

# **Equivio>Relevance Putting What Matters First**

---

*White Paper*

*by*

*Chris Dale of the e-Disclosure Information Project*



**equivio**  
zoom in. find out



# INTRODUCTION

## EXECUTIVE SUMMARY

This is one of a series of white papers written by Chris Dale of the UK-based **e-Disclosure Information Project**. Its purpose is to describe a software application called Equivio>Relevance which enables automated prioritisation of documents and keywords. Many UK lawyers are reluctant to rely on technology to cull and filter document collections, which they see as delegating functions to a machine whose output they cannot check; they imagine that this is falling short of their duties to the court and clients under the UK Civil Procedure Rules.

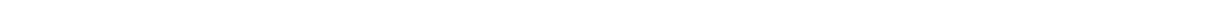
This paper considers what those duties actually are, and shows that parties are not required to look under every stone for disclosable documents. Their response must be proportionate to the object to be achieved, and the large volumes of documents involved in modern litigation cannot be handled proportionately (or, often, at all) by large teams of people starting at the earliest document and working to the end. Quite apart from what the courts require, the clients will not pay for the high manual input of expensive lawyers. They will go to other firms, take the work in house, or not litigate. Finally, the paper describes briefly what Equivio>Relevance does, and shows that its output is capable of being checked in ways not possible with the alleged “gold standard” of manual review.

Equivio>Relevance is described at [www.equivio.com/product.asp?ID=7](http://www.equivio.com/product.asp?ID=7)

## THE E-DISCLOSURE INFORMATION PROJECT

The e-Disclosure Information Project disseminates information for those with an interest in electronic disclosure in the UK courts, including judges, practitioners, suppliers and corporate clients. Its aim is the reduction of the expense of litigation. It is run by Chris Dale, a former commercial litigation solicitor and adviser on all aspects of electronic disclosure, including the court rules, the practical issues which arise and the solutions which exist to tackle them.

The main expectations of such a project are that it is knowledgeable, independent and objective. It has no client and can exist only if it is funded by sponsorship. The sponsors have in common that they are interested in and knowledgeable about aspects of e-Disclosure wider than their own commercial advantage, and that they are willing to pool resources in this indirect way to raise understanding of the issues. Equivio is a sponsor of the Project.





## DISCLAIMER

This white paper is written by Chris Dale in conjunction with Equivio as an informational resource only. It is not to be relied upon as a source of legal or technical advice.

## THE TARGET AUDIENCE FOR THIS PAPER

The specific focus of this paper is on litigation, primarily from a UK perspective. The aim is to show how the functions of Equivio>Relevance can be harnessed in the context of the formal requirements of the Civil Procedure Rules to reconcile the court's expectations with the clients' commercial expectations at a time when document volumes have made litigation too expensive for many companies. The primary messages – that this is a cost-effective, proportionate and lawyer-led response to the problem – are applicable also in litigation in other jurisdictions and in areas such as response to regulatory demands and internal investigations.

The target audience for this paper is people who fit one or more of the following criteria: those who appreciate that large volumes of electronic documents cannot be ignored; those who know the rules but find it hard to relate them to the real-life requirements of both courts and clients; those whose conscientiousness conflicts with budgetary constraints; and those whose willingness to grasp the technology is hampered by the difficulty of relating it to the conventional way of giving disclosure.

The focus on conscientiousness and conventional working is derived from the observation that many lawyers, even those who recognize that large volumes of documents cannot be handled by eye alone, are afraid of delegating part of their function to a machine. Their dominant drivers are doing a good job for their clients and fulfilling their duties to the court, and their fear is of falling short of the standard of care expected of them. This fear derives in part from the tension between the traditional idea that one must read everything and the brute practicality that one cannot do so with present-day volumes of documents. It is necessary to some extent to "let go". This involves two components: an acceptance that the rules provide a framework within which a wide discretion operates, and sufficient understanding of the technology to compare it with the alleged "gold standard" of human review. I take it as a given for these purposes that human review of large volumes of documents is neither practicable in any commercial sense nor, in fact, a "gold standard" at all. It is, nevertheless, seen as such by many conscientious lawyers.

