UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

RIO TINTO PLC,		
Plaintiff,	:	14 Civ. 3042 (RMB)(AJP)
-against-	:	OPINION & ORDER
VALE S.A., et al.,	:	
Defendants.	:	

Predictive Coding a.k.a. Computer Assisted Review a.k.a. Technology Assisted Review (TAR) - Da Silva Moore Revisited

ANDREW J. PECK, United States Magistrate Judge:

It has been three years since my February 24, 2012 decision in Da Silva Moore v. Publicis Groupe & MSL Grp., 287 F.R.D. 182 (S.D.N.Y. 2012) (Peck, M.J.), aff'd, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012). In Da Silva Moore, I stated:

> This judicial opinion now recognizes that computer-assisted review [i.e., TAR] is an acceptable way to search for relevant ESI in appropriate cases.

Da Silva Moore, 287 F.R.D. at 183. I note that while the terms predictive coding and computer assisted review still are used, technology assisted review, or TAR, now seems to be the preferred term of art. I concluded the Da Silva Moore opinion by stating:

> This Opinion appears to be the first in which a Court has approved of the use of computer-assisted review. That does not mean computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that utilize computer-assisted review. Nor does this Opinion endorse any vendor . . . , nor any particular computer-assisted review tool. What the Bar should take away from this Opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the

> producing party (or both parties) significant amounts of legal fees in document review. Counsel no longer have to worry about being the "first" or "guinea pig" for judicial acceptance of computer-assisted review. As with keywords or any other technological solution to ediscovery, counsel must design an appropriate process, including use of available technology, with appropriate quality control testing, to review and produce relevant ESI while adhering to Rule 1 and Rule 26(b)(2)(C) proportionality. Computer-assisted review now can be considered judicially-approved for use in appropriate cases.

Da Silva Moore v. Publicis Groupe & MSL Grp., 287 F.R.D. at 193 (emphasis added).

In the three years since Da Silva Moore, the case law has developed to the point that it is now black letter law that where the producing party wants to utilize TAR for document review, courts will permit it.[1/] The recent Tax Court decision in Dynamo Holdings Ltd. P'Ship v. Comm'r of Internal Revenue, 143 T.C. 9, 2014 WL 4636526 (T. C. Sept. 17, 2014), is instructive. The Tax Court's response to being asked to approve the use of TAR was that courts leave it to the parties to decide how best to respond to discovery requests:

> [T]he Court is not normally in the business of dictating to parties the process that they should use when responding to discovery. If our focus were on paper discovery, we would not (for example) be dictating to a party the manner in which it should review documents for responsiveness or privilege, such as whether that review should be done by a paralegal, a junior attorney, or a senior attorney. Yet that is, in essence, what the parties are asking the Court to consider – whether document review should be done by humans or with the assistance of computers. Respondent fears an incomplete response to his discovery. If respondent believes that the ultimate discovery response is incomplete and can support that belief, he can file another motion to compel at that time.

1/ In contrast, where the requesting party has sought to force the producing party to use TAR, the courts have refused. See, e.g, In re Biomet M2A Magnum Hip Implant Prods. Liab. Litg., No. 3:12-MD-2391, 2013 WL 1729682 & 2013 WL 6405156 (N.D. Ind. Apr. 18 & Aug. 21, 2013); Kleen Prods. LLC v. Packaging Corp. of Am., 10 C 5711, 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012). The Court notes, however, that in these cases, the producing parties had spent over $1 million using keyword search (in Kleen) or keyword culling followed by TAR (in Biomet), so it is not clear what a court might do if the issue were raised before the producing party had spent any money on document review.

