Neutral citation [2023] CAT 63

IN THE COMPETITION APPEAL TRIBUNAL

Case No: 1404/7/7/21

Salisbury Square House
8 Salisbury Square
London EC4Y 8AP

19 October 2023

Before:

SIR MARCUS SMITH
(President)
EAMONN DORAN
PROFESSOR ANTHONY NEUBERGER

Sitting as a Tribunal in England and Wales

BETWEEN

DAVID COURTNEY BOYLE

Class Representative

-and-

(1) GOVIA THAMESLINK RAILWAY LIMITED
(2) THE GO-AHEAD GROUP LIMITED
(3) KEOLIS (UK) LIMITED

Defendants

-and-

SECRETARY OF STATE FOR TRANSPORT

Intervener

Heard at Salisbury Square House on 12 October 2023

RULING (MODALITY OF TRIAL)

APPEARANCES

Mr Charles Hollander KC and Mr David Went (instructed by Maitland Walker LLP) appeared on behalf of the Class Representative, Mr Boyle.

Mr Paul Harris KC and Ms Anneliese Blackwood (instructed by Freshfields Bruckhaus Deringer LLP) appeared on behalf of Govia Thameslink Railway Limited, The Go-Ahead Group Limited, and Keolis (UK) Limited.

Ms Anneli Howard KC and Mr Nicolas Gibson (instructed by Linklaters LLP) appeared on behalf of the Secretary of State for Transport.

1. By a collective proceedings order made on 5 October 2022 (the **CPO Order**), the above-named proceedings were certified as collective proceedings and Mr Boyle authorised to act as class representative in those proceedings. The ruling setting out the Tribunal's reasons for making the CPO Order is dated 25 July 2022 and is cited under neutral citation number [2022] CAT 35 (the **First Ruling**).

2. The application for a collective proceedings order was – amongst other things – supported by four expert reports from a Mr James Harvey of Economic Insight Limited (respectively, **Harvey 1**, **Harvey 2**, **Harvey 3** and **Harvey 4**). The extent to which we relied upon these reports is set out in the First Ruling, which we do not repeat here.

3. On 14 October 2022, a case management conference took place, at which the structure of the trial of the collective proceedings was debated, and the Tribunal handed down a ruling (dated 14 October 2022 – the **Second Ruling**, [2022] CAT 46) directing a split trial, with questions of quantum hived off to **Trial 2**, all other questions being dealt with at a **Trial 1**: Second Ruling at [32]. Trial 1 – ironically, given the date of <u>this</u> Ruling – was listed for a four-week hearing "after the summer vacation in 2023".

4. For reasons which are articulated in our **Third Ruling**, [2023] CAT 19, it was necessary for the Class Representative to replace Mr Harvey as the expert instructed for the class in the collective proceedings. The new expert was a Dr Peter Davis of the Brattle Group. This change in expert inevitably meant that Trial 1 had to be adjourned. Dr Davis obviously had to acquaint himself with the case, and articulate his support for the collective proceedings, and for an amended collective proceedings claim form to be filed (the **Revised Claim Form**). That has occurred, and the Revised Claim Form is supported by Dr Davis in a number of reports (**Davis 1**, and so on). Whilst these reports draw on the work of Mr Harvey, it must be accepted that (because Dr Davis must give his own views as expert) the work of Mr Harvey has already faded into the background. Quite how much of Mr Harvey's work will be useable in the future is a matter of speculation.

5. The Third Ruling anticipated a yet further case management conference, at which any objections to the Revised Claim Form from the Defendants could be heard, and where directions (if so advised) to trial could be considered. This case management conference took place on 12 October 2023 and this Ruling (the **Fourth Ruling**) sets out how these collective proceedings are to proceed.

6. It is clear law that if collective proceedings are (i) arguable and (ii) triable they should proceed to trial. There is no elevated "merits" threshold, as the Supreme Court has made very clear in *Mastercard Inc v Merricks*, [2020] UKSC 51. The question of whether proceedings are arguable is one well-known in civil procedure law, and the "remedy" of strike-out is a well-known tool enabling a court, either of its own motion or on application, to avoid an expensive and unnecessary trial by killing proceedings off at an interlocutory stage.