Section 5 of this paper describes what the process is, that is, explains how lawyer and software work together to produce a result which is quickly achieved, which satisfies the lawyer that decisions as to relevance have been made with adequate accuracy, and which will satisfy the court that the rules have been complied with. It is necessary first to explain briefly what the court's expectations are, because they are commonly



misunderstood. If many lawyers routinely ignore the obligation to consider electronic documents completely, many approach disclosure as if their duty is to disclose every relevant document. That is not what is required by a proper application of the rules and case law.

## WHAT THE COURT EXPECTS FROM LAWYERS

### WHAT THE RULES REQUIRE

It is helpful to list in summary form some of the factors which define the target -- the crossover, that is, between a lawyer's professional duties (not least as an officer of the court), the need to find "smoking gun" documents, and the practical and commercial realities of time and cost. In no particular order, these include the following:

- The definition of a disclosable document Rule 31.6 CPR is not "relevant" but the potentially much narrower test that documents are supportive of or adverse to the cases of the giver or any other party.
- The duty of search is required to be "reasonable". Rule 31.7 CPR and the Practice Direction to Rule 31 set out factors to be considered in deciding what is reasonable.
- There is an express obligation in Paragraph 2A of the Practice Direction to discuss difficulties arising from electronic sources of documents and to try and agree keywords for identifying disclosable documents. You cannot discuss what you cannot identify. This process takes place before the Case Management Conference which decides the scope of disclosure, putting a premium on early identification and assessment of what is involved.
- Parties may agree, subject to the court's view, to give no disclosure at all. Disclosure may therefore range between 0% and 100% of the potentially disclosable documents. This is far removed from the conventional idea that one must list every document.
- There is a difference between disclosing the existence of sources and putting every disclosable document from those sources into the ring. The disclosure statement allows one to fulfill the duty of candour by identifying sources whilst explaining that their estimated value when set against the cost of extraction makes them disproportionately burdensome to disclose.



## WHAT THE CASE LAW SAYS

Case law emphasises that the duty to disclose is circumscribed by proportionality. Morgan J said this in *Digicel (St. Lucia) Ltd & Ors v Cable & Wireless Plc & Ors* [2008] EWHC 2522 (Ch)<sup>1</sup> at [46]:

*it must be remembered that what is generally required by an order for standard disclosure is "a reasonable search" for relevant documents. Thus, the rules do not require that no stone should be left unturned. This may mean that a relevant document, even "a smoking gun" is not found. This attitude is justified by considerations of proportionality. This point is well made by Jacob LJ in *Nichia Corporation v Argos Limited* [2007] EWCA Civ 741 at [50] to [52].*

Those paragraphs in *Nichia* read as follows:

*50. There is more to be said about the change to standard disclosure and indeed to the express introduction of proportionality into the rules of procedure. "Perfect justice" in one sense involves a tribunal examining every conceivable aspect of a dispute. All relevant witness and all relevant documents need to be considered. And each party must be given a full opportunity of considering everything and challenging anything it wishes. No stone, however small, should remain unturned. Even the adversarial system at its most expensive in this country has not gone that far....*

*51 But a system which sought such "perfect justice" in every case would actually defeat justice. The cost and time involved would make it impossible to decide all but the most vastly funded cases. The cost of nearly every case would be greater than what it is about. Life is too short to investigate everything in that way. So a compromise is made: one makes do with a lesser procedure even though it may result in the justice being rougher. Putting it another way, better justice is achieved by risking a little bit of injustice.*

*52 ....the rules now sacrifice the "perfect justice" solution for the more pragmatic "standard disclosure" and "reasonable search" rules, even though in the rare instance the "right" result may not be achieved. In the vast majority of instances it will be, and more cheaply so.*

These factors clearly do not justify a cavalier attitude to disclosure, but they warrant – indeed, they require – a proportionate one. The key component is transparency, that is, reasoned explanations as to the decision-making process. Before one can be transparent with opponents and the court, one must oneself be clear as to how the final selection has been arrived at, as to what is being discarded, and how those decisions were made.

