

United States District Court, N.D. Illinois, Eastern Division.
IN RE BROILER CHICKEN ANTITRUST LITIGATION
This Document Relates To: All Actions

Case No. 1:16-cv-08637

January 3, 2018

Opinion

ORDER REGARDING SEARCH METHODOLOGY FOR ELECTRONICALLY STORED INFORMATION
HON. JEFFREY T. GILBERT U.S. MAGISTRATE JUDGE

*1 This Order Regarding Search Methodology for Electronically Stored Information (“Search Methodology Order”) shall govern the Parties in the above-captioned case whether they currently are involved or become so in the future, and any related actions that may later be consolidated with this case (collectively, the “Litigation”).

I. DOCUMENT SOURCE DISCLOSURES

A. Transparency: With the goal of permitting requesting Parties an appropriate level of transparency into a producing Party’s electronic search process, without micromanaging how the producing Party meets its discovery obligations and without requiring the disclosure of attorney work product or other privileged information, the Parties will endeavor to be reasonably transparent regarding the universe of documents subject to targeted collections or culling via search terms and/or TAR/CAL.

B. Pre-Search Deduplication & Culling of Collected Data:

1. De-Duplication: Before running either of the Search Processes below in II, data should be de-duplicated by hash value across all agreed or Court-ordered document custodians.

2. Email Threading: If the producing Party’s search software has the capability, then the producing Party may choose to only include inclusive emails in the data set subject to the Keyword and/or TAR/CAL Search Process, including the data set against which keyword searches are tested. A producing Party will disclose whether or not they are testing search terms in a set of data that excludes non-inclusive emails. Non-inclusive emails do not need to be searched, reviewed, or produced in this matter. See ESI Protocol, Dkt. 459, IV(D) (1).

3. Email Domains: Should the requesting Party want certain email domains excluded from the data set against which search terms are tested, the requesting Party must provide a list of such domain names to the producing Party ahead of the producing Party’s testing of search terms. Likewise, if the producing Party identifies domains that it believes should be eliminated, it will produce a list of those domain names to the requesting Party. See ESI Protocol, IV(D)(2). Plaintiffs’ analysis to date of the Florida AG productions made by certain Defendants indicates there are a large variety of industry email newsletters that can be culled (*i.e.*, excluded from a Defendant’s review and production of documents) where there are no internal forwards of such documents after their receipt. Plaintiffs agree to provide a list of these domain names prior to Defendants undertaking a search of their data.

4. Targeted Collections: Only documents a producing Party intends to subject to electronic searching parameters should be included in the data set against which search terms are tested. As an example, a centralized, non-custodial folder of responsive Agri Stats reports that a party intends to produce in its entirety (to the extent not privileged) should not be included in the data set against which search terms are tested.

5. Exception Reporting: For any documents not otherwise identified as system or operating files, the producing Party must disclose processing exceptions that are unresolved at the end of the discovery period, such as documents that cannot be opened due to encryption or other issues.

*2 6. Disclosure of Other Culling Parameters Required: A producing Party is permitted to cull data using the agreed-upon custodial and non-custodial sources, agreed-upon date parameters, and agreed-upon search terms (if applicable), and a producing Party is permitted to remove known system or operating files, such as

those that appear on the National Software Reference Library (NSRL) hash list. As such, the Parties may cull entire file directories from computer hard drives that contain Program Files, Program Data, SWTOOLS, Windows Operating System files, etc. For those excluded directories, the Parties will only conduct searches on user-created content that is reviewable and likely to yield relevant content. To the extent a producing Party elects to use additional culling parameters, those parameters will be disclosed.

II. SEARCH METHODS

The following TAR/CAL and Keyword Search Processes govern how collected data may be electronically culled in this matter.

A. TAR/CAL Search Process:

1. Use of Search Terms with TAR/CAL:

a. No later than December 22, 2017, the requesting Party will propose to a producing Party a limited number of Document Custodians for whom, across their email only, it requests that no search term pre-culling be used prior to applying TAR/CAL during the review process. However, the other data culling parameters described in I(B) may be applied to these Document Custodians, including to their email.

b. No later than January 12, 2018, the Parties will meet and confer on any issues or disputes regarding the requesting Party's proposals.