7. The question of triability is much more nuanced. No case should be untriable, if it is arguable. Triability is all about case management and not about killing off proceedings. However, because competition claims in general, and collective proceedings (which are, by definition, competition claims) in particular, are difficult to try, a body of case law has evolved regarding the management of these claims. The process began in *Merricks* (see [135] and [150]) and was further evolved in: *London & South Eastern Railway Ltd v Gutmann*, [2022] EWCA Civ 1077 at [52]*ff*; *MOL (Europe Africa) Ltd v McLaren Class Representative Ltd*, [2022] EWCA Civ 1701 at [44]*ff*; *Gormsen v Meta Platforms Inc*, [2023] CAT 10; *Mark McLaren Class Representative Ltd v MOL (Europe Africa) Ltd*, [2023] CAT 25; and *UK Trucks Claim Ltd v Stellantis NV*, [2023] EWCA Civ 875. As can be seen from this list, which is not comprehensive, what has come to be termed the **Microsoft Pro-Sys Test** – from the Canadian decision *Pro-Sys Consultants Ltd v Microsoft Corp*, [2013] SCC 67 – arises in every application to certify collective proceedings.

8. Unsurprisingly, it arose at this case management conference, even though the proceedings had already been certified: see [1] above. It is right that it did: the necessary substitution of Dr Davis for Mr Harvey so early on in the proceedings made this inevitable. Before we turn to the specifics of this case, and without in

any way seeking to recap what has already been said in earlier cases about the Microsoft Pro-Sys Test, we make the following points:

(1) The Microsoft Pro-Sys Test is not intended as a barrier to trial, but as a case management tool to ensure an effective trial of the issues can take place. It will only be in the most extreme case, where a class representative discloses after repeated efforts a claim that simply cannot be tried, that the Tribunal may be driven to striking out a claim. The Microsoft Pro-Sys Test is not, and never has been, an *ersatz* test on the merits.

(2) It will always be necessary for a class representative, at the certification stage, to have an understanding of how the claim, if certified, is to progress. The extent of the detail that should be vouchsafed will vary from case-to-case, and the extent to which the Tribunal will probe will be acutely context sensitive. These are case management decisions where the Tribunal's margin of appreciation must necessarily be great. In *Stellantis* at [102], the Chancellor of the High Court rightly stated:

> "At the certification stage, the [Tribunal] has in each case to determine what level of detail it requires from the parties and their experts and, as Green LJ said in argument, that implies that the [Tribunal] has a broad margin of discretion in relation to certification with which this Court should not interfere unless a clear error of law is identified. In my judgment, it is not for the [class representative] to produce an expert methodology which addresses every conceivable issue or defence which the defendants say they will or may run. To go down that route would be to encourage a plethora of expert evidence addressing every conceivable argument that might be raised, and a long drawn out and expensive certification process as in the United States. It is important that the [Tribunal] and this Court discourage that approach. As Lord Briggs JSC made clear in *Merricks SC* at [41], the [Microsoft ProSys Test] is not intended to be onerous. It sets a fairly low threshold and simply does not require the [class representative's] methodology to anticipate and address at the certification stage every point that might be raised in defence."

(3) It might be said – and fairly said – that already too much encouragement has been given in overloading what is intended as a straightforward test of triability, turning the Microsoft Pro-Sys Test into something coming close to a mini-trial. We agree with the Chancellor that such an approach is to be discouraged.

(4) Nor is it the case that the Microsoft Pro-Sys Test is necessarily "once-and-for-all". Plainly, it is necessary, at certification, for that test to be passed. But, as all practitioners in this Tribunal know, cases in the Tribunal are docketed to a panel, and that panel actively manages that case from filing through to trial or final hearing. Cases are not left to proceed towards trial unsupervised. The supervision accorded to a case will vary in intensity according to need, as determined by the Tribunal panel or Chair in the given case. Whether that supervision is labelled "case management" or a continued application of the Microsoft Pro-Sys Test is a sterile debate, which we eschew. This is precisely what the Court of Appeal said in *McLaren* at [45]:

> "The duty on the CAT as gatekeeper in collective proceedings is proactive as well as reactive. Once the CAT has decided to make a CPO that is not the end of the gatekeeper role. A CPO "… is neither the beginning or the end of measures whereby the CAT may case manage collective proceedings" (Merricks (ibid) paragraph [28]). A class representative might not have to overcome a very high hurdle to obtain a CPO but the CAT should nonetheless ensure that from the certification stage the case proceeds efficiently to trial. This role might well entail the CAT imposing substantial case management burdens on the parties at an early stage."

