October 2023

THE FUTURE SOCIETY

IEEE

Model Protocol

for Electronically Stored Information (ESI)

Text and Commentary

Note on the Preparation of these Documents

These documents grew out of conversations between Nicolas Economou and Bruce Hedin following the publication, in 2019, of the first edition of *Ethically Aligned Design*, the flagship publication of the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems. Their conversations centered on the question of how to enable practitioners to act on the principles articulated in that document (to which both Mr. Economou and Dr. Hedin had contributed). Recognizing the benefits of approaching the question through a specific application, they focused on AI-enabled systems applied to legal discovery and sought to identify a mechanism that would enable the use of those systems to be grounded in an informed trust (and, more specifically, in accordance with IEEE's recommendations, grounded in the key principles of effectiveness, competence, transparency, and accountability). That discussion led to the vision of creating an ESI Protocol that would instantiate the IEEE trust principles.

Mr. Economou and Dr. Hedin initially shared their vision with interested parties from the IEEE (Konstantinos Karachalios), The Future Society (Nicolas Miailhe), and NYU's Center on Civil Justice (David Siffert, Arthur Miller, the late Peter Zimroth). Having received positive responses from those parties, they sought to gain additional validation of the relevance of their vision by outlining their ideas to a wider circle of stakeholders (a group of judges, practitioners, consultants, and academics with an interest in legal discovery, most of whom eventually became a standing review group for the documents).

After those initial steps, the project remained largely at the outline stage for a couple of years as other matters intervened. In 2022, however, thanks to support from the IEEE and further guidance from The Future Society and NYU's Center on Civil Justice, the project was resumed in earnest. Dr. Hedin took the lead in drafting; The Future Society's Samuel Curtis took on the role of project manager. In keeping with the initial vision, Dr. Hedin drafted three documents: a model protocol, a line-by-line commentary on that protocol, and a handbook for practitioners seeking a deeper understanding of the procedures prescribed in the protocol. The review group was revived in 2023 and generously gave their time to review multiple drafts of the documents. As the project approached the final draft stage, Hon. John M. Facciola, U.S. Magistrate Judge (Ret.) agreed to add a preface to the documents.

This is the story of the documents before you. As will be evident, there are many individuals to whom thanks are due; please see the *Acknowledgements* section for an effort at recognizing them. While the documents have undergone multiple reviews, it is to be expected that more will be learned when they are put into practice. Feedback, comments, and suggestions are welcome; please send them to the following addresses.

Bruce Hedin
Hedin B Consulting
bhedin@hedinb.com

Samuel Curtis
The Future Society
samuel.curtis@thefuturesociety.org

Table of Contents

Foreword

John M. Facciola, U.S. Magistrate Judge (Ret.)

Those of us who labored in the field of improving lawyers' competence and thereby improving the quality of their work tried to explain to lawyers and judges how technology professionals had established indubitably, more than a decade ago, that technology-assisted review ("TAR") was more efficient and more effective than manual review and reached a result that could be easily defended. [1] Thus, it saved the client a lot of money and it worked.

One would think that would do it. But, like many of my colleagues, I always faced the inevitable question once I finished my presentation on the effectiveness and efficiency of machine learning; "I don't get all this math stuff, but I still say a lawyer has to look at every document to do the job."

I would bite my tongue and resist the temptation to say to my interlocutor: "I wish you a long and happy life, but neither you nor anyone else on the face of the earth has enough lifetime left to review a petabyte of data, document by document." Instead, I apologized for my failure to convince the person, and moved on, recalling the wise advice of someone who told me that, like the character portrayed by Dustin Hoffman in *The Graduate,* I should have gone into plastics. I would guess the person who asked the question returned to their office and read the same email 25 times.

Convincing lawyers of doing something revolutionary is, therefore, not the easiest way to spend an afternoon. Fortunately, lawyers are the ultimate pragmatists. When the bills for reviewing documents arrived in the early days of discovery, one could almost hear the screams. The bill was, of course, more than what many clients must have thought was the settlement value of the case. The clients surely thought there had to be a better way when they saw how technology was affecting so profoundly the other aspects of their businesses. Lawyers had to find that better way and it was inevitably the product of the sophisticated application of technology. Simultaneously, their desire to cut costs while meeting judicial requirements generated an entire industry of vendors who served their needs.

I see the publication of the IEEE protocol as the culmination of the two processes that have generated a nearly universal acceptance of a central theorem: lawyers and judges should cooperate to create an e-discovery process that is proportionate to the parties' needs.

The first process was generated by the Sedona Conference. It saw that a significant impediment to e-discovery that was efficient was the perception of most lawyers that the discovery process was supposed to be as adversarial as the rest of the litigation. Thus, if a lawyer did not fight every discovery demand with the same zeal as they brought to the trial process, they were being unfaithful to the ethical requirement that they zealously represent their client.

---

[1] *See* Gordon V. Cormack and Maura R. Grossman, *Navigating Imprecision in Relevance Assessmsnts on the Road to Total Recall: Roger and Me*, in Proc. of the 40th Int'l ACM SIGIR Conference on Research and Dev. In Info. Retrieval (SIGIR '17) 5-14 (2017), available at https://dl.acm.org/doi/10.1145/3077136.3080812; Maura R. Grossman and Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, 17 Rich. J.L. & Tech. 11 (2011), available at https://scholarship.richmond.edu/jolt/vol17/iss3/5/.

While there is little evidence that the drafters of the Federal Rules of Civil Procedure in 1938 shared the lawyers' perception that discovery is supposed to[2] be eternally adversarial, there was a solution simpler than searching that history. Instead, all that Sedona hoped to accomplish was to have the lawyers agree that it could not possibly be an ethical violation to save your client money by cooperating with opposing counsel when that cooperation found a simple solution to a discovery problem, which did not surrender a client's rights or impact the final resolution of the case.

Sedona accomplished that by producing the Sedona Cooperation Proclamation, which was then endorsed by more than a hundred judges, signaling that they expected lawyers to cooperate in resolving discovery disputes.

The second process was generated by the amendments to the Federal Rules of Civil Procedure that imposed the obligation that counsel meet and confer to arrive at a proposed discovery plan to be approved by the judge.

It is therefore not surprising that we have entered the age of ESI protocols. In nearly every case where it is worthwhile, counsel work together to create a protocol that will govern every aspect of how they conduct e-discovery. Courts insist on this process and take it seriously. Indeed, in a recent case, a judge chastised a lawyer for violating the protocol to which the lawyer had agreed. Thus, despite its humble origins, the Sedona Conference exceeded the expectations of those of us who worked through it to the point where "meeting and conferring" now occurs in every civil case.

For too long, the attitude towards discovery was animated by traditional negotiation techniques. Judge Paul W. Grimm described them as lawyers asking in discovery for "the sun, the moon, the stars, and a pony" and seeing what happened. Most lawyers would have been happy with the pony, but they thought that they had to demand everything else before they "settled" for the pony. But the time it took to get there was always a significant impediment to efficient discovery, particularly because everyone knew unreasonable demands were an obvious and transparent ploy, borrowed from commercial negotiations.

The Federal Rule of Civil Procedure may have encouraged this kind of behavior by a definition of the scope of the discovery. At one point, the rule pertaining to discovery permitted discovery of information that was "reasonably calculated to lead to discovery of admissible evidence." The Advisory Committee on the Federal Rules modified the definition of discovery in 2015; now discovery is limited to information that is "relevant to a party's claim or defense and proportional to the needs of the case."[3] Thus, the correct question is whether the benefit to be gained by the discovery that is sought is justified by the time, effort and money needed to accomplish it. Proportionality is the touchstone for all discovery.

The IEEE protocol is animated by this principle universally. It makes constantly clear that the perfect is unachievable (unless we bankrupt litigants by insisting upon it) and that the proper standard is whether scientific evidence establishes that the procedure under the protocol reaches a reasonable result. Thus, we

[2] Cooperation Proclamation. 10 Sedona Conf. J. 331 (2009 Supp.)
[3] Cole's Wexford Hotel, Inc. v. Hightower, Inc., 209 F. Supp. 3d 810, 817-823 (W.D. Pa, 2016).

avoid bad guesses about whether we could have done better and replace them with scientific evidence showing that we have done the best we can do, with reasonable effort, and doing more cannot be justified. To say that is welcome is an understatement. Confronted with complaints about a discovery process, judges exercise their discretion in applying the factors identified in the Federal Rules of Civil Procedure to ascertain whether the discovery response attacked as insufficient is reasonable and proportionate to the needs of the case. While judges exercise that discretion all day long, they will welcome the parties agreeing to a protocol that will achieve a response to a discovery demand that can be objectively validated by a scientific process to which the parties have agreed.

Finally, like everything else in our world, the discovery process will have to meet the new challenges of artificial intelligence ("AI"). We are already seeing vendors advertising that they have or will make AI available in their e-discovery systems.

Fortunately, two stalwarts of the impact of technology on evidence collection, Judge Grimm and Dr. Maura R. Grossman, have already described the new challenges the federal courts will have to meet before they admit into evidence the product of AI.[4] More specifically, they show how Federal Rules of Evidence will require that a proponent of such evidence will have to establish its validity and reliability by first establishing that the process yielding the result meets the demanding standards imposed by the Federal Rules of Evidence before it is admitted into evidence. Thus, when evidence, produced by technological means, using a scientific process, is offered, Grimm and Grossman insist the court should not admit it unless (1) the process has been verified through independent means; (2) the analysis has been published and subject to peer review; (3) the error rate is known and not unacceptably high; (4) standard testing methods and protocols have been used in the creation of the evidence; and (5) the methodology used to create it is generally accepted within a profession or scientific discipline.

While I am certain that the IEEE did not anticipate it, use of the protocols it recommends, will, in my view, meet the requirements imposed by the Federal Rules of Evidence when the result the protocol achieved is attacked as insufficient. I appreciate that there is a world of difference between the process of admitting scientific evidence and a determination that a search methodology is reasonable for the discovery process. Still, the work of the IEEE has given us another, albeit inadvertent, blessing by supporting the process it creates with the strongest empirical evidence possible.

The IEEE has provided lawyers and judges with a remarkably beneficial solution to the problem of the use of machine learning in discovery. Now, we get to use it and test it in the best possible way. We have to hope that it will thereby reduce costs and increase efficiency thereby increasing the access to justice, which should be the goal of all technical innovation that is designed to assist lawyers and judges.