Dynamo Holdings Ltd. P'Ship v. Comm'r of Internal Revenue, 2014 WL 4636526 at *3.[2/] Reaching the merits, the Tax Court "disagree[d]" with the IRS's position that TAR was an "'unproven technology,'" holding: "In fact, we understand that the technology industry now considers predictive coding to be widely accepted for limiting e-discovery to relevant documents and effecting discovery of ESI without an undue burden." Dynamo Holdings Ltd. P'Ship v. Comm'r of Internal Revenue, 2014 WL 4636526 at *5 (citing articles and cases, including Da Silva Moore). For other judicial decisions approving the producing party's use of TAR, see, e.g., Green v. Am. Modern Home Ins. Co., No. 14-CV-04074, 2014 WL 6668422 at *1 (W.D. Ark. Nov. 24, 2014); Aurora Coop. Elevator Co. v. Aventine Renewable Energy - Aurora W. LLC, No. 12 Civ. 0230, Dkt. No. 147 (D. Neb. Mar. 10, 2014); Edwards v. Nat'l Milk Producers Fed'n, No. 11 Civ. 4766, Dkt. No. 154: Joint Stip. & Order (N.D. Cal. Apr. 16, 2013); Bridgestone Am., Inc. v. IBM Corp., No. 13-1196, 2014 WL 4923014 (M.D. Tenn. July 22, 2014)[3/]; Fed. Hous. Fin. Agency v. HSBC N.A. Holdings, Inc., 11 Civ. 6189, 2014 WL 584300 at *3 (S.D.N.Y. Feb. 14, 2014); EORHB, Inc. v. HOA Holdings LLC,

---

[2/] That view is consistent with Sedona Principle 6: "Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information." The Sedona Principles: Second Edition, Best Practices Recommendations & Principles for Addressing Electronic Document Production, Principle 6 (available at www.TheSedonaConference.org).

[3/] In Bridgestone, Magistrate Judge Brown allowed Bridgestone to "switch horses in midstream" from a keyword and manual review stipulation to keywords followed by TAR (similar to Biomet). Compare Progressive Cas. Ins. Co. v. Delaney, No. 11-CV-00678, 2014 WL 3563467 (D. Nev. July 18, 2014), where because the parties had stipulated to a keyword then manual review protocol, Magistrate Judge Leen would not allow Progressive to use TAR only on the positive keyword hits (Progressive was unwilling to start over and use TAR without first using keyword culling). While holding Progressive to the protocol the parties had negotiated, Judge Leen agreed that TAR "is far more accurate" than human review or keywords, and "the Court would not hesitate to approve a transparent, mutually agreed upon ESI protocol" that used TAR. Progressive Cas. Ins. Co. v. Delaney, 2014 WL 3563467 at *8-9.

No. Civ. A. 7409, 2013 WL 1960621 (Del. Ch. May 6, 2013)[4]; In re Actos (Pioglitazone) Prods. Liab. Litig, No. 6:11-MD-2299, 2012 WL 7861249 (W.D. La. July 27, 2012) (Stip. & Case Mgmt. Order); Global Aerospace Inc. v. Landow Aviation LP, No. CL 61040, 2012 WL 1431215 (Va. Cir. Ct. Apr. 23, 2012).

One TAR issue that remains open is how transparent and cooperative the parties need to be with respect to the seed or training set(s). In Da Silva Moore, defendant MSL volunteered such transparency, confirming that "'[a]ll of the documents that are reviewed as a function of the seed set, whether [they] are ultimately coded relevant or irrelevant, aside from privilege, will be turned over to' plaintiffs." Da Silva Moore, 287 F.R.D. at 187; see also id. at 192 ("This Court highly recommends that counsel in future cases be willing to at least discuss, if not agree to, such transparency in the computer-assisted review process."). In In re Actos, 2012 WL 7861249 at *4, the parties' protocol had "experts" from each side simultaneously reviewing and coding the seed set. In Bridgestone, 2014 WL 4923014 at *1, the plaintiff had offered to provide the responsive and non-responsive seed set documents to IBM and Judge Brown stated that he "expects full openness in the matter."[5] And in Fed. Hous. Fin. Agency v. HSBC, in a decision from the bench on July 24, 2012,