(5) It follows that where the Microsoft Pro-Sys Test is <u>failed</u>, the response of the Tribunal will be a nuanced one, dependent on the reason for the failure. It may be that the issue is one readily capable of being fixed, in which case the Tribunal will identify with the parties <u>what</u> needs fixing, and the class representative will proceed to fix the problem. In such a case, it is to be expected that a further hearing will not be necessary, and that the manner in which the case progresses can either be agreed or else resolved on the papers. On the other hand, it may be that the methodology is so deficient that the class representative must be sent off to try again. That has only occurred once, to our knowledge, in *Gormsen*.

9. With this very broad guidance in mind, we turn to the present case:

(1) As we have noted, this case has <u>already</u> been certified as a collective action, and this case management conference has only arisen because of the unfortunate change of expert immediately after certification. This is a rare case, and we would hope not one that will be repeated.

(2) It was, we consider, appropriate to permit the Defendants to take what points they wished in relation to the Revised Claim Form and the post-certification reports of Dr Davis. Otherwise, there was an inevitable risk of the case proceeding without further scrutiny, having been certified on the basis of superseded evidence, namely the reports of Mr Harvey.

(3) In a document entitled "Defendants' Applications" (the **Defendants' Applications Document**), the Defendants articulated a number of issues in relation to the Revised Claim Form. None of these – as Mr Harris, KC, for the Defendants readily accepted – amounted to a barrier to the case proceeding, either in terms of a strike out or in terms of a comprehensive failure of the Microsoft Pro-Sys Test. In these circumstances – whatever the merits of these points (and we say nothing on this score) – the focus needed to be on driving the action forward, rather than spending at least two days resolving a series of points having little (if anything) to do with triability.

(4) Accordingly, we indicated to the parties at the beginning of the hearing a provisional intention to proceed to manage the case to trial – precisely as had been done in our Second Ruling, but recognising that far more work had been done (in the form of Dr Davis' reports) since that Second Ruling had been handed down. It seemed to us pointless to consider matters not going to case management unless they were so fundamental to the case as to preclude an effective trial. Since not even the parties articulating these points – the Defendants – suggested that their points were so fundamental, and since this represented our view also, it is self-evident that managing the case to trial had to be the first order of business.

(5) The question, then, was how to take the case on to trial. Mr Hollander, KC, for the Class Representative, advocated for a "traditional" procedure: disclosure, factual witness statements, expert reports, trial. We are in no doubt that this "traditional" procedure will be inefficient (in terms of both time and money spent) and highly unlikely to produce either a manageable trial or a manageable process leading up to trial. It

is important to understand why this is so in the case of collective proceedings. Collective proceedings are unique in this jurisdiction in that loss arising from a claim for infringement of the Chapter I or Chapter II prohibition is established not on an individual basis but on a class-wide basis: *Michael O'Higgins FX Class Representative Ltd v Barclays Bank plc*, [2022] CAT 16 at [226]. Individual loss needs neither to be pleaded nor proved; it is enough to plead and prove class wide loss. This places the expert economist "centre stage": the effort and cost of demonstrating a class wide loss based upon a general and class wide theory of harm will generally be far less than seeking to establish and aggregate a series of individuated claims and losses. Naturally, such an approach is more consistent with the collective action regime in any event.

(6) It is for this reason that the Microsoft Pro-Sys Test is generally approached through the production of an expert report from an economist. As the Chancellor noted in *Stellantis* (see [8(2)] above) what is required is – contained in a single and brief expert report – a short articulation of the theory of harm said to arise out of the infringement, and how the expert proposes, on a class wide basis, to demonstrate in a court of law how that infringement was causative of loss, and how that loss will be quantified. It may be that the defendant chooses to challenge the feasibility of this approach; and if that is done, then the defendant will generally have to produce their own expert report in response (for this is a matter of expert opinion, not legal submission). It may be that reply evidence will, in such a case, be necessary, but that will be for the Tribunal in any given case to make that determination. The Microsoft Pro-Sys Test is a low order test for a blueprint to trial: no more, and no less.