[4] Paul W. Grimm, Maura R. Grossman, and Gordon V. Cormack. *Artificial intelligence as evidence*, 19 Nw. J. Tech. & Intell. Prop. (2021). See also Maura R. Grossman, Paul W. Grimm, Daniel G. Brown, and Molly (Yiming) Xu, *The GPTJudge: Justice in a Generative AI World*, 23 Duke L. & Tech. Rev. A preprint is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4460184.

THE FUTURE SOCIETY

IEEE

Model Protocol for Electronically Stored Information (ESI)

Protocol And Commentary

September 2023

Introduction

The potential that advanced review technologies (Technology-Assisted Review or "TAR") hold for advancing the Rule 1 objective of a "just, speedy, and inexpensive determination of every action and proceeding"[1] has been clear to litigants and the courts for over 15 years.[2] The advent, more recently, of a range of different legal applications based on Large Language Models ("LLMs") may bring additional hope for advancing the cause of effectiveness and efficiency. Attempts to realize the potential, however, often encounter frustrating obstacles. For example:

- **A Responding Party**, hoping to adopt TAR to meet its discovery obligations in a more efficient manner, instead finds itself bogged down in a costly and time-consuming meet-and-confer process, in which it must defend positions, such as that it has no obligation to disclose non-responsive documents, that would never be questioned in the absence of TAR.
- **A Requesting Party**, having no objection in principle to a Responding Party's plan to use TAR, but seeking some basis for trusting in the results of the technology, finds that its requests for minimal transparency into the effectiveness of the process are met with suspicion and obstinacy and are answered only after lengthy negotiation and even court intervention.

[1] Fed R. Civ. P. 1.

[2] For early case law supporting the use of advanced review technologies for purposes of legal discovery, see, for example, *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012), adopted sub nom. Moore v. Publicis Groupe SA, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012) and *Rio Tinto Plc v. Vale S.A.*, 14-Civ. 3042 (RMB)(AJP) (S.D.N.Y. Mar. 2, 2015); see also *United States v. O'Keefe*, 537 F. Supp. 2d 14 (D.D.C. 2008) and *Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 260, 262 (D. Md. 2008). For more recent case law on advanced review technologies, see Sedona 2023. For early academic studies demonstrating the potential effectiveness of such advanced technologies, see, for example, Hedin et al. 2009 and Grossman & Cormack 2011.

- **A Judge**, faced with a busy docket and hoping that the Responding Party's adoption of TAR will enable a speedier discovery process, instead finds the Court called upon to referee multiple battles of experts over minutiae of the TAR process.
- **A Citizen**, given a window on the proceedings, finds that the US system of civil justice, despite the potential improvements that TAR offers, remains burdened with inefficiencies, liable to produce outcomes determined more by the cost of discovery than by the merits of a case, and subject to barriers to participation that impede injured parties from seeking justice.

The goal of this model protocol (together with the companion commentary and practitioners' guidelines)[3] is to remove, or at least lower, the obstacles in the way of the effective use of TAR in legal discovery, thereby realizing the technology's potential to find the facts that matter in a lawsuit or investigation more efficiently and effectively.

The Protocol does so by addressing the issue at the root of the obstacles: **a lack of trust**. New technologies have introduced new sources of distrust: in the explainability of the technology, in the effectiveness of the technology, in the competence of the operators of the technology, and in the significance of tests conducted to validate the effectiveness of the technology. Drawing upon research into the sources of trust in technology,[4] and in particular upon the **trust conditions** articulated in the Law Chapter of the IEEE's report on paths to the responsible design, development, and use of advanced technologies,[5] the Protocol seeks:

- To enable the gathering of the evidence needed to assess the **effectiveness** of a document review effort (whether technology-assisted or manual);[6]
- To equip practitioners, of all levels of experience with TAR, with the **competence** needed to execute (or oversee the execution of) a meaningful validation exercise and to assess the results of such an exercise;[7]
- To provide for **transparency** into a Responding Party's review process where trust requires it (*and only where trust requires it*);
- To ensure that lines of **accountability** are not broken by gaps in the flow of understandable and meaningful information to attorneys of record;
- To accomplish all of the above in a manner that is streamlined and **efficient**, while also respecting operative rules of procedure, case law, and the principle of **proportionality**.

[3] Henceforth "Protocol," "Commentary," and "Guidelines."
[4] See, for example, Ferrario et al. 2019; Desai & Kroll 2017; Taddeo 2017; Spiekermann 2015; Coeckelbergh 2012; Taddeo & Floridi 2011; Taddeo 2010.
[5] IEEE 2019 (*EADe1*); for the chapter on Law, see pp. 211-281. The cited version of EAD (edition 1) is a later (2019), revised and expanded, version of an earlier (2017) public-comment version that IEEE identified as "version 2" of its public-comment drafts. Readers are directed to the Law chapter as published in the 2019 (edition 1) version of EAD.
[6] The excitement generated by the advent of powerful, publicly accessible, applications based on LLMs should not be allowed to obscure the fact that effectiveness still matters. We still need to know whether a given application works and that means *testing*, both of the generic kind (as in benchmarks) and of the local kind (as in a test of results in a specific circumstance).
[7] Realizing a wider distribution of the competence needed to conduct meaningful validation of the results generated by advanced review technologies is a key objective of the Protocol (and supporting materials). A wider distribution of competence will advance both the goal of making the conduct of meaningful validation exercises a routine step in the use of advanced technologies and the goal of leveling the playing field between resource-rich and resource-poor parties to litigation.

If the Protocol achieves these intermediate goals, it will also help to (a) **reduce the waste and unproductive wrangling** that currently slow fact-finding in civil litigation, (b) **shift the focus of proceedings from discovery questions and costs to the legal and factual issues** being litigated, and (c) **level the discovery playing field for all parties**, including those with limited resources and those with little experience in the use of advanced discovery technologies.

In order to meet these goals, the Protocol and associated materials have been designed and drafted as follows.

- **A Model Protocol**: an adaptable model ESI protocol that addresses the key issues that currently trouble parties in the discovery phase of litigation. The Protocol, in the form of a stipulation and order augmented by three appendices, focuses on gathering the evidence needed to have an informed trust in the results of a review; its provisions are shaped by the principles of proportionality and evidence-based decision-making.
- **A Commentary**: a line-by-line commentary on the Protocol. The Commentary is designed to provide legal authority, justification, interpretive guidance, and tutorial background for the Protocol's provisions.
- **A Set of Guidelines for Practitioners**: a companion document that provides further guidance on the sampling and measurement procedures specified in the Protocol and a glossary of terms of art used in validation. Focused on the validation component of the Protocol, but not linked as a line-by-line commentary, the Guidelines for Practitioners is intended to serve as an in-depth resource for practitioners (and other stakeholders) seeking a more thorough grounding in the concepts and methods underlying the approach to validation adopted in the Protocol.

Note on the intended use of these materials

Topics covered in the Commentary and Guidelines. To achieve its objectives, the Protocol must at times address topics (such as sampling and measurement) that are not typical components of legal training. The Commentary and Guidelines are designed to provide background and guidance for procedures and subject matter with which lawyers may be less familiar. The intended audience for these supporting materials is twofold.

- **Advanced practitioners** seeking to execute the Protocol's provisions in a sound and confident manner. In this category are lawyers, consultants, and vendors who already have a reasonably solid grounding in the use of advanced review technologies, but who need the additional know-how to design, execute, and defend a sound validation exercise.
- **Stakeholders** seeking to understand the resources, time, and cost required to gain a well-grounded trust in the results of a review effort. In this category are parties, lawyers, and judges who, while not themselves engaged in the execution of validation procedures, do need, in order to arrive at well-informed answers to the questions they face, a basic understanding of what a sound validation exercise requires.

To meet the needs of this audience, the supporting materials, while intended to be generally accessible, cover the topics of testing and measurement in some depth. As a practical matter, it is not expected that all

users of the Protocol will need the same level of depth on all topics. The Commentary and Guidelines are therefore designed, not as text to be read end-to-end, but as resources to be drawn upon as needed.[8]

Adaptation. The model protocol is a *model* that is designed to be adapted to the specific circumstances that occasion its use. Two points regarding its adaptability are worth highlighting.

- **Jurisdictional adaptation.** In its design and provisions, the Protocol models an agreement to which parties might stipulate in a civil proceeding in a federal court and therefore reflects the requirements of the Federal Rules of Civil Procedure. The Protocol is intended, however, to be used in other types of proceedings and other jurisdictions (including state courts) in which other rules are operative. It is expected that practitioners will adapt the model protocol to suit the requirements of their specific jurisdiction and proceeding.
- **Principled adaptation.** While it is expected that the Protocol will be adapted to circumstance, it is also expected that any adaptation will be in keeping with the general principles and reasoning, laid out in the Commentary and Guidelines, that have shaped the model.

Practical matters. Finally, two practical notes to aid in using the materials

- **Hyperlinks.** In the Protocol, words and phrases on which there is elaboration in the Commentary are indicated by double underlining. The Commentary's comment on any word or phrase so indicated is accessible via hyperlink. A hyperlink also allows passage from the entry of a comment back to the prompt in the text of the Protocol.
- **Commentary vs. Guidelines.** The Commentary, as a line-by-line annotation to the Protocol, treats topics as they occur in the Protocol. While this presentation of the material is useful for users reading the Protocol, it may not serve the needs of a user seeking a more systematic treatment of a given topic. Such a user may find the Guidelines better suited to the need. Chapter 3 of the Guidelines, for example, contains a glossary of terms of art used in validation; while much of the content of the glossary could be retrieved from the Commentary, it may be more easily retrieved from alphabetic presentation in the Guidelines (at least by readers not seeking elaboration of a specific point in the Protocol).
- **Location.** The Protocol begins on p. 11 of this document. The Commentary to that protocol begins on p. 22 of this document. The Guidelines are published as a separate document.[9]

[8] Given the calibration of the supporting materials to the needs of a reader with intermediate familiarity with the use of advanced review technologies, some readers (those with less familiarity with review technologies) may still be in need of some introductory guidance on the technologies (i.e., a "primer" on ESI and ESI review technologies). The materials presented here, while they do give attention to background concepts and considerations, are not designed to be such a primer. Readers seeking additional introductory material are directed to the various resources published by The Sedona Conference (e.g., Sedona 2023; Sedona 2018; Sedona 2014), by the Bolch Judicial Institute of Duke Law School (e.g., Bolch 2019), and by EDRM (see https://edrm.net/). Readers may also find helpful grounding in Grossman & Cormack 2021 and Baron et al. 2016.

