

Technology-Assisted Review and Other Discovery Initiatives at the Antitrust Division

Tracy Greer
U.S. Department of Justice



Technology-Assisted Review and Other Discovery Initiatives
at the Antitrust Division

Tracy Greer¹
Senior Litigation Counsel E-Discovery

The Division has moved to implement several discovery initiatives that have a significant impact on lawyers and their clients who are involved in Division investigations. This paper discusses a number of those initiatives and some best practices for those involved in Division matters.

TECHNOLOGY ASSISTED REVIEW

Protocols for the Use of Technology-Assisted Review

In March 2012, the Division revised Instruction Q of its [Model Second Request](#) to require parties to disclose any “software or technology used to identify or eliminate potentially responsive documents.” This modification of the Model Second Request was part of a Division initiative to explore the possibility of allowing parties receiving Second Requests to use technology assisted review (“TAR”) to prepare responses to Second Requests. The use of TAR offers the promise of reducing the costs incurred by merging parties responding to Second Requests and the size of the document productions received by the Division, without undermining the ability of the Division to conduct an appropriately thorough investigation. But the use of TAR in investigations raises a number of difficult issues that require careful consideration by the Division and exploration with merging parties.

Although the use of TAR in litigation has been much discussed, its use in the investigative context has not drawn much attention. We believe that using TAR in government investigations, particularly in merger investigations, raises unique issues. First, in a merger investigation, in which the Division must explore the likely impact of a merger on future market conditions, the scope of discovery frequently will be more open-ended and less defined than in litigation, when the allegations in the parties’ pleadings may circumscribe the inquiry. Second, particularly at the outset of an investigation, the merging parties likely possess significantly greater information and knowledge about how the merging parties operate, the jargon used by the merging parties, the identities of key employees, and other peculiar features of their businesses than the Antitrust Division. Because the use of TAR requires making initial assessments of the relevance of a set of materials used by the TAR system to evaluate the responsiveness of the remainder of the collection, this information asymmetry between the Division and the parties raises issues in an investigation likely not present in post-

¹ The views expressed in this paper are my own and do not necessarily reflect those of the Department of Justice or the Antitrust Division.

complaint discovery, when litigating parties are on more equal footing and can both make accurate initial relevance determinations. Third, the scope of merger investigations can be broad and Division staff, by necessity, must explore varied, changing, and sometimes contradictory theories or facts until they obtain information sufficient to permit them to discard or narrow their earlier hypotheses.

After the March 2012 revisions of the Model Second Request, the Division began working with merging parties that expressed interest in using TAR to develop procedures that would account for these issues. To date, the Division has negotiated, or is in the process of negotiating, TAR protocols in approximately a dozen instances. We expect to release a model TAR protocol in the coming months. In anticipation of that release, I want to discuss the broad outlines of the agreements we have been negotiating and the results we have seen to date.

Negotiation of a TAR protocol with the Division requires the discussion of the following issues:

- (1) Issues and Theories of the Investigation;
- (2) Collection;
- (3) Characteristics of the Document Collection;
- (4) Workflow; and
- (5) Validation

Issues and Theories of the Investigation. Although not a part of the TAR protocol *per se*, the issues and theories in the investigation represent the foundation upon which the protocol is based. As in all negotiations to modify the scope of a Second Request, Division staff and counsel for the merging parties must discuss the potential issues and theories to be explored in the investigation. Division Staff and the producing party must take the broadest view of potential product and geographic markets and theories of competitive harm. This broad view will be critical to the success of a TAR protocol because, if the initial reviewer does not code a category of information as relevant, that category of information will not be produced to the Division. Thus, a TAR protocol will be predicated on a discussion of the information responsive to the Second Request and an embrace of the prevailing e-discovery best practices embodied in the Sedona Conference's Cooperation Proclamation.² This discussion also is an opportunity to narrow the scope of the production; accordingly, in the investigations in which the Division has negotiated a TAR protocol, we have encouraged the merging parties to identify categories of information and documents responsive to the Second Request, but that are not particularly relevant to the dispositive issues in the investigation. Production of these categories of information could be deferred or foregone entirely.

² The Sedona Conference Cooperation Proclamation, 10 Sedona Conf. J. 331, 333 (2008).