(7) It follows from this that, since the author of the blueprint is the expert economist, it is the expert economist who is likely to be best placed to explain how it is envisaged the evidence will be developed to trial. In this regard, the Tribunal will pay close attention to what the expert says they need, and what is the most efficient and proportionate way to

proceed to trial. We doubt very much that a traditional disclosure exercise is efficient or proportionate in the case of collective actions. Traditional disclosure – lists of issues, where the standard and scope of search for relevant documents is articulated, drawn up by reference to the pleadings – is emphatically not what the expert economist needs. The expert economist generally needs data and information, not reams of documents (whether paper or electronic) which then must be sifted and analysed and turned into usable data and information. Suppose, in a hypothetical case, the prices and volumes of widgets sold over time are important. It would be undesirable for there to be a search for invoices showing price and volume over time, and for those invoices to be produced to the receiving party by the disclosing party, for the experts on each side then independently to produce a schedule of widget prices and volumes. What should happen is for the expert on the claimant's side to articulate the need for the data or information, and for that data or information to be produced (in schedule form) by the expert on the defendant's side. It may be that a subsequent "audit" of the data is called for – again, a matter for the Tribunal – but (given the duties of each expert to the Tribunal) data produced ought, in the first instance, to be regarded as reliable without more, particularly if an explanation as to how it has been compiled is given.

(8) Dr Davis – doubtless having been told how the ordinary process works – suggested a course of action for disclosure that is inconsistent with this approach and (with all due respect) completely indefensible in the context of these proceedings. We append to this ruling some extracts from Davis 2 indicating the keyword searches that Dr Davis would want the Defendants to undertake. These paragraphs contain more than a hint of the lawyer-driven process inapposite for these proceedings. We do not intend to criticise Dr Davis: rather we want to encourage him – and other experts in similar cases – not to define search terms or search processes for documents, but to articulate the information and data needed to bring the case to trial.

(9) Given that we consider the MicroSoft Pro-Sys Test to be satisfied, the question is what needs to be produced by the Class Representative to take these proceedings to trial. We consider that Dr Davis should, in these circumstances, articulate in full the Class Representative's case, and that he should do so by no later than 31 July 2024. We regard this as a generous date, and ideally the job can be done well before then. We will leave it to the parties to frame an appropriate order, but what we envisage in general terms is this:

(i) We see no point, at this stage, in requiring the Defendants to respond to what is merely a statement of methodology or blueprint to trial (albeit one contained in a regrettably large number of expert reports). That is to fight a "phoney war" incurring cost to no benefit. The Class Representative, through Dr Davis, must first produce a body of evidence which, assuming no pushback or evidence in response from the Defendants subsequently, is sufficient to persuade a probing and appropriately sceptical Tribunal of the merits of the claim, including as to quantum, so that the Tribunal could make a final order in a given, ascertainable, quantum. Of course, we appreciate that these proceedings will be defended, and the Class Representative's path will not be so straightforward. We are seeking to articulate, however, what the Class Representative must produce by 31 July 2024, in anticipation of a response from the Defendants.

(ii) Of course, Dr Davis will require significant data and information to do this. We make clear that if Dr Davis frames appropriately clearly and narrowly the information and/or data he seeks from both the Defendants <u>and</u> the Intervener, and explains briefly why he wants it, the Tribunal will expect the Defendants and the Intervener to co-operate through their own experts. If this process does not work, the Tribunal's Chair will make himself available on short notice by way of remote hearing to deal with any disagreement. This offer is made in the expectation that it

can and will be used. Given that these proceedings have already been certified as collective proceedings, the Tribunal will not be sympathetic to requests for more time in the summer of 2024 if requests for data or information could and should have been made sooner.

(iii) We expect Dr Davis to need more than just information or data from the Defendants and the Intervener. There may be a need for third party disclosure; or expert evidence from a discipline other than the economic. We cannot anticipate, and it may be that a formal application will have to be made in such cases. But the Chair of the Tribunal again stands ready to deal with such matters remotely and on short notice.

(iv) The Second Ruling directed a split trial. Whilst that split may still be appropriate for trial, we want to be clear that the material that must be produced by the Class Representative should cover all issues, not just those issues relating to Trial 1.

(v) It may be that this material would benefit from a short "position statement" drawn by the Class Representative's legal team, pulling together the strands in what will, doubtless, be a voluminous body of evidence. We leave that to the discretion and good judgement of the Class Representative.

10. This paves the way for a case management conference on the far side of the 2024 long vacation at which the Defendants will articulate the directions they need for trial, and the Intervener will state their own position. A trial at some point in 2025 should then be possible. We make clear that we expect the Defendants and the Intervener to review the Class Representative's work over the long vacation and to have considered – and ideally begun work on – their responses.