2. Producing Party TAR/CAL Disclosures:

a. No later than January 19, 2018, a producing Party that elects to use TAR/CAL will disclose the following information regarding its use of a TAR/CAL process: (a) the name of the TAR/CAL software and vendor, (b) a general description of how the producing Party's TAR/CAL process will work, including how it will train the algorithm, such as using exemplars, keyword search strings, or some other method, (c) a general description of the categories or sources of the documents included or excluded from the TAR/CAL process, and (d) what quality control measures will be taken.

3. Requesting Party Response:

a. Within 7 days of receiving a producing Party's TAR/CAL Disclosures, the requesting Party may raise with the producing Party any concerns with the proposed TAR/CAL process or categories of documents that it proposes should be excluded from the TAR/CAL process. A requesting Party may also propose any exemplars it proposes be used to train a TAR/CAL tool or narrow keyword search strings it proposes be used to generate exemplars to train a TAR/CAL tool. A producing Party retains the right to reject and oppose any such requests, subject to resolution by the Special Master and/or the Court.

4. Cooperation: The Parties agree to work together in good faith to resolve any differences that they may have over the producing Party's use of TAR/CAL and its processes, recall, and validation proposals. If an agreement cannot be timely reached, then the Parties agree to raise this issue with the Special Master and to follow her direction absent the showing of good cause to the contrary, and subject to the Parties' rights to petition the Court for review of or relief from any decision or guidance provided by the Special Master.

B. Keyword Search Process:

1. Iterative Process: Developing efficient keyword search terms is an iterative process and will require transparent and cooperative efforts by both the producing and requesting Party; however, it is important to set certain limits in order to effectively and efficiently manage time and expense.

2. Search Software Disclosures: No later than January 12, 2018, the producing Parties will disclose any search software they have decided to use (including version number) and that software's default stop/noise words and search language syntax. Additionally, the Parties should use best efforts to disclose information that answers these questions regarding their search tool:

*3 a. What stop words have been excluded from the index (if different than the default stop words for the tool)?

b. Can searches be constrained by upper- and lowercase?

c. Can numbers and single letters be searched?

d. Are there characters that cannot be searched or are treated as spaces or ignored?

e. How are diacritics resolved?

f. Can searches be run on metadata fields?

g. Are proximity-limited search terms subject to an evaluation order, e.g., will terms structured X w/5 Y yield hits if the text reads Y w/5 X?

h. Does the tool offer synonym searching?

i. How does the tool account for common misspellings?

3. First Phase Search Term Proposals:

a. *Producing Party Proposes an Initial Set of Search Terms* : No later than January 19, 2018, the producing Party will propose a set of search terms. The producing Party's proposal will include, to the extent known, semantic synonyms and common spellings of the keywords proposed. Where a producing Party seeks to exclude false positives (aka, "noise hits") by modifying or excluding certain keywords, then it will supply contextual examples of such false positives to explain why they must be excluded.

b. *Requesting Party's Proposed Revisions*: Within 12 days of receiving the initial proposed search terms, the requesting Party will provide any proposed revisions to the producing Party's search terms.

c. *Producing Party Provides Information Sufficient to Support Its Objections*: Within 8 days of receipt of the requesting Party's proposed revisions, the producing Party will provide information sufficient to support its objections to specific search terms, which could include, for example, estimates of the incremental number of false positive hits and the incremental number of true positive hits introduced by the disputed search terms, as well as examples of the false positive hits.

d. *Cooperation*: The producing Party and the requesting Party will work together in good faith to reasonably narrow the number of documents returned via search term hits and narrow the number of irrelevant documents captured as a result of the search terms. To the extent any disputes remain concerning the sufficiency of the producing Party's information in support of its objections and/or the use of specific search terms after good faith negotiations have occurred, either Party may request the assistance of the Special Master in resolving such disputes.

4. Second Phase Search Term Proposals:

a. *Requesting Party Proposes an Additional Set of Search Terms* : The Parties agree that Plaintiffs collectively and Defendants collectively may propose additional search terms to a producing Party one time. No later than May 14, 2018 (60 days after the Court-ordered March 15, 2018 deadline for "Rolling Production of Documents, Prioritizing Custodians as Agreed by the Parties or Ordered by the Court" (Dkt. 574)), the requesting Party may propose a set of additional search terms. The requesting Party will explain generally the basis for the additional requested terms, which could include, for example, identifying by Bates number exemplar documents that support the request.

b. *Producing Party Provides Information Sufficient to Support Its Objections* : No later than 10 days after the requesting Party provides an additional set of proposed search terms, the producing Party will provide information sufficient to support its objections to specific additional search terms, which could include, for example, estimates of the incremental number of false positive hits and the incremental number of true positive hits introduced by the disputed additional search terms, as well as examples of the false positive hits.