[9] https://thefuturesociety.org/model-protocol-for-esi-guidelines/

UNITED STATES DISTRICT COURT
[X] DISTRICT OF [Y]

ABC INC., Plaintiffs, -against- XYZ INC. Defendants.	Case No. x:xx-aa-xxxxx **STIPULATION AND ORDER**

The provisions contained in this Stipulation and Order (collectively, the "Protocol") shall govern the manner in which they discuss and conduct the review of electronically stored information ("ESI") and hard copy documents in this case (the "Litigation").

I. DEFINITIONS[14]

Key terms used in the Protocol are defined as follows.

A. **Document:** Either a set of one or more tangible pages or an electronic file, produced manually or by a software application, constituting a logical single communication of information.

B. **Attachment:** Any document associated with another document for purposes of communication, storage, retention, or routine records management.

C. **Collected Set:** The set of documents identified by the Responding Party as potentially subject to review for responsiveness.

D. **Review Set:** The set of documents, within the Collected Set, actually designated for a review for responsiveness.

II. PROVISIONS

The Parties agree to the following provisions.

A. Defining the Review Set

1. The Responding Party may use the following methods to reduce the volume of data from the Collected Set to be included in the Review Set:

[14] Terms not defined in this section are used in accordance with the definitions provided in the *Commentary on the Protocol* or in the *Guidelines for Practitioners* (specifically, Chapter 3: *Glossary of Terms of Art Used in Validation*) that accompanies the Protocol. Terms of art not defined in either the Commentary or the Guidelines are used in accordance with the standard definitions provided in Grossman & Cormack 2013 and Sedona 2020.

a. Deduplication;

b. De-NISTing;

c. Email threading.

2. In addition, the Responding Party may, provided it adheres to the validation provisions specified below (II.A.3), use the following methods to identify portions of the Collected Set to be excluded from the Review Set:

a. Identification based on metadata values (e.g., Custodian, DateSent);

b. Identification through the application of search terms to document text.

3. Applications of the exclusionary methods identified in II.A.2 above shall be validated as follows.

a. If there is a dispute regarding a proposed exclusion based on metadata values, the Parties shall endeavor to resolve the dispute, in the first instance, based on *a priori* considerations. If *a priori* considerations are not dispositive, the proposed exclusion shall be validated in accordance with the procedures specified in Appendix A to the Protocol.

b. Exclusions based on the results of applying search terms to document text shall be validated in accordance with the procedures specified in Appendix A to the Protocol.

B. **Review for Responsiveness – Initial Disclosures**

At the outset of the review for responsiveness, the Responding Party shall disclose:

1. The identity and general character of the processes (both TAR and non-TAR) to be used in assessing the documents in the Review Set for responsiveness, noting, for each process identified, the specific parts of the Review Set to be reviewed using that process;

2. The identity and qualifications of the individuals accountable for the results of the review processes, including, for each individual identified, the specific processes, or parts of processes, for which the individual is accountable.

C. **Review for Responsiveness – Training Documents**

1. The Responding Party shall not be required to disclose documents used as training material for its review processes (whether manual or technology-assisted review processes) and is not obliged to claim such documents as privileged.

2. The Requesting Party, if it has access to documents it believes will be included in the Review Set (or to documents it believes are similar to documents that will be included in the Review Set), shall be permitted to provide up to 500 such documents (along with its assessments of the responsiveness of the documents) to the Responding Party

for potential inclusion in training sets. The Responding Party is not obligated to include such documents in its training sets (or, if it includes them, to use the responsiveness designations supplied by the Requesting Party) but is obligated to disclose whether it included them.

D. Review for Responsiveness – Validation

At the conclusion of its assessment of the responsiveness of the documents in the Review Set, the Responding Party shall validate the results of its review processes in accordance with the procedures specified in Appendix B to the Protocol.

E. Documents Subject to Privilege or Other Protections

1. In accordance with the provisions included in Appendix C to the Protocol, any production of ESI made pursuant to the Protocol, whether an interim production or a final production, shall not be deemed the waiver of any privilege.
2. The Responding Party shall disclose the identity and general character of the processes used in creating logs of documents to be withheld on grounds of privilege (or other grounds). Included in the disclosure, the Responding Party shall describe the methods used to validate the results of such processes and summarize the results of those tests.

F. Resolution of Disputes

The Parties shall endeavor to resolve any disputes in a manner that:

1. Furthers the realization of an effective response to the Requesting Party's discovery requests;
2. Respects the Responding Party's autonomy in deciding the manner best suited to respond to the requests for production;
3. Is consistent with the principle of proportionality;
4. Prioritizes evidence-based resolution of questions;
5. Makes, as appropriate, early use of expert support;
6. Is efficient in its use of the Court's time; and
7. Complies in all respects with this Court's order, rules, and procedures pertaining to the resolution of discovery disputes. This includes the obligation to meet and confer with each other before filing any motion with the Court seeking relief.

Appendix A: Validation of Exclusionary Steps

An exclusionary step is the identification of a set of documents to be excluded from the Review Set, whether based on the results of applying search terms to the text of documents or based on the values of metadata fields. Validation of an exclusionary step involves two components: (a) a quantitative component, in which the number of responsive documents in the subset designated for exclusion is compared with the number in the subset designated for inclusion in the Review Set, and (b) a qualitative component, in which the nature of any responsive documents designated for exclusion is assessed.

If the quantitative component shows that the number of responsive documents designated for exclusion is small relative to the number to be included in the Review Set, *and if* the qualitative component shows that any responsive documents designated for exclusion provide information that is either marginal to the issues being litigated or is recoverable from other documents that are to be included in the Review Set, the exclusionary step shall be considered sufficiently validated and may be applied in defining the Review Set.

This appendix specifies the procedures to be followed in validating exclusionary steps as well as the disclosures expected at the conclusion of the validation exercise. It does so first for exclusions based on the application of search terms and then for exclusions made on the basis of metadata values.

Validating Exclusions based on Search terms

Quantitative Assessment

1) Define the test domain. Apply the exclusionary step(s) under evaluation to the Collected Set.
    a) Define the *Positive Set* as the set of documents to be retained in the Review Set, together with any associated family members of such documents; the number of documents in this set will be denoted as N_+.
    b) Define the *Negative Set* as the remainder of documents in the Collected Set (i.e., the set of documents designated for exclusion from the Review Set); the number of documents in this set will be denoted as $N_\circ$.
2) Draw and review the validation sample.
    a) Draw the Positive Sample. From the Positive Set, draw a simple random sample of 400 documents; this will be called the *Positive Sample*, and the number of documents in this sample will be denoted as n_+.

b) Draw the Negative Sample. From the Negative Set, draw a simple random sample of 6,000 documents; this will be called the *Negative Sample* and the number of documents in this sample will be denoted as $n_{\circ}$.

c) Create the Validation Sample. Combine the documents in the Positive and Negative Samples into a single sample, to be called the *Validation Sample*, with no indication either of the set (Positive, Negative) from which the documents were drawn or of any prior coding (responsive, non-responsive) associated with the documents.

d) Review the Validation Sample. Conduct a manual review of the entire Validation Sample, coding each document in the sample as responsive or non-responsive; this review will be called the *Validation Review*.

3) Obtain an estimate of the number of responsive documents in the Positive Set.

a) Using the results of the Validation Review (i.e., the results from Step 2(d)), find the number of documents from the Positive Sample that the Validation Review found to be responsive; this number will be denoted as r_{+}.

b) Using the numbers obtained in Steps 1(a), 2(a), and 3(a) (i.e., the values for N_{+}, n_{+}, and r_{+}), calculate the estimated number of responsive documents in the Positive Set; this number will be denoted as t_{+}.

i) Find the estimated proportion of responsive documents in the Positive Set (p_{+}).

$$p_{+} = \frac{r_{+}}{n_{+}}.$$

ii) Find the estimated number of responsive documents in the Positive Set.

$$t_{+} = p_{+} \times N_{+}.$$

4) Obtain an estimate of the number of responsive documents in the Negative Set.

a) Using the results of the Validation Review (i.e., the results from Step 2(d)), find the number of documents from the Negative Sample that the Validation Review found to be responsive; this number will be denoted as $r_{\circ}$.

b) Using the numbers obtained in Steps 1(b), 2(b), and 4(a) (i.e., the values for $N_{\circ}$, $n_{\circ}$, and $r_{\circ}$), calculate the estimated number of responsive documents in the Negative Set; this number will be denoted as $t_{\circ}$.

i) Find the estimated proportion of responsive documents in the Negative Set ($p_{\circ}$).

$$p_{\circ} = \frac{r_{\circ}}{n_{\circ}}.$$

ii) Find the estimated number of responsive documents in the Negative Set.

$$t_{\circ} = p_{\circ} \times N_{\circ}.$$

5) Compare $t_{\circ}$ to t_{+}. If the estimated number of responsive documents in the Negative Set ($t_{\circ}$) is small relative to the number of responsive documents in the Positive Set (t_{+}), where a reasonable *prima facie* threshold for "small" is 10% or less of the number of responsive documents in the Positive Set, the Parties agree, and the Court should find, that the exercise provides empirical support for the proposition that the exclusionary step is meeting its intended objective of reducing the size of the Review Set without any meaningful loss of responsive information.

Qualitative Assessment

The purpose of the qualitative assessment is to supplement the results of the quantitative assessment with an evaluation of the *nature* of the documents designated for exclusion. The procedures for this assessment are as follows.

1) Review all documents from the Negative Sample found to be responsive and assess them against the criteria of *importance* (Do they contribute information that may help decide issues being litigated?) and *uniqueness* (Do they contribute novel information that is not recoverable from documents that are retained in the Review Set?).
2) If none of the responsive documents identified in the Negative Sample are both important and unique, the Parties agree, and the Court should find, that the exercise provides empirical support for the proposition that the exclusionary step reduces the size of the Review Set without meaningful loss of responsive information.
3) If one or more of the responsive documents in the Negative Sample is both important and unique, that will indicate that the exclusionary step is not meeting its intended objective and should either be modified in some way (e.g., the scope of the search terms should be expanded) or not taken at all.