Collection. In the context of a TAR protocol, the Division must be sure that the process of collecting information does not undermine the effectiveness of the TAR software. For example, we are reluctant to permit a responding party to use search terms to collect documents that thereafter will be processed by a TAR platform. The use of search terms, with their well-known limitations,³ has the potential to exclude many responsive documents from the collection and, thus, to render ineffective the TAR platform. For those collections too voluminous to process cost-effectively, the Division would prefer a more objective means of narrowing the collection, such as narrowing date ranges, deduplicating the collection, eliminating certain domain names from email collections, or eliminating file types.

Characteristics of the Document Collection. Before discussing a TAR protocol with the Division, counsel should understand the characteristics of the documents and information to be collected. TAR is not an appropriate tool to capture certain categories of information demanded by a Second Request. For example, the production of databases and data collections should be negotiated separately and should not be included in the TAR process. Similarly, TAR would not be an effective means to capture responsive materials from websites, corporate intranets, or other sources of information not associated with an identifiable custodian. Because TAR considers only the text of the document collection, it will not reliably capture spreadsheets, pictures, images,⁴ drawings, audio or video files, or other non-text based information in the collection. If the responding party's collection likely contains responsive, non-text documents, the party should discuss with the Division an appropriate manual review of those materials. In investigations in which TAR has been employed, we have not been entirely satisfied that the TAR process works effectively with foreign and mixed-language documents (i.e., documents containing more than one language). If a company has such documents in its collection, we would be open to discussion about the appropriate course, but we have not yet settled on a procedure.

Workflow. The Division's TAR protocol discussions will also include how the producing party will structure the "workflow"⁵ for its TAR process. One key question is who will be responsible for the training and quality control review of the collection. The Division believes that the use of subject-matter experts to conduct the review is superior to using armies of less informed lawyers. Productions to the Division too often are rife with facially nonresponsive documents and information that is costly to load and review. Finally, in accordance with the recommendations of most vendors, we expect a producing party to deduplicate the collection before applying TAR, but not use other analytical tools, such as consolidating email threads, on the TAR collection.

³ See, e.g., Gregory L. Fordham, *Using Keyword Search Terms in E-Discovery and How They Relate to Issues of Responsiveness, Privilege, Evidence Standards, and Rube Goldberg*, 15 Rich. J.L. & Tech. 8 (2009).

⁴ Market share charts that typically appear in PowerPoint presentations are an obvious example. Such charts clearly would be responsive to the Second Request.

⁵ "Workflow" is used by vendors to describe how counsel should structure the application of the TAR to the document collection.

The discussions of the use of the TAR platform also will explore the means by which counsel plans to identify the “seed set,” i.e., the set of documents used to “train” the software to identify responsive and nonresponsive documents. The most common method of identifying the seed set is the use of statistical samples or a small collection of documents identified by subject-matter experts. Each TAR platform tracks different statistics or metrics after each round of review. The Division asks the producing party to provide some of this information in order to assure the Division that the TAR platform is performing as expected. In particular, we find the “overturn rate” – the frequency in which the subject-matter expert disagrees with the responsiveness determination of the TAR software – is a particularly important metric that gives the Division some insight into the workings of the TAR protocol.

Finally, our discussions will address the “confidence level.”⁶ To date, we have used confidence measures at the 90% level, or higher. We are aware of some criticism of using such high confidence measures, particularly as compared to the likely confidence level for manual review. We are open to discussing other confidence measures, depending upon the richness of the document collection. Applying a lower threshold probably would involve additional discussions and, possibly, some sampling to assure the Division that relevant information is being produced.

Validation. The Division consistently has asked the responding party to provide a statistically significant sample of nonresponsive documents to ensure that facially responsive documents were not excluded from the collection. Typically, we ask the responding party to generate five to seven statistically significant samples from the set of documents deemed nonresponsive. A Division representative selects one or more of the samples, and the samples (minus any documents coded as privileged) are made available to a Division lawyer for a quick review. Division staff reviews the samples over a secure Web-based viewer or in the offices of the producing party’s counsel. Generally, the Division agrees to complete its review of the samples in one to two days. The focus of the review is to look for obviously responsive documents so material that their exclusion would undermine confidence in the TAR process. All parties that successfully have negotiated TAR protocols have agreed to this process.

The Division’s Experience to Date

The Division has negotiated or is negotiating TAR protocols in approximately a dozen instances. These instances almost exclusively have been in the context of Second Request productions, but we also have negotiated a TAR protocol in a filed civil case.

⁶ Confidence level reflects the statistical accuracy of the review process. See Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, Vol. 7 Fed. Cts. L. Rev. 1, 12 (2013).

