million.

4. By contrast, except in unusual cases, in the case of paper disclosure, parties usually know what paper they have. Often the problem is merely locating it physically and going through it to produce the documents required by the standard disclosure test. The problem with ESI is that, because of the matters mentioned above in paragraph 1, parties often do not know how much ESI they have, or where it is. They might have a idea as to which servers it is on or which personal computers it is on, or which back-up tapes it is on, but without a great deal more information, it is very difficult for them to know how much documentation will be revealed by searches of the media on which their ESI is stored and how much it is going to cost to search it and what the end result is going to be. A further issue might be that not all forms of ESI are searchable. Therefore, it has to be accepted that any search is not necessarily conclusive as to whether a particular document exists. Equally often the parties do not know where to begin their searches. In the case, for example, of email, the relevant servers are often not in their possession and sometimes not even in the jurisdiction. An ill considered search for ESI may produce far too few documents for review but more likely will produce such volumes that human review of every document is neither proportionate nor practical. Because of this a substantial industry has developed to handle the identification, collection, reduction and organisation for review of ESI. Often, this is carried out electronically, with technology aiding and supplementing human review.”

14. In that case, it was proposed that initially there should be a keyword search of a large electronic database to see how many documents were turned up by it. Then the court could consider what should be done next. The Senior Master said this:

“26. That searching, because it is going to be done in a comparatively simple way, without using specialist software at this stage, is just going to give us the potential numbers of documents. Similarly, doing the same type of search in respect of the MEDS system for the 31 terms but only in respect of each of the key witnesses, will give us the potential number of documents in respect of that as well. It is at that stage, when that crude way of finding out what documents might be in existence is completed, that a service provider will have to be agreed between the parties, and will have to be instructed to look at what the next stage of the exercise should involve and how much it is going to cost, in order to produce a corpus of documents which is reviewable by both parties.

27. At the moment we are just staring into open space as to what the volume of the documents produced by a search is going to be. I suspect that in the long run this crude search will not throw up more than a few hundred thousand documents. If that is the case, then this is a prime candidate for the application of software that providers now have, which can de-duplicate that material and render it down to a more sensible size and search it by computer to produce a manageable corpus for human review – which is of course the most expensive part of the exercise. Indeed, when it comes to review, I am aware of software that will effectively score each document as to its likely relevance and which will enable a prioritisation of categories within the entire document set.”

So the Senior Master certainly contemplated that specialist software might be brought into play to score the hundreds of thousands of anticipated e-documents for relevance and therefore possible disclosure in the proceedings. But so far as I am aware he said no more than that.

15. In the present case, and in accordance with the rules, Electronic Disclosure Questionnaires were exchanged in February 2015. By paras 11 and 12 of my order of 7 August 2015, the parties were to attempt to reach agreement on the scope of e-disclosure, failing which any outstanding issues were to be determined at a CMC. Commendably, several rounds of correspondence between the parties have resulted in large measures of agreement. There are or will be some areas of disagreement which will or may have to be determined later, but for now the parties have, subject to the approval of the court, agreed on the (automated) method to be employed in the Second Claimant’s e-disclosure exercise, and also the scope of the keywords to be employed.
16. That method to be employed involves ‘predictive coding’, and the purpose of this judgment is to explain shortly why the court approved its use in this case. Amongst the Defendants, the lead on this issue was taken by the Fourth Defendant. Mr Edward Spencer, an associate solicitor from the Fourth Defendant’s solicitors, Taylor Wessing LLP, made a witness statement on 17 December 2015, explaining the Fourth Defendant’s proposal that the Second Claimant use ‘predictive coding’ in the e-disclosure exercise. This judgment owes a considerable debt to that witness statement.

Predictive coding

17. Mr Spencer explains in his statement that the term ‘predictive coding’ is used interchangeably with ‘technology assisted review’, ‘computer assisted review’, or ‘assisted review’. It means that the review of the documents concerned is being undertaken by proprietary computer software rather than human beings. The software analyses documents and ‘scores’ them for relevance to the issues in the case. This technology saves time and reduces costs. Moreover, unlike with human review, the cost does not increase at the same rate as the number of documents to be reviewed increases. So doubling the number of documents does not double the cost.
18. I should say, by way of footnote, that the ideas underpinning this process are not completely new. Primitive versions of this kind of process were being demonstrated

to (sometimes sceptical) litigation lawyers in the mid-1980s. I was one of them. But this was before the advent of personal computers, let alone of tablets and smartphones. There was no everyday or home computer culture then, and especially not amongst English lawyers. Now computers and computer technology are much more accepted as the norm, and, crucially, the technology is vastly better, for example in terms of storage size, portability of hardware and storage media, processor speed and programming, amongst other matters. A number of computer software companies now offer predictive coding software for use by lawyers.