---

[4] In EORHB (known as the Hooters case), at the end of an October 15, 2012 hearing on motions, Vice Chancellor Laster sua sponte ordered the parties to use TAR, stating: "This seems to me to be an ideal non-expedited case in which the parties would benefit from using predictive coding. I would like you all, if you do not want to use predictive coding, to show cause why this is not a case where predictive coding is the way to go." EORHB, 10/15/12 Conf. Tr. at 66. The parties subsequently agreed that defendant would use TAR but plaintiff would not (based on their representation that there would not be sufficient cost savings from TAR because of their low volume of ESI), and Vice Chancellor Laster approved their approaches. EORHB, Inc. v. HOA Holdings LLC, 2013 WL 1960621 at *1.

[5] In January 2015, the parties informed Judge Brown that "on review some of the [seed set] documents listed as nonresponsive were, in fact, responsive." (Bridgestone, No. 13-cv-01196, Dkt. No. 108: 2/5/15 Order at 4.) Judge Brown"remind[ed] both parties that (continued...)

Judge Cote required transparency and cooperation, including giving the plaintiff full access to the seed set's responsive and non-responsive documents (except privilged). In contrast, in the second Biomet decision, 2013 WL 6405156 at *1, 2, Judge Miller said that he could find no authority that would allow him to require Biomet to share seed set documents with plaintiffs' counsel, but suggested that Biomet rethink its opposition to doing so. Thus, where the parties do not agree to transparency, the decisions are split and the debate in the discovery literature is robust. See, e.g., John M. Facciola & Philip J. Favro, Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled To Work Product Protection, 8 Fed. Cts. L. Rev. 1 (2015).

If the TAR methodology uses "continuous active learning" (CAL) (as opposed to simple passive learning (SPL) or simple active learning (SAL)), the contents of the seed set is much less significant. See generally Gordon V. Cormack & Maura R. Grossman, Evaluation of Machine Learning Protocols for Technology-Assisted Review in Electronic Discovery, in Proceedings of the 37th Int'l ACM SIGIR Conf. on Research & Dev. in Info. Retrieval (SIGIR '14), at 153-62 (ACM New York, N.Y. 2014), http://dx.doi.org/10.1145/2600428.2609601; Maura R. Grossman & Gordon V. Cormack, Comments On "The Implications of Rule 26(g) on the Use of Technology-Assisted Review," 7 Fed. Cts. L. Rev. 285, 298 (2014) ("Disclosure of the seed or training set offers false comfort to the requesting party . . . ").

5/ (...continued)
to the extent they use predictive coding, he expects full transparency in how the predictive coding is established and used." (Id.)

In any event, while I generally believe in cooperation,[6/] requesting parties can insure that training and review was done appropriately by other means, such as statistical estimation of recall at the conclusion of the review as well as by whether there are gaps in the production, and quality control review of samples from the documents categorized as non-responsive. See generally Grossman & Cormack, Comments, supra, 7 Fed. Cts. L. Rev. at 301-12.

The Court, however, need not rule on the need for seed set transparency in this case, because the parties agreed to a protocol that discloses all non-privileged documents in the control sets. (Attached Protocol, ¶¶ 4(b)-(c).) One point must be stressed – it is inappropriate to hold TAR to a higher standard than keywords or manual review. Doing so discourages parties from using TAR for fear of spending more in motion practice than the savings from using TAR for review.

The Court has written this Opinion, rather than merely signing the parties' stipulated TAR protocol, because of the interest within the ediscovery community about TAR cases and protocols. The Court is approving the parties' TAR protocol, but notes that it was the result of the parties' agreement, not Court order. And as in Da Silva Moore, the Court's approval "does not mean . . . that the exact ESI protocol approved here will be appropriate in all [or any] future cases that utilize [TAR]. Nor does this Opinion endorse any vendor . . . , nor any particular [TAR] tool." Da Silva Moore v. Publicis Groupe & MSL Grp., 287 F.R.D. at 193. Indeed, the Court informed counsel that their stipulated protocol was somewhat vague and generic, which is why they felt the need to accompany it with a cover letter to "provide the Court with a brief summary of those [TAR]