Parties have inquired about use of TAR in other contexts, including in response to Civil Investigative Demands, but we have not reached agreement on a TAR protocol in those instances. I have a few observations based on our experience:

1. TAR produced smaller, more responsive document productions. The TAR productions have contained much more relevant information and less that obviously is not responsive. Division staff have experienced substantial benefit, and the producing parties have reported substantial time and cost savings.⁷
2. TAR requires expecting the unexpected. Because TAR is a new method of responding to Second Requests, open communication and transparency about the process are essential components of a successful experience. In one instance, for example, the producing party observed unexpected results as the collection was coded. Counsel discovered that the search for information relevant to two independent theories of competitive harm relating to two different product markets produced contradictory results. The producing party proposed, and we agreed, that the collection would be processed twice, each time coding separately for one of the two theories.
3. Producing parties sometimes prefer manual review. In some instances, the producing party preferred to use traditional manual review, rather than TAR, for the documents of the most senior executives. Although the Division did not object to this approach, judgments about responsiveness during manual review are less accurate and almost certainly are not consistent among reviewers. Similarly, counsel for a producing party sometimes have preferred the use of search terms in connection with a TAR protocol despite the well-known limitations of search terms. Ideally, however, a TAR protocol would apply to the entire collection so that the increased accuracy and consistency of TAR would be realized.
4. Privilege review continues to be costly and time consuming.⁸ Because counsel for producing parties understandably may be risk averse with regard to client privileges, all potentially responsive documents typically are subjected to a page-by-page privilege review. This type of review can undermine some of the cost and time savings of TAR. Consistent with Fed. R. Civ. P. 26, the Division's policy is to sequester or return inadvertently produced privileged documents.

⁷ See Joe Palazzolo, *Software: The Attorney Who Is Always on the Job*, WALL ST. J., May 6, 2013, at B1 (reporting savings of 50% for a Second Request production using a TAR protocol compared to a Second Request production using manual review).

⁸ See generally Hon. John M. Facciola & Jonathan M. Redgrave, *Asserting and Challenging Privilege Claims in Modern Litigation: The Facciola-Redgrave Framework*, 4 FED. CTS. L.R. 19 (2009).

5. Opportunities to narrow the production further exist. The Division has encouraged parties using a TAR protocol to identify categories of documents that, while technically responsive to the Second Request, are not essential to resolving the competitive concerns at issue in the investigation. Parties should view this as an opportunity to reduce further the size of the production.

Going Forward

As the Division gains additional experience with TAR in the merger context, we will consider opportunities to expand the use of TAR to other types of investigations, such as grand juries or civil conduct investigations. TAR might be completely unsuitable for some types of investigations – for instance, in a conspiracy investigation, the fact that a meeting took place at a particular time and place might be significant but might not be revealed through use of TAR. We caution against using TAR without disclosing its use to the Division. Also, as we expand our experience using TAR, we encourage counsel to provide suggestions and feedback about TAR process and protocols.

OTHER DISCOVERY INITIATIVES

Search Terms

Although TAR is the topic most discussed, the vast majority of producing parties continues to use search terms to narrow document productions. The Division, as well as the Department, has devoted significant effort to educating its lawyers about the limitations of search terms and the ways in which search terms can be used more effectively. To this end, the Division has addressed search terms in Instruction Q of the Model Second Request:

... If search terms will be used, in whole or in part, to identify documents and information that are responsive to this Request, provide the following: (1) a list of the proposed search terms; (2) a word dictionary or tally list of all the terms that appear in the collection and the frequency with which the terms appear in the collection (both the total number of appearances and the number of documents in which each word appears); (3) a glossary of industry and company terminology (including any code words related to the Transaction); (4) a description of the search methodology (including the planned use of stem searches and combination (or Boolean) searches); and (5) a description of the applications that will be used to execute the search. The Department strongly recommends that the company provide these items prior to conducting its collection of potentially responsive information and consult with the Department to avoid omissions that would cause the company's response to be deemed deficient.

We added these recommendations in order to avoid the “Go Fish” nature of most search term negotiations.⁹ Lawyers must recognize that they use words differently from most document custodians and that they are ill-suited to selecting search terms. Join a group of non-lawyers in conversation and just listen for a few minutes (this will be challenging!). At the end of this experiment, ask yourself if lawyers having the same conversation would use the same phrases and words.