11. We should also make clear that if the Class Representative's expert and the Defendants' expert can constructively engage before 31 July 2024, so as to

make Dr Davis' job easier, and potentially shorten the path to an effective trial, so much the better. But that is a course we cannot and do not compel.

12. It is theoretically possible that the Class Representative's evidence, when produced, discloses either a demurrable (strikeable) claim or else one that discloses an unmanageable trial. We regard these both as theoretical possibilities. No-one has suggested a strike-out; and on two occasions now (in our First Ruling and at this case management conference) we have considered the Microsoft Pro-Sys Test and found that it is met. It would therefore be quite extraordinary for this position to change. But it is a theoretical possibility, and we are not closing out the making of an application for summary disposal of these proceedings.

13. There was, before us, an application by the Defendants for the costs thrown away by them because of the change of expert, which we have referred to. The point made – and it is a fair one – is that even if the change of expert was not the "fault" of the Class Representative (and it was not) the Defendants should not bear the costs of that change. We adjourn that application to the 2024 post-vacation case management conference. We are reluctant to make any order as to costs in the context of an on-going process (see, for example, [4] above), such as the one which we have directed in this case.

14. This Ruling is unanimous.

Sir Marcus Smith President	Eamonn Doran	Professor Anthony Neuberger

Charles Dhanowa O.B.E., K.C. (*Hon*) Registrar	Date: 19 October 2023

d. There is of course a question of what the false negative error rate threshold (x%) should reasonably be. With a two-step process (whereby all documents are permanently determined as not relevant if they are identified by the search criteria), it would be appropriate to use a low false negative error rate threshold, keeping in mind that a false-negative error rate of 0.1% (0.01%) of 4 million documents would lead to 4,000 (400) relevant documents being incorrectly omitted.

391. It will make sense to use the random samples collected when evaluating the search terms as a part of the training data for any application of the CMML Model.

392. If it proves too hard to achieve a sufficiently low false negative error rate using the Defendants' manual first-stage 'search term' approach, the alternative would be to retain the full set of c. 4 million documents when developing and refining the CMML model.

D.2.a.ii Further targeted searches

393. I understand that previously the CR had indicated in the letter to Freshfields of 16 February 2023 that "improvements [are] required to ensure that the search is effective. It should be emphasised that Annex 1 contains an initial analysis and we may have further comments on the thematic searches following further review."[582]

394. As I have already mentioned, it is difficult for either me or the CR to design the 'right' search terms without access to the underlying pool of documents, as it is hard to evaluate the performance of keyword searches in the abstract. However, I do note that the most natural use of the Defendants' search term step is to seek to exclude documents that are definitely *not* relevant – that is, leaving documents that are possibly relevant for consideration as part of the model (otherwise such possibly relevant documents would be discarded a priori by not meeting the search criteria adopted – and, at the moment, I understand that no checks at all are performed on whether documents dropped in this way are dropped correctly or in error).

---

[582] **Exhibit PD-0203**, Letter from Maitland Walker to Freshfields Bruckhaus Deringer LLP dated February 16, 2023, *David Courtney Boyle v. Govia Thameslink Railway Limited, The Go-Ahead Group limited, Keolis (UK) Limited*, p. 3.

395. If my previous section's process of testing the keyword search as a part of the process cannot be carried out, then I would suggest that the search criteria are set to allow a broad selection of documents to pass to the second stage. For reference, Annex 1 to the letter to Freshfields of 16 February 2023 included a column called "Improved search term" which still included restrictive search terms (e.g. a search term could require multiple words to be present for a document to be identified as potentially relevant).[583] Below, I have split out such multi-part search terms: instead of requiring multiple words to be present in a document, a document can instead be identified by including a single relevant word below. I also added further search terms that could result in identifying further relevant documents. I would suggest including documents that would be captured by any of the below categories of search terms (with each not case sensitive) categories:

a. Relevant franchise/route: Franchise OR TSGN OR Southern OR Thameslink OR "Great Northern" OR GN OR "Gatwick Express" OR GX OR GEX OR "Gat* Exp*" OR Brighton OR BML OR

b. Fare amendment / ticket types

   i. fare* OR pric* OR charg* OR reserv* OR upgrad* OR oyst* OR PAYG* OR contactless* OR "fare inconsistenc*" OR "fare* harmonis*" OR "brand restric*" OR "simplify fare* structure*" OR "remove anomalies" OR "anomal*" OR "align" OR "ticket valid*" OR "acceptance" OR