6/ See, e.g. Da Silva Moore, 287 F.R.D. at 192 ("This Court was one of the early signatories to The Sedona Conference Cooperation Proclamation, and has stated that 'the best solution in the entire area of electronic discovery is cooperation among counsel. This Court strongly endorses The Sedona Conference Cooperation Proclamation (available at www.The SedonaConference.org).'" (quoting William A. Gross Constr. Assoc., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 136 (S.D.N.Y 2009)(Peck, M.J.))).

processes as provided by the parties' respective vendors." (Dkt. No. 181: 2/13/15 Letter to Court, at 1.) With that caveat, for whatever benefit it may be to subsequent cases, the parties' cover letter (Dkt. No. 181) and approved protocol (Dkt. No. 181-1) are attached to this Opinion as an Appendix.

SO ORDERED.

Dated: New York, New York
March 2, 2015

Andrew J. Peck
United States Magistrate Judge

Copies **by ECF** to: All Counsel
Judge Berman

quinn emanuel trial lawyers | washington, dc
777 Sixth Street NW, 11th Floor, Washington, District of Columbia 20001-3706 | TEL (202) 538-8000 | FAX (202) 538-8100

WRITER'S DIRECT DIAL NO.
(202) 538-8166

WRITER'S INTERNET ADDRESS
mikelyle@quinnemanuel.com

February 13, 2015

Hon. Andrew J. Peck
United States Magistrate Judge, Southern
District of New York
Daniel Patrick Moynihan Courthouse
500 Pearl Street, Courtroom 20D
New York, New York 10007

Re: Rio Tinto v. Vale et al, Civil Action No. 14-cv-3042 (RMB) (S.D.N.Y.)

Dear Judge Peck:

Plaintiff Rio Tinto plc ("Plaintiff") and Defendant Vale S.A. ("Vale") write jointly to provide the Court with a revised proposed Predictive Coding Protocol. Pursuant to the Court's February 6, 2015 Order, the parties have continued to meet and confer with respect to certain suggested revisions to the proposed protocol and believe that we have resolved the Court's concerns with respect to various aspects of the protocol. In addition, the parties respective vendors have reviewed the proposed protocol and believe it is consistent with and can be applied to the parties respective predictive coding processes. While the proposed Predictive Coding Protocol requires the parties to exchange details about their respective predictive coding process, we also take this opportunity to provide the Court with a brief summary of those processes as provided by the parties' respective vendors.

Rio Tinto

As discussed at the parties February 6, 2015 hearing, Rio Tinto and its vendor, Precision Discovery, will be using Relativity Assisted Review ("RAR"). Using RAR, Rio Tinto will first create a Control Set by randomly sampling from the document universe. The legal subject matter expert will then review the control set for responsiveness. The Control Set is not used to train the set, it is only for validation. Following the control set review, Precision Discovery will note the Control Set's percentage of responsive documents. This number will serve as a benchmark throughout the project. It will also affect the size of the seed set. We will then

perform a seed set training review. The seed set may be created using random sampling, keyword searching, and/or conceptual ranking. After the first round of seed set review, Precision Discovery will use the coding from the seed set to categorize the document universe, including the control set. After categorization, the Control Set will have both computer coding and human coding. In comparing the human coding with computer coding, Precision Discovery will check for coding volatility, precision, recall, and F1 metrics to track progress and would expect to see an increase in precision, recall, and F1 from round to round. Volatility is also expected to decrease from round to round. These last two steps (the seed set training round and check for coding volatility, precision, recall and F1 metrics) will be repeated until Quinn Emanuel, in consultation with Precision Discovery, is satisfied with the metrics achieved. When the metrics of precision, recall and F1 are achieved, we will perform a last step of validation. For this final step, documents coded as non-responsive will be sampled. If the level of overturns within this sample is within the margin of error, the predictive coding project is complete. If it is above the margin of error, we will again repeat those same steps until the level of overturns is within the margin of error.