Disclosures

If the Responding Party believes that the validation exercise provides evidence that the exclusionary step is effective and therefore can be used in defining the Review Set, it will provide the Requesting Party with the following information from the validation exercise.

1) N_{+}, n_{+}, and r_{+}; i.e., the number of documents in the Positive Set, the number of documents in the Positive Sample, and the number of responsive documents found in the Positive Sample.
2) $N_{\circ}$, $n_{\circ}$, and $r_{\circ}$; i.e., the number of documents in the Negative Set, the number of documents in the Negative Sample, and the number of responsive documents found in the Negative Sample.
3) Copies of any non-privileged responsive documents found in either the Positive Sample or the Negative Sample, with those copies being provided in the most convenient format (typically digital).

Validating Exclusions based on Metadata values

The procedures and disclosures for validating exclusions made based on metadata values are the same as those for validating exclusions made on the basis of the results of applying search terms to document text, with the following exceptions.

1) When a metadata-based exclusion is designed for application to a specific subset of the Collected Set (e.g., when a date exclusion is designed for application to data recovered from a particular custodian or set of custodians), the Responding Party may choose to conduct a validation exercise that is focused on that subset.
2) For a validation focused on an exclusion based on metadata values, the default size of the Negative Sample is 1,200 documents. (The default size of the Positive Sample is 400 documents, as it would be for the validation of search terms.)

Appendix B: Validation of Review Processes

The validation of a review process involves two components: (a) a quantitative component, in which an estimate of the *recall* achieved by the review process is obtained, and (b) a qualitative component, in which the nature of the responsive documents missed by the review process is assessed.

The review process can be considered to have been reasonably effective if (1) the quantitative component shows that the review process has captured a high percentage of the responsive documents in the Review Set, *and* (2) the qualitative component shows that such documents as are missed by the review process provide information that is either marginal to the issues being litigated or is recoverable from other documents that are captured by the review process, the review process can be considered to have been reasonably effective.

This appendix specifies the procedures to be followed for both components as well as the disclosures expected at the conclusion of the validation exercise.

Quantitative Assessment

1) Define the test domain. Aggregate the results of all review processes applied to the Review Set (for purposes of this appendix, the *Review Process* shall be defined as the aggregate of all review processes applied to the Review Set).
    a) Define the *Positive Set* as the set of documents identified as responsive by the Review Process, together with any associated family members of such documents; the number of documents in this set will be denoted as N_+.
    b) Define the *Negative Set* as the remainder of documents in the Review Set (i.e., the set of documents neither deemed responsive by the Review Process nor associated, by family relation, with a document deemed responsive by the Review Process); the number of documents in this set will be denoted as $N_\circ$.
2) Draw and review the validation sample.
    a) Draw the Positive Sample. From the Positive Set, draw a simple random sample of 400 documents; this will be called the *Positive Sample* and the number of documents in this sample will be denoted as n_+.
    b) Draw the Negative Sample. From the Negative Set, draw a simple random sample of 3,400 documents; this will be called the *Negative Sample* and the number of documents in this sample will be denoted as $n_\circ$.
    c) Create the Validation Sample. Combine the documents in the Positive and Negative Samples into a single sample, to be called the *Validation Sample*, with no indication

either of the set (Positive, Negative) from which the documents were drawn or of any prior coding (responsive, non-responsive) associated with the documents.

d) Review the Validation Sample. Conduct a manual review of the entire Validation Sample, coding each document in the sample as responsive or non-responsive; this review will be called the *Validation Review*.

3) Obtain an estimate of the number of responsive documents in the Positive Set; this is the number of *True Positives* yielded by the Review Process.

a) Using the results of the Validation Review (i.e., the results from Step 2(d)), find the number of documents from the Positive Sample that the Validation Review found to be responsive; this number will be denoted as r_+.

b) Using the numbers obtained in Steps 1(a), 2(a), and 3(a) (i.e., the values for N_+, n_+, and r_+), calculate the estimated number of responsive documents in the Positive Set; this number will be denoted as t_+.

i) Find the estimated proportion of responsive documents in the Positive Set (p_+).

$$p_+ = \frac{r_+}{n_+}.$$

ii) Find the estimated number of responsive documents in the Positive Set.

$$t_+ = p_+ \times N_+.$$

4) Obtain an estimate of the number of responsive documents in the Negative Set; this is the number of *False Negatives* yielded by the Review Process.

a) Using the results of the Validation Review (i.e., the results from Step 2(d)), find the number of documents from the Negative Sample that the Validation Review found to be responsive; this number will be denoted as $r_\circ$.

b) Using the numbers obtained in Steps 1(b), 2(b), and 4(a) (i.e., the values for $N_\circ$, $n_\circ$, and $r_\circ$), calculate the estimated number of responsive documents in the Negative Set; this number will be denoted as $t_\circ$.

i) Find the estimated proportion of responsive documents in the Negative Set ($p_\circ$).

$$p_\circ = \frac{r_\circ}{n_\circ}.$$

ii) Find the estimated number of responsive documents in the Negative Set.

$$t_\circ = p_\circ \times N_\circ.$$

5) Calculate an estimate of the recall achieved by the Review Process.

$$Recall = t_+/(t_+ + t_\circ).$$

6) Evaluate the recall estimate. If the estimate is high, where a reasonable *prima facie* threshold for "high" is 75% or greater, the Parties agree, and the Court should find, that the

validation exercise provides empirical support for the proposition that the Review Process has achieved a reasonably complete retrieval of documents responsive to the operative document requests.

Qualitative Assessment

The purpose of the qualitative assessment is to supplement the results of the quantitative assessment with an evaluation of the *nature* of any documents missed by the Review Process. The procedures for this assessment are as follows.

1) Review all documents from the Negative Sample found to be responsive and assess them against the criteria of *importance* (Do they contribute information that may help decide issues being litigated?) and *uniqueness* (Do they contribute information that is not recoverable from documents that are captured by the Review Process?).
2) If none of the responsive documents identified in the Negative Sample are both important and unique, the Parties agree, and the Court should find, that the exercise provides empirical support for the proposition that the Review Process has achieved a reasonably complete retrieval of documents responsive to the operative document requests.
3) If one or more of the responsive documents in the Negative Sample is both important and unique, that will indicate that the Review Process may need to be supplemented to account for such documents.

Disclosures

If the Responding Party believes that the validation exercise provides evidence that the Review Process has been effective and therefore can be used in defining the set of documents, pending review for privilege, to be produced to the Requesting Party, it will provide the Requesting Party with the following information from the validation exercise.

1) N_+, n_+, and r_+; i.e., the number of documents in the Positive Set, the number of documents in the Positive Sample, and the number of responsive documents found in the Positive Sample.
2) $N_\circ$, $n_\circ$, and $r_\circ$; i.e., the number of documents in the Negative Set, the number of documents in the Negative Sample, and the number of responsive documents found in the Negative Sample.
3) Copies of any non-privileged responsive documents found in either the Positive Sample or the Negative Sample, with those copies being provided in the most convenient format (typically digital).

Appendix C: On disclosure of privileged material

1) The production of privileged or work-product protected documents, ESI, or information, whether inadvertent or otherwise, is not a waiver of the privilege or protection from discovery in this case or in any other federal or state proceeding. This provision shall be interpreted to provide the maximum protection against waiver allowed by Federal Rule of Evidence 502(d).

2) Nothing contained herein is intended to or shall serve to limit a party's right to conduct a review of documents, ESI, or information (including metadata) for relevance, responsiveness and/or segregation of privileged and/or protected information before production

3) The parties intend the provisions of this document shall displace the provisions of Federal Rule of Evidence 502(b)(1) and (2). That is, all documents and other information produced in accordance with this document shall be considered to have been produced inadvertently and the Responding Party shall be deemed to have taken reasonable steps to prevent disclosure.

Commentary

Double underlining. Example of hyperlinked comment. (Click on lemma to return to the location of the prompt in the text of the Protocol.)

Protocol. This is, of course, not the first ESI protocol developed. Many others have been developed over the years, most in the context of specific litigation, but some also as models for use in a range of different circumstances. Among those we would note are the protocols adopted in *In Re: Actos (Pioglitazone) Products Liability Litigation*, MDL No. 6:11-md-2299 (W.D. La., Jul. 2, 2012); *Rio Tinto Plc v. Vale S.A.*, 14-Civ. 3042 (RMB)(AJP) (S.D.N.Y. Mar. 2, 2015); *In Re: Blue Cross Blue Shield Antitrust Litigation*, MDL No.: 2406, Master File No.: 2:13-CV-20000-RDP (N.D. Ala., Southern Division, Apr. 16, 2015); *In Re Broiler Chicken Antitrust Litigation*, Case No. 1:16-cv-08637 (N.D. Ill., Eastern Division, Jan. 3, 2018); and *In Re: Valsartan, Losartan, and Irbesartan Products Liability Litigation* (D.N.J. Dec. 2, 2020). We would also note the model protocols proposed in D. Md. 2007; W.D. Wash. 2020; Grossman & Cormack 2021; Lau & Lee 2017; DOJ 2018; Payne & Six 2020; and Ball 2023. Among these previous protocols, the ones most closely aligned in approach with the model protocol proposed here, and from which we have drawn, are the *Broilers* protocol and that proposed by Payne & Six 2020. Ball 2023 is in many ways complementary to the model proposed here. Whereas our protocol focuses on the review step (and how to gain trust in the results of that step); the Ball protocol covers technical questions related to the processing and production of ESI in some depth; practitioners may benefit from having Ball 2023 available for reference.

Document. This definition is adapted from that provided in Sedona 2020.[15] It should be noted that this definition includes not only the "canonical" types of communications (emails, word documents, spreadsheets, etc.) that until recently have made up the bulk of discoverable ESI, but also types that are now becoming increasingly favored in the workplace (text messages, both ephemeral and not, social networking posts, communications on collaboration platforms, etc.).

It should further be noted, with regard to the question of what constitutes a document subject to discovery, that the Protocol is designed to be adaptable across different matters and over time. The landscape for data subject to discovery is rapidly changing and is introducing challenging questions with regard to the cost and feasibility of recovering different types of data. While the Protocol cannot provide a one-size-fits-all answer to all such questions, it does require that practitioners, in adapting the model to their specific circumstances, resolve their disputes in keeping with the principles of evidence-based decision making, early use of expert support, and proportionality (see Protocol II.F).