---

<sup>1</sup> <http://www.bailii.org/ew/cases/EWHC/Ch/2008/2522.html>



## WHAT THE CLIENTS WANT

It is necessary also to involve the client in the decision-making – it is all very well for a judge to say, as Morgan J said in *Digicel*, that a course which may miss the “smoking gun” is acceptable, but that does not help the lawyer vis à vis his client if he does actually miss a key document.

It is the client, more than anyone, who is worried about the costs. The lawyer must be able to make a risk assessment in terms which weigh the risk, such as it may be, of using a technological solution such as Equivio>Relevance against the costs to be saved by doing so, and then explain it to the client. The processes and methodology described below do not in fact imply any greater risk than that implicit in human selection. Humans are more fallible and less consistent. The Equivio>Relevance approach combines the best of human intelligence with the power of a software application with, in addition, a repeatable, auditable and transparent methodology.

## IS THE TECHNOLOGY GOOD ENOUGH FOR MY CLIENTS?

Since part of the purpose of this paper is to address lawyer concerns about the quality of results obtained by a technology solution like Equivio>Relevance, it is worth looking at the results of tests and trials in which Equivio>Relevance has been pitted against other solutions and against human review.

## HOW ACCURATE IS EQUIVIO>RELEVANCE TECHNOLOGY?

Lawyers proposing to delegate any part of the document selection process to a software application will inevitably ask whether it can be trusted. They mean by that whether it can be trusted to do at least as good a job as they, the lawyers, would do.

This is ground which has been well covered in other white papers whose purpose is to report on various exercises by which the precision and recall of Equivio>Relevance has been tested. These are:

How to Jump-Start Early Case Assessment: A Case Study based on TREC 2008

Am Law 100 Firm Uses Equivio>Relevance to Find More Relevant Documents and to Find Them Faster – an Epiq-Equivio Case Study

Both of these papers are available at <http://equivio.com/downloads.asp>



The Epiq paper describes a trial in which an expert using Equivio>Relevance was pitted against a human review team over a document population of 47,000. In summary, the software and the human review team agreed on the relevant/not-relevant designations on 91.4% of the documents. A human “oracle” determined by samples taken from the remainder that Equivio>Relevance was correct in most of the documents on which the software and the human team were in disagreement. The Equivio->Relevance process took only 18 man-hours of “expert” time and an hour of processing thereafter.

Comforting though these statistics appear to be (and there is no reason to doubt them), tests of this kind are persuasive rather than conclusive as to whether Equivio>Relevance (or any other software application) is going to have any specific degree of accuracy over any particular set of documents. That is not just to do with the software algorithms, but reflects amongst many other factors the absence of an objective yardstick by which to measure accuracy. No two humans, still less teams of humans, are going to make the same relevance decisions consistently. Extensive sampling and duplicated effort is needed to validate wholly human exercises.

Tests such as those described in these papers give a large measure of comfort as to the likelihood that Equivio>Relevance will give an acceptable measure of accuracy. It is not necessary, however to rely on such tests. The next section of this paper describes the iterative process by which legal teams and the software make relevance decisions including, crucially, the ability to check the accuracy of the software’s relevance estimates on a continuing basis.

It follows from what is said above about the UK disclosure rules that they do not require an idealised (and commercially unacceptable) standard of perfection, but a solution which is proportionate to what is at issue.

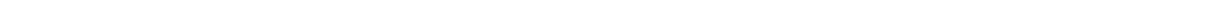
It is worth observing in addition that focus on the formal requirements of the rules ties in with what is actually more important – being able to tell the client, as quickly as possible, and at the lowest possible cost, whether he is likely to win or lose the case, and what the costs implications are going forward.

## WHAT DOES EQUIVIO->RELEVANCE DO?

This section sets out what the purpose is of Equivio>Relevance and then considers the elements which a lawyer might expect from an application to which he is committing this early stage of his client’s disclosure. The section concludes by outlining the process by which the lawyer uses Equivio>Relevance to prioritise documents for review.

### THE STATED PURPOSE OF EQUIVIO>RELEVANCE

I cannot improve on the description given on Equivio’s website which is as follows:





Equivio>Relevance enables automated prioritisation of documents and keywords. Based on initial input from a lead lawyer, Equivio>Relevance uses statistical and self-learning techniques to calculate graduated relevance scores for each document in the data collection.