Vale

First, an initial Control Set will be created by drawing a statistically-valid random sample of documents from the review population. This will be used as the "gold standard" and should be coded by a member of the team who has a thorough understanding of the matter. Second, a Seed Set will be coded by a team of human reviewers. These documents can be selected at random or based on judgmental sampling. Third, Deloitte's Dynamic Review, which utilizes the LibLinear library as its document classification algorithm, will use the review team's coding to build a predictive model to categorize the rest of the documents in the Document Universe. Fourth, the predictive model will be used to assign a responsiveness score to the rest of the documents in the Document Universe. Fifth, the Control Set is used to determine the effectiveness of the model. Sixth, additional Training Sets are drawn from the Document Universe and reviewed. Documents in the Training Sets are generally drawn based on the responsiveness scores assigned by the model, but they can also be drawn at random or based on judgmental sampling. Seventh, in consultation with Deloitte, the legal team reviews the results of the model using precision and recall rates to make strategic decisions with respect to the unreviewed documents in the Document Universe. Depending on the results of this analysis, Steps 2 through 7 are repeated until precision and recall rates reach an appropriate level. Eighth, as a final validation, a randomly selected Validation Set will be pulled and reviewed to verify results are as predicted by the model.

****

The parties respectfully submit the attached proposed Predictive Coding Protocol for the Court's review.

Very truly yours,

/s/Michael Lyle
Michael Lyle

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

Rio Tinto plc, Plaintiff, -against- Vale S.A., Benjamin Steinmetz, BSG Resources Limited, VBG–Vale BSGR Limited aka BSG Resources (Guinea) Ltd. aka BSG Resources Guinée Ltd, BSG Resources Guinée SARL aka BSG Resources (Guinea) SARL aka VBG-Vale BSGR, Frederic Cilins, Mamadie Touré, and Mahmoud Thiam, Defendants.	14 Civ. 3042 (RMB) (AJP)

STIPULATION AND ORDER RE:
USE OF PREDICTIVE CODING IN DISCOVERY

WHEREAS, the Plaintiff Rio Tinto and the Defendant Vale (collectively the "Parties" and each a "Party") in the above captioned litigation ("Action") agree to the use of predictive coding for the search, review, and production of documents in this Action and to enter a stipulation ("Stipulation") to memorialize their agreement;

IT IS HEREBY STIPULATED AND AGREED, by and between the undersigned, as attorneys of record for the Parties, as follows:

1. Definitions

(a) "Precision" means the fraction of documents identified as likely responsive by the Predictive Coding Process that are in fact responsive.[1]

[1] Definitions of "Precision" and "Recall" are adapted from the "Grossman-Cormack Glossary of Technology-Assisted Review," 7 Fed. Cts. L. Rev. 1, 25, 27 (2013).

(b) "Predictive Coding Process" means the use by the Parties of Predictive Coding Software to categorize documents into those that are likely responsive and those that are likely non-responsive.

(c) "Predictive Coding Software" means the software a Party elects to use to perform the Predictive Coding Process.

(d) "Recall" means the fraction of responsive documents that are identified as likely responsive by the Predictive Coding Process.

(e) "Statistically Valid Sample" means a random sample of sufficient size and composition to permit statistical extrapolation with a margin of error of +/- 2% at the 95% confidence level.[2]

2. Scope of this Stipulation

(a) The procedures described in this Stipulation govern the use of predictive coding to assist in the production of documents by the Parties to this Action. In this Stipulation, predictive coding shall mean and refer to a process for selecting and ranking a collection of documents using a computerized system that incorporates the decisions that lawyers have made on a smaller set of documents and then applying those decisions to the remaining universe of documents.