Attachment. This definition is adapted from that provided in Sedona 2020.[16] It should be noted that the relation "associated with" in this definition includes relations in which a copy of a file is coupled with

[15] Sedona 2020: 299: "**Document (or Document Family)**: A collection of pages or files produced manually or by a software application, constituting a logical single communication of information, but consisting of more than a single stand-alone record. Examples include a fax cover, the faxed letter, and an attachment to the letter, the fax cover being the 'Parent,' and the letter and attachment being a 'Child.'" Also, Grossman & Cormack 2013: 14: "**Document**: In the context of Electronic Discovery, a discrete item of Electronically Stored Information that may be the subject or result of a search or review effort."

[16] Sedona 2020: 270*f.*: "**Attachment**: A record or file associated with another record for the purpose of retention, transfer, processing, review, production, and routine records management. There may be multiple attachments

another for purposes of sending or storage (the familiar form of an email attachment) as well as relations in which a file is linked to another via a URL included in the parent file. In the case of a document shared via URL, the pertinent version should, to the extent it can be recovered (and recovered with due regard given to considerations of proportionality), reflect the state of the document at the time at which it was shared.[17]

As observed with regard to the definition of "document," it should be noted, with regard to the question of what constitutes an attachment subject to discovery, that the landscape for data subject to discovery is rapidly changing and is introducing challenging questions with regard to the cost and feasibility of recovering different types of data. While the Protocol cannot provide a one-size-fits-all answer to all such questions, it does require that practitioners, in adapting the model to their specific circumstances, resolve their disputes in keeping with the principles of evidence-based decision making, early use of expert support, and proportionality (see Protocol II.F). In the case of questions as to what constitutes a discoverable attachment, adherence to these principles will mean sampling of the data types at issue, expert testimony on cost and feasibility questions, and an informed application of proportionality.

Collected Set. The focus of this protocol is on steps subsequent to the collection of documents potentially subject to a review for responsiveness (i.e., subsequent to the identification of the Collected Set). The same principles that inform the Protocol's provisions, however, should guide the parties in discussing issues related to collection (see the Protocol's provisions under II.F on *Resolution of Disputes*). More specifically, in keeping with the principle of transparency, the Responding Party should disclose the terms (dates, custodians, sources, etc.) that define the scope of its collection efforts. To the extent that there are disputes about the scope of collection, the parties should seek solutions that (a) give due weight to considerations of proportionality and (b) are informed by both empirical evidence (e.g., samples that show whether or not a given source is likely to contain meaningful amounts of responsive information) and expert guidance (e.g., subject-matter experts who are familiar with the information-governance policies operative in the organization from which data is being collected).

Deduplication. A method of volume reduction that achieves its result by using hash values to identify sets of exact duplicate documents and then retaining for review a single instance of each set. Identifiers (e.g., custodian names) of the sources of the members of the set other than that single instance are retained in, and retrievable from, metadata records.

De-NISTing. A method of volume reduction that achieves its result by excluding from further review a range of file types that, given their function, are unlikely to have any content responsive to a production request (system files, executable files, etc.). The method identifies such files based on the Reference Data Set that NIST maintains as a component of its National Software Reference Library (NSRL).

Email threading. A method of volume reduction that achieves its result by grouping emails that are part of the same conversation and then retaining for review just the inclusive node in each conversation (the email that contains all other nodes in the conversation), setting aside from review non-inclusive nodes in

associated with a single 'parent' or 'master' record. In many records and information management programs or in a litigation context, the attachments and associated record(s) may be managed and processed as a single unit. In common use, this term often refers to a file (or files) associated with an email for retention and storage as a single message unit."

[17] Questions around the treatment of linked documents are still not fully settled. For an opinion that holds that linked documents should *not* be treated as attachments, see *Nichols v. Noom, Inc*., No. 20-CV-3677 (LGS) (KHP) (S.D.N.Y. Mar. 11, 2021).

the conversation (emails that are duplicated (with respect to both text and attachments) in the inclusive node).

Metadata values. Under this approach, the values in the metadata fields associated with a document (e.g., the values in a date field) are used to identify documents that can be set aside from further review as being presumptively non-responsive. For example, if a given custodian was not engaged in the processes or transactions at issue in the litigation until a certain date, emails collected from that custodian that had a *CreationDate* value prior to that date could be candidates for exclusion from the Review Set.

Search terms. Under this approach, a party develops a set of Boolean queries (search terms, often taking the form of compound expressions in which multiple search terms are combined using logical operators (e.g., "and", "or") as well as proximity operators (e.g., "within five tokens")) that can be applied to the Collected Set either to identify documents that should be included in the Review Set (and to exclude from review documents not so identified) or to identify documents that are candidates for exclusion from further review.

***A priori* considerations.** The "*a priori* considerations" referred to here include any characteristics of a specific subset of data that would permit the formation of a reasonable expectation as to the likelihood that the subset would contain responsive information significant in either quantity or quality. Examples of such characteristics include date associated with a document, the functional role of the custodian holding the document, the use to which a given shared drive was put, and so on.

Outset of the review. It should be noted that the Protocol requires the Responding Party to make the disclosures specified under this provision at the *outset* of the review (before results have been obtained). If, once the review is under way, the Responding Party decides to change either review processes or review operators, the Responding Party should provide the Requesting Party with updated information in a timely manner and, in providing the update, should include the reasons for the change.

Identity and general character. The purpose of this disclosure is *not* to demonstrate that the review processes to be deployed by the Responding Party *will* be effective; that cannot be done before the fact (hence the need for validation at the conclusion of the process). The purpose of this disclosure is simply to establish that there are reasonable grounds for believing that the processes to be deployed by the Responding Party could be effective. This can be accomplished by providing enough information to show that the processes have been found to be effective in other circumstances (where the "other circumstances" could be other matters or scientific evaluations).

More specifically, what is expected in an "identity and general character" disclosure is, for each review process to be deployed, the following.

1) **The name of the tool or service to be used.** This should include, for vendor-provided tools or services, the name of the vendor providing the tool or service.

2) **A high-level characterization of the nature of the review process.** What is required here is not a detailed description of the process; it will suffice simply to indicate where the process in question fits in the taxonomy of review processes currently available for legal discovery. For example: *Is it manual or technology-assisted? If technology-assisted, is it statistical or rules-based? If the process employs machine learning, does it use active or passive approaches to the identification of new training data? Are the training and review phases separate or continuous? And so forth.* In most cases, it should be possible to provide the required information in one page or less.

3) **A reference to venues in which the effectiveness of the process has been demonstrated.** What is required here is not a detailed description of an evaluation; what is required is sufficient information to provide assurance that the process is not untested. For this purpose, it will suffice to indicate, for example, that the process has been deployed and found effective in some number of other matters or to indicate that the process was included in a reputable independent evaluation and found to have achieved reasonable levels of recall.

As stated in the provision, if multiple review processes are to be used, the disclosure should include an indication of the parts of the Review Set to which a given process will be deployed. For example, if the primary review process will be a technology-assisted one, but, for images with little usable text, a manual review process will used, that should be indicated in the disclosure. Similarly, if the OCR quality of documents sourced from hard-copy originals is such as to make the documents unsuitable for technology-assisted review (and therefore subject to inclusion in a manual review process), that should be indicated in the disclosure.[18]

Identity and qualifications. The purpose of this disclosure is to provide assurance that the individuals responsible for the operation and testing of the review processes have the skills and experience needed to carry out their roles effectively. What is expected is not a detailed description of the experience and qualifications of the individual in question; a summary of relevant academic qualifications and/or work experience will typically serve the purpose. An up-to-date CV, for example, should in most cases provide the expected information.

A disclosure of "identity and qualifications" is not expected for every individual who will be engaged in the operation and testing of the review process. The provision requires disclosure only for those in senior oversight roles who can be held ***accountable*** for the results of the review.

Access. The purpose of this provision is to enable the Requesting Party to bring to the Responding Party's attention specific documents, or types of documents, that it believes should be a focus of the review effort (and so, in the Requesting Party's opinion, should be included in the data used to train both manual and technology-assisted review processes). The provision permits the Requesting Party to select, from any documents to which it already has access (e.g., documents already available from productions made in related litigation or from data supplied by a witness or whistleblower), any documents it wishes for this purpose. One way in which this provision could be used productively would be for the Requesting Party to provide instances of documents it believes are likely to be of high importance but low frequency in the Review Set. By drawing attention to such documents, the Requesting Party puts the Responding Party on notice that such documents should not be overlooked in the training and conduct of the review. Of course, as stipulated in the provision, it is up to the Responding Party to decide whether to act on any such notice; a failure to act will, however, open the Responding Party to a potential challenge if identifiable gaps in the eventual production are found.

Similar to documents. There will, of course, often be circumstances in which the Requesting Party does not have access to documents that will actually be included in the Review Set. In such circumstances the Requesting Party may (a) provide the Responding Party with exemplar documents, to which it does have access, that it believes are similar to documents in the Review Set, (b) provide synthetic documents that

[18] This example is not to be understood as implying that OCRed documents are, by definition, unsuitable for TAR; TAR processes, with effective feature engineering, can be robust against OCR errors (see O'Halloran et a. 2023).

have been created for the express purpose of modeling the types of documents in which the Requesting Party is interested, or (c) provide descriptions of the documents, or types of documents, that it believes should be a focus of the review effort.[19]

Up to 500. The upper limit specified, 500 documents, should suffice, in most cases, to enable the Requesting Party to provide a set of training exemplars that represent a broad range of topics, language, and document types. A set of that size should also not be an undue burden on the Responding Party. Of course, if the circumstances of a specific matter so warrant, a higher or lower upper limit may be set.

In accordance with the procedures. For discussion of alternative sequencing of the validation exercises for exclusionary steps and the review process, see the Comment on *Validated in accordance with the procedures* above.