As an expert-guided system, Equivio>Relevance works like this:

- An expert reviews a sample of documents, ranking them as relevant or not.
- Based on the results, Equivio learns how to score documents for relevance.
- In an iterative, self-correcting process, Equivio feeds additional samples to the expert. These statistically generated samples allow Equivio>Relevance to progressively improve the accuracy of its relevance scoring.
- Once a threshold level of accuracy is achieved, Equivio ranks the entire collection, calculating a graduated relevance score for each document.

## HOW DOES EQUIVIO->RELEVANCE FIT INTO THE PROCESS?


The problem confronting lawyers and their clients is broadly the same in all jurisdictions. In the US, everything which is potentially relevant must be ready at an early stage for the meet and confer required by the rules. In the UK, there is greater latitude to leave on one side documents which, although potentially disclosable, are arguably of little or no significance. In both cases, the challenge is to identify relevant documents under severe time and cost pressures.

### IDENTIFYING THE WIDEST POOL OF DOCUMENTS TO BEGIN WITH

The first steps more involve thoughtfulness than technology. Discussions with the clients (including its IT department) narrow the range of material to be collected in the first place. Broad categorisations such as date ranges and key people serve to identify the starting point. It may in fact be a non-trivial task in these days of multiple data sources, cloud computing etc. to decide on the primary scope, but the upshot of this stage is a pool -- possibly a very big pool -- of everything which might possibly be relevant. From this collection are removed system files, duplicates and other things which cannot possibly be relevant. Where is the lawyer to start in tackling the resulting collection?

### PRIORITISATION

None of the objectives are met by a plod through all the documents in date order. It is not that manual review is unnecessary – technology’s staunchest advocates are not yet saying that, although the influential Senior Master Whitaker is willing to predict that we will see disclosure exercises in the foreseeable future which have been entirely performed by technology. What is important is that the review takes place over the smallest set of documents, that priority is given to the most important, and that



decisions can be made as to the resources needed for subsequent stages. Some documents must be reviewed, and quickly, by senior lawyers. Documents of a lesser perceived significance might be passed to more junior staff. Those which are perceived as having little or no importance might be put to one side in the hope that no one ever has to look at them.

This requirement was expressed thus by HHJ Simon Brown QC in a conference speech:

*The Judge is the end-user of all this Disclosure activity. What I want to know is this: what is the case about? Which of the pleaded issues really matter in getting to the heart of the dispute? Can we split the case up and limit disclosure to the subjects which matter, or which matter most?*

That is an argument for prioritisation and that is the primary function of Equivio>Relevance.

#### OVER-INCLUSION AND UNDER-INCLUSION


The lawyers need to walk the line between over-inclusion and under-inclusion, the first representing a wasted cost and the second a serious risk. The technical terms here are precision and recall. Precision relates to the percentage of recovered documents which are in fact relevant, whilst recall measures the number of relevant documents which were actually found. High recall is likely to decrease precision which means, in everyday terms, that casting the net widely increases the chance of finding more relevant documents but increases also the number of non-relevant documents submitted to review; too narrow a search decreases the number of non-relevant documents, but more relevant documents will be missed. The relationship between the precision and recall is called the F-Measure. As detailed in the Equivio case study based on the TREC tests, in comparison to standard keyword searches, Equivio>Relevance simultaneously reduces both over-inclusion and under-inclusion rates.

The lawyers are less interested in the technicalities than in what this means in terms of getting the job done. The purpose in setting out the courts' requirements above is to make it clear that one is not required to "look under every stone". To pursue the analogy, Equivio>Relevance helps determine which stones are worth looking under.