(b) Nothing in this Stipulation shall prevent the Party responding to discovery (the "Responding Party") from using other search, review, or coding methodologies in

[2] The size of the sample will vary depending upon several factors, and shall be calculated using the formula $n = \frac{X^2 * N * P * (1-P)}{(ME^2 * (N-1)) + (X^2 * P * (1-P))}$, where, ME is the margin of error; X is the Confidence Level (1.96 for a 95% Confidence Level); P is judgment of richness, N is the population and n is sample size. Where richness is not reasonably estimable, 0.5 may be used. Based on a Confidence Level of 95%, richness of 0.5, a Population of 1,000,000, and a margin of error of 2%, the resulting sample size is 2,395 documents.

addition to, or in place of, predictive coding to help identify documents that are responsive to the document requests from the Party seeking discovery (the "Requesting Party").

3. Initial Disclosure of Information about Predictive Coding

(a) Prior to the commencement of any review using predictive coding, the Responding Party shall disclose to the Requesting Party in writing its intention to use predictive coding and the following information:

(i) The name, publisher, version number, and a description of the Predictive Coding Software and Predictive Coding Process;

(ii) The name and qualifications of the person who will oversee the implementation of the predictive coding process (the "Technical Expert");

(iii) A description of the documents to be subjected to predictive coding (the "Document Universe"), including:

(1) Custodian / Source[3];

(2) Data types (e.g., email, electronic documents, etc.);

(3) The number of documents in the Document Universe, in total and for each Custodian / Source;

(iv) The responsiveness categories into which the Document Universe is to be categorized (the "Responsiveness Categories").

(b) The Parties shall meet and confer to address any questions or disputes about the selection of the Predictive Coding Software, the Technical Expert, the Document Universe, and the Responsiveness Categories, and the Responding Party shall make its

[3] These terms have the definitions set forth in the ESI Protocol (Dkt. No. 82).

Technical Expert reasonably available to address questions about the technical operation of the Predictive Coding Software.

4. Predictive Coding Methodology

(a) Culling the Document Universe. If the Responding Party determines it to be reasonable and appropriate, the Responding Party may use search terms and other criteria (the "Culling Criteria") to reduce the volume of the Document Universe. If it does so, the Responding Party shall promptly:

(i) Disclose in writing to the Requesting Party the Culling Criteria used and the number of documents removed by the Culling Criteria (the "Excluded Documents");

(ii) Review a Statistically Valid Sample from the Excluded Documents, disclose the size of that sample set, and produce any responsive, non-privileged documents the Responding Party identifies;

(iii) Meet and confer with the Requesting Party, if requested, to address any questions or disputes about the reasonableness and appropriateness of the Culling Criteria.

(b) Control Set Review. To aid in the Predictive Coding Process, and to determine the prevalence of responsive information from the Document Universe, the Responding Party shall review a Statistically Valid Sample of documents from the Document Universe (the "Control Set"). Prior to the commencement of Seed Set Identification, see 4(c) below, the Responding Party shall disclose the results of the review of the Control Set to the Requesting Party, including the number of documents in the Control Set and the number of documents that were coded for each of the Responsiveness Categories during the review of the Control Set. The Responding Party shall produce all

non-privileged documents reviewed in the Control Set, and for each document disclose the Responsiveness Categories, if any, to which it is responsive. The Requesting Party shall raise any disputes regarding how the documents were coded for each of the Responsiveness Categories in the Control Set within five (5) business days of the production of 4,000 documents or fewer or ten (10) business days of the production of more than 4,000 documents. The parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced from the Control Set shall be deemed "Highly Confidential" under the terms of the Stipulated Protective Order (Dkt. No. 81), shall be used only for the purpose of evaluating the accuracy of the document coding, and shall be promptly returned or destroyed after review by the Requesting Party and the resolution of any disputes.

(c) Seed Set Identification. The Responding Party may use any reasonable method, including, but not limited to, search terms, to identify a set of documents to be used to initially train the Predictive Coding Software (the "Seed Set"). Prior to commencement of Training, see 4(d) below, the Responding Party shall disclose to the Requesting Party in writing a description of the size of the Seed Set and the methodology used to identify it. The Responding Party shall produce all non-privileged documents and disclose for each document the Responsiveness Categories, if any, to which it is responsive. The Requesting Party shall raise any disputes regarding how the documents were coded within five (5) business days of the production of 4,000 documents or fewer or ten (10) business days of the production of more than 4,000 documents. The parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced from the Seed Set shall be deemed "Highly Confidential" under the terms of the Stipulated Protective Order (Dkt. No. 81), shall be used only for the purpose of evaluating

the accuracy of the document coding, and shall be promptly returned or destroyed after review by the Requesting Party and the resolution of any disputes.