Identity and general character. As was the case with the "identity and general character" disclosure regarding review processes, the purpose of this disclosure regarding processes for creating privilege logs is simply to establish that there are reasonable grounds for believing that the processes to be deployed by the Responding Party could be effective (i.e., provide information that was both accurate and sufficiently granular to serve the purposes of a privilege log). This can be accomplished by providing (1) a high-level characterization of the nature of the logging process (*Is it manual or automated? If automated, what is the name of the tool or service to be used?*) and (2) a reference to venues in which the effectiveness of the process has been demonstrated.

Methods used to validate the results of such processes. The purpose of this disclosure is to provide transparency into the methods used to validate the effectiveness of the processes used to create logs of documents withheld on grounds of privilege (or other grounds). No more than a high-level summary is required. The Protocol does not specify that a particular validation protocol be followed; it does set the expectation that meaningful testing of the methods used for creating privilege logs will be carried out.

Results of those tests. No more than a high-level summary of results is required. In most cases, such a summary will involve little more than numbers that reflect the results of the tests. No disclosure of actual documents is expected as part of this summary.

Interim production. The goals of discovery can be advanced by an allowance for "interim" (or "rolling") productions, partial productions of documents made from some subset of the Review Set and made prior to completing review of the full Review Set. The purpose of this provision of the Protocol is to ensure that such productions can be made without introducing the risk that doing so might, if it occasioned the production of a privileged document, result in a waiver of privilege.

Effective. In the case of a response to a request for production, effectiveness is a measure of the extent to which the response has met the standard set in the operative rules of procedure (e.g., the "reasonable inquiry" standard set forth in Fed. R. Civ. P. 26(g)(1)(B)). In civil proceedings conducted in federal courts, the effectiveness of a response is evaluated against the standard of a reasonable inquiry. Or, to anchor the standard more closely to the precise provisions in the rule: *Is the responsive production the result of a reasonable inquiry that is consistent with the rules and warranted by existing law per Fed. R. Civ. P. 26(g)(1)(B)(i)?*

[19] Of course, the fact that the Requesting Party is able to describe a document (or create an artificial model of a document) does not imply that the document in fact exists.

Viewed in this light, an assessment of the effectiveness of a response to a request for production will almost always be an empirical inquiry, a gathering of both quantitative and qualitative data on the accuracy and completeness of the Responding Party's response to the request. A proposed resolution of a dispute between the parties will further effectiveness if it increases the transparency into this empirical inquiry.

Autonomy. Case law and rules of civil procedure in the US have long, both prior to the advent of TAR and subsequent to its availability and adoption, recognized that the Responding Party should be given considerable latitude in deciding how to conduct its review for material responsive to a request for production. There are multiple reasons for doing so, including the fact that any effective search for responsive documents, whether involving TAR or not, will also require access to large numbers of *non-responsive* documents to which the Requesting Party is not entitled. This latitude is recognized in Sedona's Principle 6 ("Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.")[20] and has been reaffirmed in multiple rulings, including, from 2020, *Livingston v. City of Chicago* ("the court agrees with the City that as the responding party it is best situated to decide how to search for and produce emails responsive to Plaintiffs' discovery requests").[21]

The latitude granted the Responding Party is not, however, without limits. The Responding Party is not absolved from meeting the "reasonable inquiry" standard, *and demonstrating the meeting of that standard*, as specified in the governing rule of procedure (*e.g.*, in the case of proceedings governed by the FRCP, the standard that requires the Responding Party's certification, on the basis of reasonable good-faith inquiry, that its response is consistent with the rules and warranted by existing law).[22] While the Responding Party is granted a reasonable amount of space for the responsible exercise of autonomy in its decisions and actions, those decisions and actions must lead to an effective result and the Responding Party must demonstrate that they have indeed done so.[23]

Many of the discovery disputes that occupy the time of the Parties and the Court stem from the Requesting Party's efforts to curtail the Responding Party's autonomy in decision-making in the hope of gaining greater trust in the results produced. A key objective of this model protocol is to avoid such disputes by putting in place provisions that preserve, for the Responding Party, space for the responsible exercise of autonomy, but also putting in place provisions that require rigorous, meaningful, and transparent validation of the

[20] Sedona 2018. In discussing Principle 6, the authors of the commentary on the principle note additional grounds for the latitude granted to the Responding Party: (1) procedural rules and case law provide that the Responding Party should, at least in the normal course, meet its discovery obligations without direction or intervention from the Court or opposing counsel; (2) decisions about how most efficiently and effectively to respond to a request often require access to information about an organization's data infrastructure and architecture as well as an understanding of information governance (IG) systems and protocol (i.e., information to which the Responding Party has unique access); (3) the Responding Party, in the normal course, will bear the costs of the discovery effort, so should be permitted to decide on the tools and methods used to conduct the effort; and (4) the position articulated in Sedona Principle 6 has repeatedly been endorsed by the courts (see, for example, the rulings cited in footnote 97 (p. 120) of the commentary on the principles).

[21] *Livingston v. City of Chicago*, No. 16 CV 10156 (N.D. Ill. Sep. 3, 2020), p. 4. For additional case law in support of Sedona Principle 6, see the rulings cited in footnotes 92 and 97 of The Sedona Principles.

[22] Fed. R. Civ. P. 26(g)(1)(B)(i).

[23] For an opinion that recognizes *both* the latitude granted to the Responding Party *and* the requirement that the Responding Party's proposed process meet a reasonableness standard, see the *Report and Recommendation* written by Hon. James C. Francis IV (Ret.) serving as Special Master in *In re Diisocyanates Antitrust Litig.*, MDL No. 2862, 2021 WL 4295729 (W.D. Pa. Aug. 23, 2021). Judge Francis's R&R was adopted in its entirety as the opinion of the Court (*In re Diisocyanates Antitrust Litig*., MDL 2862 (W.D. Pa. Sep. 21, 2021)).

results of a review, thus providing the Requesting Party with the information needed for a well-grounded trust in the results.

Proportionality. Proportionality, balancing the value of information gained against the cost of gaining it, has long been an operative consideration in legal discovery, and the weight attached to that consideration has only increased with the advent of advanced review technologies and with the 2015 amendments to the FRCP.[24] Accordingly, the principle of proportionality has shaped the design of the Protocol and is also expected to be a governing consideration for practitioners addressing specific questions related to the implementation of the Protocol in a given matter.

It should also be noted that the principle of proportionality interacts with other requirements set for the resolution of disputes. For example, the requirement for evidence-based resolution of disputes means that questions related to proportionality should be decided, to the extent possible, on the basis of evidence gathered as to the real cost and value of obtaining the information in questions. At the same time, the requirement that evidence be gathered in resolving disputes is limited by the proportionality requirement.

Evidence-based resolution. The intention of this requirement is to ensure that Parties seek to resolve their disputes, where possible, and as early as possible, on the basis of data gathered that is specifically relevant to the dispute and avoid engaging in prolonged discussion about hypothetical scenarios for which there is no evidentiary support.

Expert support. The purpose of this requirement is to enable the expedited resolution of questions and disputes and, ultimately, the expedited arrival at a position of well-grounded trust in the results of a response effort. While some lawyers are reluctant to bring specialized experts to the meet-and-confer table, it is important to recognize that the answers to many of the questions that occur in the course of applying and testing advanced search and review technologies do require specialized expertise (*e.g.*, in machine learning, in information retrieval, in statistics). Bringing individuals with the right kinds of expertise into the discussion sooner can enable a shorter, more direct, and generally less contentious path to a well-informed answer to a question.

The use of expert support is also supported by case law. The importance of expertise in addressing issues raised by the use of advanced technologies in legal discovery was recognized in a number of early cases.[25] In a case involving a dispute about the use of search terms, for example, United States Magistrate Judge John M. Facciola ruled that "[g]iven this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread" (*United States v. O'Keefe*, 537 F. Supp. 2d 14 (D.D.C. 2008)). United States Magistrate Judge Paul W. Grimm, in another case involving a dispute about the use of search terms,[26] similarly recognized the importance of ensuring that a search system be designed and implemented by

[24] For the specific considerations that figure into a proportionality evaluation, see Fed. R. Civ. P. 26(b)(1): "Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, **considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit**" (emphasis added). See also the limitations on discovery identified in 26(b)(2)(C). For a discussion of the application of the principle of proportionality in e-discovery, in light of the 2015 amendments, see Sedona 2017.

[25] In this comment, we have drawn from the useful summary of court rulings with regard to expertise and the use of advanced technologies provided by Losey 2018.

26 Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260, 262 (D. Md. 2008).

individuals with the right qualifications: "[s]election of the appropriate search and information retrieval technique requires careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected should be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented." In a more recent case,[27] United States Magistrate Judge James C. Francis IV ruled against a party because of its failure to provide the needed expert support for a motion: "[t]he defendants have not provided the necessary expert opinions for me to assess their motion to compel search terms. The application is therefore denied."

The competence to identify the circumstances in which specialized scientific or technological expertise is required and to ensure that, in those circumstance, the appropriate kind of expertise is readily accessible is also increasingly recognized as a component of the professional ethics of attorneys. In Comment 8 to Rule 1.1[28] of its *Model Rules of Professional Conduct*, the ABA notes that "[t]o maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, **including the benefits and risks associated with relevant technology**, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject" (emphasis added).[29] State bar associations are beginning to follow suit. In 2015, the Standing Committee on Professional Responsibility and Conduct of the State Bar of California published Formal Opinion No. 2015-193, in which it, after discussing the ethical issues raised in a hypothetical in which an attorney fails to recognize the need for specialized expertise in addressing questions related to e-discovery, concluded that "[a]ttorneys who handle litigation may not ignore the requirements and obligations of electronic discovery. Depending on the factual circumstances, a lack of technological knowledge in handling e-discovery may render an attorney ethically incompetent to handle certain litigation matters involving e-discovery, absent curative assistance under rule 3-110(C), even where the attorney may otherwise be highly experienced."[30] The curative assistance to which the opinion refers consists of filling in gaps in competence by drawing on the support of an expert who does have the requisite competence (or by acquiring the requisite competence prior to performing the tasks for which it is required): "If a lawyer does not have sufficient learning and skill when the legal services are undertaken, the lawyer nonetheless may provide competent representation by (i) associating with or, where appropriate, professionally consulting another lawyer whom the lawyer reasonably believes to be