*4 c. *Requesting Party and Producing Party Will Meet and Confer Regarding Requesting Party's Proposed Additional Search Terms*: No later than 15 days after the requesting Party proposes an additional set of search terms, the Parties will meet and confer regarding any disputes or counter-proposals regarding the additional search terms. To the extent any disputes remain concerning the sufficiency of the producing Party's information in support of its objections and/or the use of specific additional search terms after good faith negotiations have occurred, either Party may request the assistance of the Special Master in resolving such disputes.

d. *Good Cause Inability of a Party to Meet the Deadlines Imposed in this Order* : While it is expected that the Parties shall make their best efforts to comply with the deadlines set forth in this Order, it is conceivable that technical (or other) issues or unanticipated volumes may interfere with a Parties' best efforts to comply. Should a Party anticipate that for good cause it may be unable to meet a deadline set forth in this Order, the Party shall promptly raise the issue with the other Parties, explain the reason for the inability to timely comply, and negotiate a reasonable extension for compliance. If the Parties are unable to immediately agree upon a revised deadline for compliance, they shall promptly raise the issue with the Special Master or the Court for resolution. This provision shall not be construed as blanket permission for a Party to modify or extend the

deadlines agreed to by the Parties and set forth in this Order without good cause, but rather, to recognize that when dealing with search and review of large volumes of electronically stored information, there are sometimes legitimate, unanticipated challenges that may interfere with a Party's best efforts to fulfill its obligations and therefore, to afford the Parties reasonable flexibility and mutual accommodation should such eventuality occur.

III. VALIDATION PROTOCOL

A. The review process should incorporate quality-control and quality-assurance procedures to ensure a reasonable production consistent with the requirements of Federal Rule of Civil Procedure 26(g). Once a producing Party reasonably believes that it has produced or identified for production substantially all responsive non-privileged documents, it shall conduct validation according to the sampling protocol described below and in Appendix A. This Validation Protocol shall apply to the review process regardless of whether technology-assisted review ("TAR") or exhaustive manual review ("manual review") was used by the producing Party.

B. The Document Collection ("Collection") is defined as including all documents identified for review for responsiveness and/or privilege following the application of keywords or other culling criteria. This Validation Protocol assumes that the completeness or adequacy of the Collection has already been established. For purposes of the three putative plaintiff classes' validation requirements under this Validation Protocol, the Collection refers to the combined set of documents of a particular proposed class of plaintiffs, rather than to each individual named representative of a particular class of plaintiffs.

C. The Collection shall be partitioned into the following two or three Subcollections, for manual review or for TAR processes, respectively:

1. Documents identified by the review as responsive to at least one Request for Production, including any privileged documents, but not including family members of responsive documents, unless those family members are deemed to be responsive in their own right ("Subcollection C(1)");

*5 2. Documents coded as non-responsive by a human reviewer, regardless of how the documents were selected for review (e.g., by TAR, manual review, or otherwise) ("Subcollection C(2)");

3. Documents excluded from manual review as the result of a TAR process ("Subcollection C(3)"). If the review process involved only manual review and no TAR, the Collection will not include Subcollection C(3).

D. A sample shall be drawn consisting of the following:

1. 500 documents selected at random from Subcollection C(1) ("Subsample D(1)");

2. 500 documents selected at random from Subcollection C(2), if TAR was used, otherwise 2,500 documents selected at random from Subcollection C(2), if manual review was used ("Subsample D(2)");

3. 2,000 documents selected at random from Subcollection 1(c) if TAR was used ("Sample D(3)"). If TAR was not used, there will be no Subsample D(3).

E. Should a producing Party believe that the sample sizes specified in Paragraph III(D) would be disproportionate or unduly burdensome under the circumstances, that Party shall promptly raise the issue with the requesting Party. To the extent a dispute remains concerning the sample sizes to be used after good faith negotiations have occurred, either Party may request the assistance of the Special Master in resolving such dispute.