(d) Training Sets. The Responding Party may use any reasonable method to train the Predictive Coding Software. Upon completion of the training, the Responding Party shall disclose in writing the results of the training to the Requesting Party, including, to the extent reasonably available, the number of documents reviewed, the number of documents coded for each of the Responsiveness Categories during training, the number of documents identified as likely responsive by the Predictive Coding Process, and the estimated rates of Recall and Precision with their associated error margins. The Responding Party shall produce all non-privileged documents used to train the Predictive Coding Software and disclose for each document the Responsiveness Categories, if any, to which it is responsive. The Requesting Party shall raise any disputes regarding how the documents were coded within ten (10) business days of their production, and the parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced shall be deemed "Highly Confidential" under the terms of the Stipulated Protective Order (Dkt. No. 81), shall be used only for the purpose of evaluating the accuracy of the document coding, and shall be promptly returned or destroyed after review by the Requesting Party and the resolution of any disputes.

(e) Uncategorized Documents. The Responding Party shall disclose the number of documents the Predictive Coding Software is unable to evaluate for any reason, including the unavailability of machine-readable text or documents that could not be ranked, and review such documents in their entirety in order to identify responsive, non-privileged documents for production.

(f) Validation Set. Prior to production, the Responding Party shall review a Statistically Valid Sample of documents in the Document Universe that are categorized as likely non-responsive by the Predictive Coding Process (the "Purported Non-Responsive Documents") in order to determine the prevalence of responsive documents that are contained therein. Prior to production, the Responding Party shall disclose to the Requesting Party in writing the number of Purported Non-Responsive Documents, the size of the Validation Set, the number of documents identified as responsive during the review of the Validation Set, and the implied rate of Recall. The Responding Party shall produce any responsive, non-privileged documents it identifies during the review of the Validation Set.

5. General Provisions.

(a) The Parties hereby agree to meet and confer in good faith over any disputes that might arise with respect to the terms and conditions of this Stipulation or any other aspects relating to discovery. The Responding Party agrees to make its Technical Expert reasonably available to the Requesting Party for questions about its use of predictive coding. Should the Parties be unable to resolve their disputes on any issues stemming from the use of predictive coding set forth in this Stipulation, they shall submit those issues to the Court for resolution.

(b) Notwithstanding the provisions set forth in this Stipulation, the Parties respectively reserve their rights regarding the instant discovery process. This includes, but is not limited to, the Requesting Party's right to object to the efforts of the Responding Party to search for, review, and produce information in response to the Requesting Party's document requests; and the Responding Party's right to withhold information pursuant to

the objections it previously interposed in response to the Requesting Party's document requests.

IT IS SO ORDERED

Date: _____

_____
Hon. Andrew J. Peck
United States Magistrate Judge

William A. Burck
Eric C. Lyttle
Michael J. Lyle
Stephen M. Hauss
QUINN EMANUEL URQUHART & SULLIVAN, LLP
777 6th Street NW, 11th floor
Washington, DC 20001
williamburck@quinnemanuel.com
ericlyttle@quinnemanuel.com
mikelyle@quinnemanuel.com
stephenhauss@quinnemanuel.com

Counsel for Plaintiff Rio Tinto plc.

Jonathan I. Blackman
Lewis J. Liman
Boaz S. Morag
CLEARY GOTTLIEB STEEN & HAMILTON LLP
One Liberty Plaza
New York, NY 10006
jblackman@cgsh.com
lliman@cgsh.com
bmorag@cgsh.com

Counsel for Defendant Vale S.A.