[27] *Greater New York Taxi Association v. City of New York*, No. 13 Civ. 3089 (VSB) (JCF) (S.D.N.Y. Sept. 11, 2017).
[28] "Rule 1.1: Competence. Client-Lawyer Relationship. A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation" (ABA Rules of Prof'l Conduct R.1.1).
[29] ABA Rules of Prof'l Conduct R.1.1 cmt. 8. More recently (2019), the ABA, in a resolution on the ethical use of AI-enabled systems in legal practice, further underscored competence as a condition for ethical use of AI-enabled systems in legal practice. The text of the resolution is: "RESOLVED, That the American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ('AI') in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI" (ABA House of Delegates, Resolution 112, adopted August 12-13, 2019). Commentary on the resolution's implications for a lawyer's duty of competence and duty to supervise can be found in the report accompanying the resolution (pp. 5-7). For further discussion of the resolution, see Economou 2019.
[30] The State Bar of California, Standing Committee on Professional Responsibility and Conduct, Formal Opinion No. 2015-193 (2015: 7).

competent, (ii) acquiring sufficient learning and skill before performance is required, or (iii) referring the matter to another lawyer whom the lawyer reasonably believes to be competent."[31]

Finally, the International Organization for Standardization (ISO), in a standard that provides a framework for evaluating tools and methods applied to legal discovery, emphasizes the importance of giving due attention to the various types of specialized expertise that may need to be called upon in conducting an effective review: "[a]n ESI review, regardless of a specific approach, is fundamentally an information retrieval exercise; an effective ESI review will therefore draw upon, as appropriate, the kinds of expertise that are brought to bear in information retrieval science (computer science, statistics, linguistics, etc.)".[32] Needless to say, it would not be reasonable to expect that lawyers themselves acquire all the requisite scientific and technological skills; they should, however, be competent to identify when those skills are needed and, when they are needed, have access to individuals with the appropriate expertise[33] and bring that expertise to bear in expeditiously resolving disputes with opposing counsel.

For these reasons, the Protocol sets the expectation that, when information is exchanged between the Parties, both the requests for information and the responses will, when appropriate, be provided, or at least reviewed for relevance and accuracy, by individuals competent in the relevant areas of expertise.[34]

Efficient. "Efficient" recourse to the Court's assistance, but also recourse that is effective at getting to the truth of the matter in dispute and reaching a fair resolution.

The Court's time. When appropriate, the Parties and the Court may find that the appointment of a Special Master charged with overseeing discovery proceedings helps the Parties to resolve disputes in a manner that is both efficient and just.

Procedures to be followed (for exclusionary steps). The Protocol assumes, as a default, a validation model in which exclusionary steps are validated at the time of their execution (Appendix A) and then a separate validation exercise is conducted for the review process (Appendix B). While this model, being in accord with the typical workflow of a response to a discovery request and allowing timely and focused testing (and, if needed, remediation) of each of the two steps (exclusions, review), will likely be preferred in most cases, it is not the only model. In some instances, for example, a party may prefer a model in which the validation of exclusionary steps is carried out simultaneously with the validation of the review process. In other instances, a party (or the Court) may prefer that the validation exercises be combined to generate

[31] The State Bar of California, Rules of Prof'l Conduct R.1.1(c)). (Rule 3-110(C), referenced in the 2015 Formal Opinion, has since (2018) been replaced by Rule 1.1(c).) For additional discussion of technological competence as a component of professional ethics, see Lat 2020; Walker 2018; Pizzi & Brickell 2014.

[32] ISO 2017: 19. This standard has sometimes received less than favorable welcome by practitioners who see its requirements as displacing, to negative effect, the legal judgment of attorneys and the norms of professional ethics. While such concerns are understandable, a reading of such requirements as are actually contained in the standard will find that those requirements are of such a general nature that they serve more as indicators of topics to consider when evaluating a review process than as narrowly prescriptive requirements to which a party must adhere. If practitioners view the standard in that light, they may find it offers a useful framework for evaluating e-discovery tools and methods.

[33] Such individuals may be in-house experts, specialists provided by a technology vendor, or independent consultants with the appropriate expertise in science, statistics, technology, and the law.

[34] There may also be occasions in which the Court, recognizing the need for expert support, appoints a Special Master (or other expert) to assist in resolving some of the scientific or technical issues raised by the use of advanced discovery technologies. *Rio Tinto Plc v. Vale S.A.*, *In Re Broiler Chicken Antitrust Litigation*, and *In re Diisocyanates Antitrust Litig.* are all examples of cases in which the Court made effective use of a Special Master to address issues related to the use and validation of TAR.

a single "end-to-end" estimate of recall.[35] Such alternative models, like the default, will still meet the objective of generating empirical evidence that will enable a sound assessment of the effectiveness of the processes in question. The Protocol is easily adapted to accommodate them.[36] Whether a practitioner prefers the default or one of the alternatives is largely a matter of pragmatic considerations (for an overview of some of these practical considerations, see the relevant discussion in the Guidelines, Section 1.2 (*On the Validation of Exclusionary Steps*), under the heading *Why not a recall number?*).

Step(s). The Responding Party may choose to conduct a single aggregate test of the result of applying multiple exclusionary steps (e.g., the result of applying multiple metadata-based exclusions or the result of both a metadata-based exclusion and a search-term-based exclusion). In the case of such an aggregate test, the Positive Set would be defined as the set of documents to be retained in the Review Set (together with any associated family members of such documents), **after applying all the exclusionary steps that are being evaluated in the exercise**. The Negative Set would then be simply the remainder of documents in the Collected Set. While such aggregate testing will provide actionable information, it should be noted that, in the case of metadata-based exclusions, tests focused on a specific exclusion may be more efficient and informative; see the Comment *Focused on that subset* below.

Family members. A "family" is defined as the set of documents formed from a parent document and any associated attachments.[37] A "family member" is any member of that set (whether a parent or an attachment).

Simple random sample. A simple random sample is one that has been drawn in such a manner that the specific combination of documents that have been drawn into the sample had the same probability of being selected as every other possible combination of documents of the same sample size. Textbook definitions are more precise: "Simple random sampling, or random sampling without replacement, is a sampling design in which *n* distinct units are selected from the *N* units in the population in such a way that every possible combination of *n* units is equally likely to be the sample selected."[38] What this means, in terms of sampling requirements, is that (i) the sample be drawn from the source population in such a way that all documents in that population are available for selection into the sample, (ii) all documents in the source population have an equal probability of being selected into the sample, and (iii) a document, once selected, is not available for re-selection.

This definition of a simple random sample has a few practical implications that should be noted. First, the requirement that every possible combination of the target number of documents have an equal likelihood of being the sample selected means that some methods of selecting the sample, even if intuitively attractive,

---

[35] For a case in which the Court ordered such an end-to-end estimate of recall, see *In re Diisocyanates Antitrust Litig*., MDL No. 2862, 2021 WL 4295729 (W.D. Pa. Aug. 23, 2021) and, in particular, the *Report and Recommendation* written by the Special Master in that matter.

[36] In the case of simultaneous validation, the accommodation is simply a matter of timing (the validation protocols remain as specified in Appendices A and B). In the case of obtaining an end-to-end estimate of recall, the accommodation is simply a matter of using the numbers generated by the exercises specified in the appendices (or, more specifically, the outcome of the review of the Negative Sample(s) used for validating exclusionary steps and the outcome of the review of both the Positive and Negative Samples used for validating the review process) as inputs to a stratified estimate of aggregate recall. For more on using stratified sampling designs to obtain aggregate estimates, see the Guidelines, Section 1.4 (*Additional circumstances and metrics*).

[37] For a definition of *attachment*, see comment on Attachment above; note that, in accordance with that definition, documents linked to a parent document via a URL included in the parent document count as an attachment.

[38] Thompson 2002: 11. See also Freedman et al. 2011 "Simple random sampling means drawing at random without replacement" (p. 340); Mendenhall & Beaver 1991: "Simple random sampling gives every different sample in the population an equal chance of being selected" (p. 120).

will not be satisfactory. A method of selection, for example, that specified that every i^{th} document on a list of documents in the source population be selected into the sample would not satisfy the randomness requirement, because, so defined, the selection protocol would *a priori* exclude from selection many otherwise possible combinations of documents (e.g., combinations that allowed for the inclusion of documents adjacent to each other on the list).

Second, not all document review and analysis tools have the functionalities required for the selection of random samples (although increasingly they do) and not all operators of review and analysis technologies have the expertise needed to distinguish a sample that meets the requirements of randomness from one that does not. A responding party should ensure that it has access to the tools and competencies required to meet the randomness requirement and to test that any samples selected do not depart from the randomness requirement.[39]

Third, given that randomness requires that every possible combination of the target number of documents have an equal likelihood of being the sample selected, any method of screening multiple samples, all of which represent the same state of the result set, to find those that will be favorable for purposes of validation will obviously fail to satisfy randomness (since the screening process, by design, excludes samples found to be unfavorable). **Any such screening process would render the results of the validation exercise invalid.**

With regard to the latter observation, if the Requesting Party has reason to believe that the Responding Party may in fact attempt to screen samples to find ones more favorable to it, the Requesting Party may require that the Responding Party disclose, **prior to the review of any sample used for validation purposes**, a list of the IDs of the documents included in the sample. Such a disclosure would serve to "freeze" the sample(s) to be used in the validation exercise, prior to the Responding Party's learning of the specific contents of the documents included in the sample, thus preventing any screening effort. The Protocol, having been designed on the assumption that most lawyers adhere to the standards of professional ethics, does not include such a disclosure as a default provision, but the Parties could easily add such a provision when it was deemed necessary.

400 documents. The purpose of the Positive Sample is to obtain an estimate of the number of responsive documents in the Positive Set (the set that will be included in the Review Set); this estimate will be used as a reference against which to compare the analogous estimate obtained from the Negative Sample (i.e., the number of responsive documents in the Negative Set). A sample of 400 documents, *regardless of the size of the source population* (i.e., in this case, regardless of the size of the Positive Set),[40] will have the following characteristics.

[39] As an example of a simple test of randomness, one might compare the distribution of documents across custodians in the source population against the same distribution in the sample. There will of course be some differences, but the differences should not be so great that they cannot be attributed to the operation of chance in the sample selection process. A simple chi-squared test can be used to detect departures from what might be attributed to chance.

[40] For more on the relation of population size to the power of a given sample size, see Chapter 2 of the Guidelines accompanying this protocol.