19. In modern times, as I understand it, 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).
20. 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.
21. 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. (These quality assurance exercises are clearly “additional techniques” contemplated by paragraph 27 of Practice Direction B to Part 31.)
22. 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. (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.

23. 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. In his statement, Mr Spencer says that he understands that in fact it should involve review of some 8 to 12 batches of documents. 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.
24. 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.

The authorities

25. In England there is not a great deal by way of guidance, and nothing by way of authority, on the use of such software as part of the disclosure process. The fleeting reference in *Goodale v Ministry of Justice* to specialist software has already been mentioned. There is some practical guidance in a short article by Celina McGregor (a practising solicitor) in *The Solicitor's Journal* for 5 August 2014, at page 27. A more detailed guide is the *Guide to eDisclosure*, produced by the Technology and Construction Solicitors' Association, and available on its website. A rare comment in an English textbook is that of Charles Hollander QC, in *Documentary Evidence*, 12th ed, 2015, at [9-20]:

“At present, the population of documents identified is normally searched for relevance by lawyers or paralegals. In the United States electronic searching is beginning to be introduced. Tests have shown that it is more reliable than review by humans. No doubt this will be with us soon.”

A footnote to this statement reads:

“There is a judgment of the US District Court for the Southern District of New York which discusses predictive coding in detail: *Moore v Publicis Groupe*, unreported, February 24, 2012. The judge was Judge Andrew Peck, who has written widely on predictive coding...”

26. The US Federal Court case of *Moore v Publicis Groupe*, 11 Civ 1279 (ALC)(AJP), referred to in that extract, was one where the Plaintiffs were female employees of the Defendants, and alleged gender and other kinds of discrimination against them on the part of the Defendants, as well as breaches of equal pay legislation (all of which the Defendants denied). The document set was in excess of 3 million documents, and the

assigned magistrate judge endorsed the use of computer technology in carrying out the discovery process. A magistrate judge is, I understand, a delegate of the judges of the federal district court, and is assigned to a particular case together with a particular district judge.

27. The issues between the parties in relation to electronic disclosure covered the questions (i) which custodians' emails would be searched, (ii) whether there should be one or two phases of such searches, (iii) the number of iterative rounds needed to stabilise the training of the software, (iv) whether accepting the Defendants' predictive coding approach would allow their lawyers to certify discovery as complete when it was not; (v) whether it was contrary to the Federal Rules of Evidence, and (vi) whether it was impossible to assess whether predictive coding would produce accurate results. The magistrate judge held that the Plaintiffs' concerns in issues (iv) to (vi) were unfounded.
28. The magistrate judge also made some general comments, of which the following seem to be the most useful from an English perspective (footnotes and citations omitted):

“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 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 it 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 ediscovery 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 review method, at cost proportionate to the 'value' of the case.
[...]

The slightly more difficult case would be where the producing party wants to use computer-assisted review and the requesting party objects. The question to ask in that situation is what methodology would the requesting party suggest instead? Linear manual review is simply too expensive where, as here, there are over three million emails to review. Moreover, while some lawyers still consider manual review to be the 'gold standard', that is a myth, as statistics clearly show that computerised searches are at least as accurate, if not more so, than manual review. [...]

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, knowledge of the responding party's 'cards' [...]

Another problem with keywords is that they often are over-inclusive, that is, they find responsive documents but also large numbers of irrelevant documents. [...]