Unicode Productions

As the Division’s investigations touch more international companies, we have seen a substantial increase in foreign-language productions. To better process these productions, the Division has developed production specifications for Unicode TIFF and TXT, an alternative to our standard ASCII TIFF and TXT production based on an English-only character set. Before beginning negotiations with Division staff, counsel for a producing party should know whether and to what extent the party’s documents are in languages other than English. This is particularly important for character-based languages (e.g., Japanese).

Translations

Division investigations, particularly criminal investigations, are becoming more international in scope and therefore encompass documents in languages other than English. To facilitate foreign- language productions, the Division has developed production specifications for translations that are produced with the original document.

Preservation

Some companies are moving aggressively to implement very short preservation policies (e.g., 30 days) for some electronic documents, particularly email. Companies with such short preservation policies must suspend these document retention policies upon learning of that they are subjects of an investigation. Issues of burden or cost associated with preservation should be raised to Division staff. Division staff typically inquire about document retention policies soon after an investigation begins and, where appropriate, send preservation notices, particularly in connection with conduct investigations. However, a company that is the subject of an investigation, should suspend document retention policies regardless of whether the Division sends a formal preservation notice to the company.

⁹ As Judge Peck recently observed, “the way lawyers choose keywords, [as Ralph Losey [suggests](#)] is the equivalent of the child’s game of ‘Go Fish.’ The requesting party guesses which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party’s ‘cards’ (i.e., the terminology used by the responding party’s custodians). Indeed, the responding party’s counsel often does not know what is in its own client’s ‘cards.’” *Moore v. Publicis Groupe*, 287 F.R.D. 182, 191 (S.D.N.Y. 2012).

Destroying Documents and Information When Investigations are Closed

For years, at the conclusion of each investigation, Division staff contacted every party that had produced documents, to inquire whether those documents should be returned or destroyed. Given that virtually all documents produced to the Division are “copies” of electronic documents that reside on corporate networks, not surprisingly, the vast majority ask the Division to destroy those documents. The Division has changed its policies to reflect that majority position.

Going forward, a producing party must identify any original paper documents by designating the documents themselves as “Original” and by providing the Division with a list of such documents at the time they are produced. These documents will be returned to the producing party at the conclusion of the investigation. For electronic copies, the producing party may ask the Division to return submitted material at the conclusion of an investigation, but the request **must** be made when the material is **produced**. If the producing party does not request the return of the material at the time of production, and it does not have an “Original” designation, the Division will destroy the material at the conclusion of the investigation, with the exception of copies (whether of original documents or an electronic copy) kept by the Division for a law enforcement purpose or as required by the Federal Records Act.

BEST PRACTICES FOR NEGOTIATING WITH THE DIVISION

“Don’t forget to fly the plane.” The **point** of producing documents to the Division is to resolve the antitrust investigation. Given the complexity of antitrust investigations, why should the producing party spend time and money on disputes about the scope or nature of electronic productions? To that end, the following best practices are recommended for counsel engaged in negotiating electronic productions with the Division:

1. Be prepared. Spend some time with your client before your first conversation with Division staff. You should be familiar with the company, including how it stores its information, the databases it keeps, as well as whether there is anything unusual about the company’s information or how it is kept (e.g., foreign-language documents, short retention schedules). Counsel should understand the client’s backup policies, particularly for email.
2. Use the current production specifications. The Division has developed specific and detailed production specifications, both for ASCII or English-only and for Unicode TIFF and TXT productions that accommodate Asian, as well as Latin, character sets. The production specifications are adjusted frequently in response to problems that arise, so ask Division staff for the current version rather than re-using specifications from a prior investigation.

The Division's Litigation Support Case Managers are available to answer questions about the production specifications.

3. Sample productions. Please take the time to provide to the Division a sample production using actual documents from the collection. We review these sample productions carefully to identify early any concerns or issues to avoid the need to reproduce materials. The Division discourages the use of a preexisting or "dummy" set of documents for the sample production, as these types of productions fail to reveal issues or problems that may exist with the actual production set.
4. Communication and transparency. Candid conversations regarding issues with or concerns about the production are the most effective means of resolving problems. Candor builds confidence and trust and makes resolving disagreements much easier.

The Division, along with the rest of the Department, has adopted policies and procedures to address electronic discovery issues. The Division has embraced the use of TAR platforms to narrow the scope of productions in response to Second Requests and has adopted other initiatives to address the ever-expanding scope of documents and information. Counsel who appear before the Division are encouraged to suggest other ways to make investigations more efficient and less costly.