F. The sample of 3,000 documents comprised of the documents from Subsamples D(1), D(2), and, if TAR was used, D(3), shall be combined into a single Validation Sample, with no indication of the Subcollection from which the documents were derived or how they were previously coded. The Validation Sample shall be reviewed and coded by a subject matter expert ("SME") who is knowledgeable about the subject matter of the litigation. This should be an attorney who is familiar with the RFPs and the issues in the case. During the course of the review of the Validation Sample, the SME shall not be provided with any information concerning the Subcollection or Subsample from which any document was derived or the prior coding of any document. The intent of this requirement is to ensure that the review of the Validation Sample is blind; it does not preclude a Party from selecting as SMEs attorneys who may have had prior involvement in the original review process.

G. Once the coding in Paragraph III(F) has been completed, the producing Party shall prepare a table listing each of the 3,000 documents in the Validation Sample. For each document, the table shall include:

1. the Bates number of the document (for documents produced), or a control/identification number (for non-produced documents);
2. the Subsample from which the document came (i.e., D(1), D(2), or, if TAR was used, D(3));
3. the SME's responsiveness coding for the document (i.e., responsive or non-responsive);
4. the SME's privilege coding for the document (i.e., privileged or not privileged). If the document is coded as non-responsive, a privilege determination need not be made. All documents in the Validation Sample coded as privileged shall be included on the producing Party's Privilege Log, as per the requirements set forth in VI of ESI Protocol (Dkt. 459).
5. for putative class plaintiffs, the named class representative associated with the document.

*6 H. The following items shall be provided to the requesting Party and to the Special Master:

1. the table described in Paragraph III(G);
2. a copy of each responsive, non-privileged document in the Validation Sample that was not previously produced or identified for production to the requesting Party;
3. the statistics and recall estimate detailed in Appendix A to this Order.

I. Once the requesting Party has received and has had an opportunity to review the items described in Paragraph III(H) and Appendix A, the Parties shall meet and confer to determine whether or not the Parties agree that the recall estimate, and the quantity and nature of the responsive documents identified through the sampling process, indicate that the review is substantially complete. If the recall estimate and the samples indicate that Subcollections C(2) and/or C(3) still contain a substantial number of non-marginal, non-duplicative responsive documents as compared to Subcollection C(1), the review and quality assurance process shall continue, and the validation process shall be repeated, as warranted. If the parties are unable to agree on whether the review is substantially complete, or whether the validation process must be repeated, the Special Master shall render a decision, subject to the Parties' rights to petition the Court for review of or relief from any decision or guidance provided by the Special Master.

SO ORDERED.

MAURA R. GROSSMAN SPECIAL MASTER

APPENDIX A

Method of Recall Estimation

An estimate of recall shall be computed to inform the decision-making process described in III(H) of the Validation Protocol; however, the absolute number in its own right shall not be dispositive of whether or not a review is substantially complete. Also of concern is the novelty and materiality (or conversely, the duplicative or marginal nature) of any responsive documents identified in Subsamples D(2) and/or D(3). The estimate of recall shall be derived as described below, depending on whether or not the review process involved the use of TAR. It should be noted that, when conducted by an SME pursuant to Paragraph III(F) of the Validation Protocol, a recall estimate on the order of 70% to 80% is consistent with, but not the sole indicator of, an adequate (i.e., high-quality) review. A recall estimate somewhat lower than this does not necessarily indicate that a review is inadequate, nor does a recall in this range or higher necessarily indicate that a review is adequate; the final determination also will depend on the quantity and nature of the documents that were missed by the review process.

Recall Estimation Method for a Review Process Involving TAR:

The number of responsive documents found \approx the size of Subcollection C(1) \times the number of responsive docs in Subsample D(1) \div 500.

The number of responsive documents coded incorrectly \approx the size of Subcollection C(2) \times the number of responsive documents in Subsample D(2) \div 500.

The number of responsive documents not reviewed \approx size of Subcollection C(3) \times the number of responsive documents in Subsample D(3) \div 2,000.

Estimated recall \approx the number of responsive documents found \div (the number of responsive documents found + the number of responsive documents coded incorrectly + the number of responsive documents not reviewed).

*7 Recall Estimation Method for a Review Process Involving Manual Review:

The number of responsive documents found \approx the size of Subcollection C(1) \times the number of responsive documents in Subsample D(1) \div 500.

The number responsive documents coded incorrectly \approx the size of Subcollection C(2) \times the number of responsive documents in Subsample D(2) \div 2,500.

Estimated recall \approx the number of responsive documents found \div (the number of responsive documents found + the number of responsive documents coded incorrectly).

End of Document.