   ii. "fares paper" OR "fare* setting" OR FSR OR "fare* meeting" OR

   iii. "dedicated far*" OR "dedicated pric*" OR "dedicated charg*" OR ("inter-availab*" OR "interavailab*") OR

   iv. "Any permitted" OR "single-brand" OR "dual-brand" OR

   v. (network AND (fare* OR pric* OR charg* OR reservation OR ticket)) OR "unlimited travel" OR

c. Costs and profits: cost* OR margin* OR profit* OR revenue* OR

d. Regulation

583 **Exhibit PD-0204**, Annex 1 to letter from Maitland Walker to Freshfields Bruckhaus Deringer LLP dated February 16, 2023, *David Courtney Boyle v. Govia Thameslink Railway Limited, The Go-Ahead Group limited, Keolis (UK) Limited* , p. 1.

    i. TSA OR (ticket* w/20 "settlement agreement") OR NRCOT OR NRCOC OR "Conditions of Travel" OR "Conditions of Carriage" OR
    ii. "Bid Fares Policy" OR BFP OR

e. Competition from other train companies or alternative modes of transport
    i. (compet* OR constrain* OR share OR "market power") w/20 (coach OR car* OR bus OR busses OR alternativ* OR bik* OR bicycle* OR cycl* OR walk* OR driv* OR taxi* OR cab* OR road OR train* OR rail*) OR
    ii. ((customer* OR passenger*) w/20 (choice OR prefer* OR behave* OR alternativ* OR response* OR switch*)) OR
    iii. ((commut* OR leisure OR business OR touris*) w/20 (market OR share OR segment OR customer* OR passenger*)) OR
    iv. ((passenger* OR customer*) w/20 (number* OR volume*)) OR
    v. (nature OR reason* OR purpose*) w/20 (journey* OR travel*) OR
    vi. experience OR congest* OR "servic* frequen*" OR interchang* OR
    vii. elasticit* of demand OR switch* OR
    viii. "barrier* to entry" OR "entry barrier*" OR

f. Penalty/Excess fares: "penalty fare" OR "excess fare" OR PF OR EF OR ((pric* OR fare*) AND differen*) OR

g. London Victoria prices: "plat* 13" OR "plat* 14" OR

h. Fare creation process incl. flows being removed/created: (new OR different OR remove* or amend* OR add OR addition OR added) w/20 (station OR stop OR city OR [include any stops that exist on the route between Brighton and London Victoria/London Bridge whether or not they are part of the TSGN routes, separating them using OR separators])" OR

i. Additional requests (as outlined below):
    i. "catchment area*" (for each of the stations on each of the potential flows) OR
    ii. timetable OR
    iii. WebTAG OR "value* of time" OR "time value" OR
    iv. GJC OR "Generalised Journey Costs" OR

v. "market power" OR "abuse of a dominant position" OR "abuse of a dominance" OR

vi. "National Rail Enquiries Fares Finder" OR

vii. "pricing tool" OR "pricing model" OR "pricing policy" OR

viii. "Passenger Demand Forecasting Handbook" OR PDFH OR

ix. "service levels" OR inefficiencies OR "loading factors" OR

x. ("customer complaint*" OR "customer satisfaction") AND pric* OR

xi. "customer survey*" OR

xii. (sale* OR purchas*) AND (business* OR self-employed) OR

xiii. passenger type/journey purpose: touris* OR commuter* OR "business user*" OR "staff" OR "shift work*" OR leisure OR

xiv. by demand patterns during the day: peak OR off-peak

396. I hope that following discussions with the Defendants' expert, the targeted search terms could be further refined but, to evaluate the search process adopted, there is no good substitute for hard data on the false negative error rates achieved by a given set of search terms.

D.2.a.iii Application of CMML (Logistic regression)

397. The question that needs to be resolved is whether the Defendants' approach to training the CMML (logistic regression) model is fit-for-purpose. At the moment, on the basis of the information so far provided by the Defendants, it is not possible to properly assess whether that is the case. In particular:

a. it is not clear how well the model predicts relevant documents either in sample or out of sample;

b. the Defendants do not provide any sense of what features the model determined were predictive of document relevance; and

c. the Defendants do not provide sufficient information as to the CMML model's predictive performance, including in particular false-positive and false-negative error rates.