It was Judge Brown, mentioned above, who gave judgment in *Earles v Barclays Bank Plc* [2009] EWHC 2500 (Mercantile)<sup>2</sup>, a case which illustrates over- and under-inclusion. He said of the 500 page trial bundle that only 10% of the documents in it were relevant. The documents actually needed to prove the central issue were not produced, the bank having apparently concluded that too much work was required to find them. The judgment, which deprived the successful defendant of half its costs, was in part a

---

<sup>2</sup> <http://www.bailii.org/ew/cases/EWHC/Mercantile/2009/2500.html>



warning to litigants to get their house in order against any foreseeable litigation, by the institution of document retention and litigation readiness policies with processes to support them. Even where that has been done, the result merely reduces the starting volumes; it is still necessary to sort the relevant from the irrelevant when litigation arises, and “relevance” is specific to the facts of a particular case. That, as its name implies, is the function of Equivio>Relevance. No amount of generic litigation readiness (which necessarily anticipates all foreseeable circumstances rather than any specific set of facts and issues) can achieve this. Equivio>Relevance goes further than the binary choice between relevant and irrelevant – it addresses also the problems of over- and under-inclusion, prioritisation and quality control specific to the case.

### MEASURING WHAT IS BEING DONE

Lawyers need also some means of measuring what they are doing. At a project management level, this goes to time estimates and cost projections, as well as comparison with alternative means of tackling similar problems. Clients complain not just of high costs but of unexpected costs. Increasingly they are requiring their lawyers to work for fixed fees, so that time- and cost-management becomes the lawyers’ own problem.

In addition, the court rules require that lawyers can show that their conduct of disclosure is proportionate to the objective, that is, that the time and cost going into the exercise is sufficient to do the job properly – and no more. Equivio>Relevance includes reports and graphs which give (amongst other things) a measure of recall and precision, allowing the user to manage cost and risk. It is possible to show, for example, what the effect would be (that is, how many more or fewer documents must be reviewed, and how many more or fewer relevant documents would be captured) of an adjustment to the target.

### QUALITY CONTROL

Lastly, the lawyers need a systematic means of quality control, that is, a means of assessing as they go along whether the results coming back are indeed representative of the true position. That requires the ability to take samples of both what has been marked as relevant and what has not, and to feed back the results into the system. Equivio>Relevance allows that iterative process to take place with the minimum of additional work.

## WORKING WITH EQUIVIO>RELEVANCE

### THE INTERCHANGE BETWEEN LAWYER AND SYSTEM

Equivio->Relevance hands out a manageable sample of say 40-50 documents which are passed to a lawyer - the human expert - for a decision as to their relevance. The lawyer makes a document-by-document decision and the system then passes him or her another batch of documents. The selection of these documents is influenced by the



decisions made as to the first batch, that is, the system determines on the basis of the information already provided what next to pass to the lawyer.

This process continues in iterative cycles, with the system constantly calculating relevance scores by reference to prior decisions. After a number of such cycles - perhaps 35 to 50 of them - the system determines that it has enough information to draw relevance conclusions over the whole of the rest of the body of documents. The user can determine where to set the bar, that is, decide what mix of recall and precision is desirable for the case, and this can be changed as suggested by the outcome. This process is far from a "black box" scenario in which the system throws out its selection of documents leaving the lawyers to take on trust what decisions have been made.

The user can set a "culling percentage", which may be fixed by reference to past experience and fine-tuned depending on results of sampling. A filter can be applied to include or exclude specific categories of documents if that is required.


One of the resulting reports shows a "Recommended action" - Don't review, Optional review or Must review which shows what percentage of documents fall into each category and what the recall and precision is for each category. This gives a quick overview of what the task ahead involves and may, for example, suggest that the target is over-ambitious relative to the time and budget. This provides a kind of sliding scale, allowing adjustment appropriate to the context. Your main driver may be to reach the narrowest set or it may be to catch anything which might possibly be relevant. The requirement may change as the case develops or be different for different collections. User input at this point is a function of skill, intellect and experience, to which the technology is a biddable servant.

This is enhanced by the next stage. As the resulting selection is reviewed, discrepancies may emerge between Equivio's conclusions and those of the subsequent manual review. These discrepancies may go in either direction - documents given a high priority may prove in fact to be irrelevant, and irrelevant documents may have been marked as if they were important. The result of this stage can be fed back into the system allowing an iterative fine-tuning of the results.

Discrepancies between individual reviewers can be tracked down. These may indicate one of three things - that some reviewers are better than others, that there is a genuine conflict of views which, if not trapped, may lead to inconsistent coding, or that the user guidelines should be adjusted to take account of systematic errors.