Computer-assisted review appears to be better than the available alternatives, and thus should be used in appropriate cases. ”

29. The Plaintiffs objected to the decision of the magistrate judge, and it was subject to review by the assigned district court judge, Judge Andrew L Carter Jr. Under the applicable rules, the judge could modify or set aside any part of the order of the magistrate judge “that is clearly erroneous or contrary to law”. However the judge in his judgment of 25 April 2012 referred to authorities which held that “magistrates are afforded broad discretion in [non-dispositive] disputes and reversal is appropriate only if their discretion is abused,” and that matters concerning discovery were considered non-dispositive of the litigation. Judge Carter then affirmed the decision, saying this:

“Mindful of this highly deferential standard of review, the Court adopts Judge Peck's rulings because they are well reasoned and they consider the potential advantages and pitfalls of the predictive coding software.”

30. The judge considered various arguments about the reliability of the software, and concluded:

“There is simply no review tool that guarantees perfection. The parties and Judge Peck have acknowledged that there are risks inherent in any method of reviewing electronic documents. Manual review with keyword searches is costly, though appropriate in certain situations. However, even if all parties here were willing to entertain the notion of manually reviewing the documents, such review is prone to human error and marred with inconsistencies from the various attorneys' determination of whether a document is responsive. Judge Peck concluded that under the circumstances of this particular case, the use of the predictive coding software as specified in the ESI protocol is more appropriate than keyword searching. The Court does not find a basis to hold that his conclusion is clearly erroneous or contrary to law. Thus Judge Peck's orders are adopted and Plaintiffs' objections are denied.”

31. Closer to home, the Irish High Court has also endorsed the use of predictive coding, in *Irish Bank Resolution Corporation Ltd v Quinn* [2015] IEHC 175. There, unlike the present case, the use of the technique was not agreed between the parties. The judge, Fullam J, considered the *Moore* case, and pointed out that the rules of court (still based on the English RSC 1883, Ord 31) did not require that a manual review of documents be carried out. He further said:

“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.”

32. So far as I am aware, no English court has given a judgment which has considered the use of predictive coding software as part of disclosure in civil procedure. At all events, a search of the BAILII online database for “predictive coding software” returned no hits at all, and for “predictive coding” and “computer-assisted review” only the Irish case referred to above.

Decision

33. In the present case, the factors in favour of approving the use of predictive coding technology in the disclosure process seemed to me to be these:

- (1) Experience in other jurisdictions, whilst so far limited, has been that predictive coding software can be useful in appropriate cases.

- (2) There is no evidence to show that the use of predictive coding software leads to less accurate disclosure being given than, say, manual review alone or keyword searches and manual review combined, and indeed there is some evidence (referred to in the US and Irish cases to which I referred above) to the contrary.
- (3) Moreover, there will be greater consistency in using the computer to apply the approach of a senior lawyer towards the initial sample (as refined) to the whole document set, than in using dozens, perhaps hundreds, of lower-grade fee-earners, each seeking independently to apply the relevant criteria in relation to individual documents.
- (4) There is nothing in the CPR or Practice Directions to prohibit the use of such software.
- (5) The number of electronic documents which must be considered for relevance and possible disclosure in the present case is huge, over 3 million.
- (6) The cost of manually searching these documents would be enormous, amounting to several million pounds at least. In my judgment, therefore, a full manual review of each document would be “unreasonable” within paragraph 25 of Practice Direction B to Part 31, at least where a suitable automated alternative exists at lower cost.
- (7) The costs of using predictive coding software would depend on various factors, including importantly whether the number of documents is reduced by keyword searches, but the estimates given in this case vary between £181,988 plus monthly hosting costs of £15,717, to £469,049 plus monthly hosting costs of £20,820. This is obviously far less expensive than the full manual alternative, though of course there may be additional costs if manual reviews still need to be carried out when the software has done its best.
- (8) The ‘value’ of the claims made in this litigation is in the tens of millions of pounds. In my judgment the estimated costs of using the software are proportionate.
- (9) The trial in the present case is not until June 2017, so there would be plenty of time to consider other disclosure methods if for any reason the predictive software route turned out to be unsatisfactory.
- (10) The parties have agreed on the use of the software, and also how to use it, subject only to the approval of the Court.

There were no factors of any weight pointing in the opposite direction.

34. Accordingly, I considered that the present was a suitable case in which to use, and that it would promote the overriding objective set out in Part 1 of the CPR if I approved the use of, predictive coding software, and I therefore did so. Whether it would be right for approval to be given in other cases will, of course, depend upon the particular circumstances obtaining in them.