#### WHAT HAPPENS NEXT?

The output of this phase is two bodies of documents, those determined to be relevant and those which, provisionally at least, are to be left on one side. The latter are still available, and the fact that they are ranked as having low priority may be re-opened;



this may be because later developments warrant a fresh set of decisions or because, as Judge Brown suggested in the passage quoted above, a deliberate decision is made to deal only with certain issues or categories of documents first in the hope that it proves unnecessary to come back to them.

In some cases, this stage is the end of the story – the upshot of this exercise may be the discovery of documents which can be used to persuade opponents to settle, or which show the clients that success is uncertain and the prospective outlay too great. Equivio>Relevance will have served its purpose if any of these outcomes becomes obvious.

If the action is to proceed, the relevance rankings can be exported into any document review system and become part of what is known about the documents for further review. This means that future searches can include the Equivio relevance ranking. This portability is important, because it means that the decision to use the power of Equivio>Relevance does not commit the lawyers to any particular application (or, indeed, to any). This point is well illustrated by Allen & Overy's approach which was the subject of a recent article in The Metropolitan Corporate Counsel headed Prioritization: Saving E-Discovery Costs While Speeding Review<sup>3</sup>. It is worth reading anyway for its description of how Allen & Overy use Epiq Systems' IQ Review (Epiq's integration of Equivio>Relevance) to save time and money for its clients, but the point in this context is that Allen & Overy can take what has been processed, analyzed and prioritised in IQ Review into their in-house application. That happens to be FTI's Ringtail, but almost any review application can take the Equivio prioritisation data in the same way.

## SUMMARY

Like Equivio's other applications – their near-duplication and e-mail thread technology for example – Equivio>Relevance is relatively easily understood in terms of what lawyers and their clients need. It takes no great technical knowledge to see that time and money is to be saved by any of them – who can argue with the idea that grouping all similar or related documents together under one eye at one time is not only a time-saver but leads to consistency?

Equivio>Relevance adds another dimension to this. If "relevance" appears to be a more subjective component than near-duplicates or e-mail threads, it is no less grounded in science. The lawyer is not, however, required to take much on trust. The system allows for both audit and cross-checking and it is possible to show how a given result was arrived at.

---

<sup>3</sup> <http://www.metrocorpcounsel.com/current.php?artType=view&EntryNo=10282>



The ability to make an early assessment of volumes is critical and will become more so for two reasons. One of these is the UK courts (and others) are increasingly wanting to manage cases, something which will become more important with the proposed new Practice Direction and E-Disclosure Questionnaire; that management requires an assessment of what parties have got, what is important, and what the time and costs implications are; that dictates, or at least influences, what the scope is of the subsequent disclosure exercise. Equivio>Relevance is designed to enable this.

The other reason is perhaps more important – clients will only litigate if they can get an early idea of both likely outcome and projected costs, will instruct only those who can provide this, and will increasingly expect their lawyers to share the risk with them. An application which does what is described above and which, as a bonus, leaves the client free to choose any review application, is a powerful tool in the hands of lawyers or, indeed, of the clients themselves.

Contact Equivio by e-mail to Warwick Sharp [warwick.sharp@equivio.com](mailto:warwick.sharp@equivio.com) or [info@equivio.com](mailto:info@equivio.com)



## ABOUT CHRIS DALE

© Chris Dale 2009

T: +44 1865 463033

M: +44 777 580640

Email: [cjdale@ntlworld.com](mailto:cjdale@ntlworld.com)

Web: [www.chrisdalelawyersupport.co.uk](http://www.chrisdalelawyersupport.co.uk)

Blog: [www.chrisdale.wordpress.com](http://www.chrisdale.wordpress.com)

## ABOUT EQUIVIO

Equivio develops text analysis software for e-discovery. Users include the DoJ, the FTC, KPMG, Deloitte, plus hundreds of law firms and corporations. Equivio offers Zoom, an integrated web platform for analytics and predictive coding. Zoom organizes collections of documents in meaningful ways. So you can zoom right in and find out what's interesting, notable and unique. Request a demo at [info@equivio.com](mailto:info@equivio.com) or visit us at [www.equivio.com](http://www.equivio.com).

**Zoom in. Find out.**