- It will enable the Parties, in gauging the uncertainty associated with the estimate, to obtain a margin of error[41] that is no greater than ± 5%.[42]
- It will almost always[43] include at least one instance of any type of document that is represented in at least 0.75% of the source population (i.e., occurs at a frequency of at least 1 out of every 133 documents).
- It will, in most cases,[44] bring into view a meaningful range of the sorts of responsive documents included in the set designated for downstream review.

Estimated number. The procedures specified in this Appendix provide guidance as to how to arrive at a *point estimate* for the target metric (in this case, the prevalence of responsive documents in the Positive Set). As a practical matter, this is the statistic that is most immediately actionable by practitioners in assessing the results of a validation exercise and will be the focus of attention for both Parties. Hence its calculation is all that is required by the Protocol.

It is always good practice, however, when using statistical estimates, also to calculate a gauge of the sampling-based uncertainty (or *sampling error*) that is associated with a point estimate (i.e., the variability that is a result of the operation of chance in the random selection process). Typically, this is done by the calculation of a *margin of error* or *confidence interval* associated with the estimate.

The procedures for calculating these gauges of uncertainty (which, although involving some steps beyond the calculation of the point estimate, are in fact not overly complex) are provided in the Guidelines accompanying this document. The inputs to these procedures are included in the disclosures already provided for in the Protocol; no additional disclosures are needed. Practitioners seeking to obtain the margin of error associated with an estimate may follow the procedures provided in the Guidelines or may choose to call on the support of individuals with the appropriate statistical expertise (whether these individuals are in-house experts, support specialists provided by a technology vendor, or independent consultants with a practice in legal discovery).

Simple random sample. See earlier comment on *Simple random sample*.

6,000 documents. The purpose of the Negative Sample is to obtain an estimate of the number of responsive documents in the Negative Set (the set that will be excluded from the Review Set); this estimate will be compared against the analogous estimate obtained from the Positive Sample in order to assess the quantitative impact of the exclusion. A sample of 6,000 documents, *regardless of the size of the source*

[41] Calculated at a level of 95% statistical confidence.

[42] To express the size of the margin of error in a manner that allows easy comparison from one circumstance to another, we assume the estimate of responsive documents in the Positive Set is expressed as a percentage (rather than an absolute number). **Note that we do so only for the purpose of enabling a reasonable assessment of the size of the *margin of error*.** When assessing the implications of the quantity of documents missed (or captured) by a process, it is important to express that quantity in terms of numbers of documents. For more on this point, see the Guidelines accompanying this protocol (and, specifically, the discussion under the heading *Why a number, not a percentage?*).

[43] To be specific: 95% of the time.

[44] The exceptions being cases in which the prevalence of responsive material in the Review Set is very low. When such cases occur, they can be addressed by either increasing the size of the Positive Sample or re-visiting the exclusionary technique in question (e.g., search terms) to make the technique more effective at avoiding the inclusion of false positives.

population (i.e., in this case, regardless of the size of the Negative Set),[45] will have the following characteristics.

- It will enable the Parties, in gauging the uncertainty associated with the estimate, to obtain a margin of error[46] that is no greater than ± 1.3%.
- It will almost always[47] include at least one instance of any type of document that is represented in at least 0.05% of the source population (i.e., occurs at a frequency of at least 1 out of every 2,000 documents).
- It will allow a view of the sorts of responsive documents that would be excluded from downstream review (assuming that such documents exist).

These characteristics, together with those of a Positive Sample of 400 documents, are sufficient to enable a meaningful assessment of the impact of the exclusion being evaluated in most circumstances.[48] As a practical matter, moreover, the cost of reviewing samples of the specified sizes, while not insignificant, is not unduly burdensome, given the potential impact of excluding data from further review. If the cost of reviewing a total of 6,400 documents outweighs the savings that would be realized by taking a given exclusionary step (such as using search terms to define the review set), then it is best simply to skip the exclusionary step and include the data in the Review Set. If the savings that would be realized by taking a given exclusionary step do outweigh cost of reviewing 6,400 documents, then taking the exclusionary step is a reasonable means for improving the efficiency of the review (assuming, of course, that the results of the validation exercise support the exclusion).

If a Party has reason to believe that the prevalence of responsive material in the Collected Set is relatively high (e.g., 1% or higher), and so a Negative Sample of lower power (hence smaller size) than that specified in the Protocol would suffice to detect any meaningful number of misses, a party could choose to utilize a smaller Negative Sample (e.g., a sample of 3,000 documents, which would have the power to include at least one instance of any type of document that is represented in at least 0.1% of the source population (i.e., occurs at a frequency of at least 1 out of every 1,000 documents). In making such adjustments, the Party would benefit from expert support.[49]

Combine. The purpose of this step is to ensure that the validation review is "blind," i.e., conducted in a manner that prevents the reviewers from knowing whether a given document was drawn from the Positive Set or from the Negative Set (or had received any prior responsiveness coding), thus shielding the review from the impact of bias. For more on the importance of ensuring that validation reviews be conducted in a blind manner, see Grossman & Cormack 2021.

---

[45] For more on the relation of population size to the power of a given sample size, see Chapter 2 of the Guidelines accompanying this protocol.
[46] Calculated at a level of 95% statistical confidence.
[47] To be specific: 95% of the time.
[48] Of course, as provided for in the Protocol, quantitative measures need to be supplemented by qualitative analysis to obtain a complete view of the impact of a given exclusion.
[49] A competent expert should be able to provide the required guidance at reasonable cost. Of course, one motivation for creating the Protocol (and supporting materials) is to promote a wider distribution of the required skills (or at least a wider distribution of an understanding of what skills are required). To the extent that it is successful at meeting that objective, it should increase the supply of individuals with the requisite expertise and therefore bring down the cost of obtaining it. The Protocol, if successful, should also increase demand, which should further incentivize the acquisition of the requisite skills (skills that, though specialized, are not difficult to acquire), again increasing supply and reducing cost.

Manual review. The review of the documents in the Validation Sample may be carried out by one or more reviewers. The reviewers shall be chosen by the Responding Party and should be attorneys who are familiar with the operative requests for production, with the issues that prompted the requests, and with the subject matter related to those issues.

Estimated number. See earlier comment on *Estimated number*.

Expanded. By identifying search terms (among those already included in the set) that need to be further generalized (to expand coverage) or search terms (not yet included in the set) that need to be added.

Following information. With regard to these disclosures, two observations are in order. First, all that is required to be disclosed under Items (1) and (2) on the list are numbers (they do not require disclosure of documents or any other protected or proprietary information). The only documents that must be disclosed are those specified in Item (3), all of which are **non-privileged responsive** documents (and all of which are disclosed under the protections provided for in Appendix C). **The Protocol does not require the disclosure of non-responsive documents.** Second, the numbers to be disclosed under Items (1) and (2) are just the inputs to the metrics used for evaluating the effectiveness of the search terms; as long as those inputs (six numbers) are disclosed, either party can, with the help of the procedures specified in the Appendix (and, if a party finds it helpful, with the further instruction provided in the accompanying Guidelines), calculate the derived metrics (including the margins of error associated with estimates).

Focused on that subset. Metadata-based exclusions are often designed for application to only a specific subset of the Collected Set (e.g., a date-based exclusion may be intended for a particular custodian or set of custodians, but not for the Collected Set as a whole). When that is the case, we obtain a clearer view of the effect of the exclusion if we focus the validation exercise on just the part of the Collected Set for which the metadata-based restriction is intended. What this means, in practical terms, is that, when a metadata-based exclusion is defined for application to a specific subset of the Collected Set, that subset should define the domain of the validation exercise: the Positive Set will be defined as the set of documents, *within the relevant subset*, designated for inclusion in the Review Set (and the Positive Sample is drawn from the Positive Set so defined) and the Negative Set will be defined as the set of documents, *within the relevant subset*, designated for exclusion from the Review Set (and the Negative Sample is drawn from the Negative Set so defined).

1,200 documents. Because metadata-based exclusions operate via simpler criteria than do search-term-based exclusions and because metadata-based exclusions are often designed for application to only specific subsets of the Collected Set, we can typically gain the information needed to validate a metadata-based exclusion with a smaller Negative Sample than is required for a search-term-based exclusion. The Protocol's default size for the Negative Sample, when validating metadata-based exclusions, is 1,200 documents. A sample of this size, *regardless of the size of the source population* (i.e., in this case, regardless of the size of the Negative Set),[50] will have the following characteristics.

- It will enable the Parties, in gauging the uncertainty associated with the estimate, to obtain a margin of error[51] that is no greater than ± 2.8%.

[50] For more on the relation of population size to the power of a given sample size, see Chapter 2 of the Guidelines accompanying this protocol.
[51] Calculated at a level of 95% statistical confidence.

- It will almost always[52] include at least one instance of any type of document that is represented in at least 0.25% of the source population (i.e., occurs at a frequency of at least 1 out of every 400 documents).
- It will allow a view of the sorts of responsive documents that would be excluded from the Review Set (assuming that such documents exist).

These capabilities, together with those of a Positive Sample of 400 documents, will typically be sufficient to enable a meaningful assessment of the impact of a metadata-based exclusion in most circumstances.[53]

Recall. Recall is a measure of the completeness of a review effort. It answers the question: *Out of all the responsive documents in the Review Set, what percentage did the Review Process succeed in identifying?* The lower the percentage (i.e., the lower the level of recall), the less complete the result set; the higher the percentage (i.e., the higher the level of recall), the more complete the result set.[54]

Procedures to be followed (for review processes). The procedures specified in this appendix are designed for use in the canonical case in which there is a single Review Set and the validation exercise is conducted when the review of that set is complete (i.e., when the Review Set can be exhaustively divided into one Positive Set and one Negative Set). The procedures can be easily adapted, however, to more complex cases in which there may be multiple Review Sets or multiple Positive or Negative Sets within a single Review Set (e.g., when review and production is done on an interim or rolling basis). The procedures could also be adapted to a circumstance in which a Party (or the Court) wanted to arrive at an estimate of "end-to-end" recall (a recall number that accounted for the effects of both the exclusionary steps and the review process).[55] All that is required is to assume a stratified sampling design and obtain the inputs and estimates accordingly. The procedures for doing so are provided in the Guidelines accompanying this protocol and commentary.

All review processes. I.e., if distinct review processes have been applied to distinct parts of the Review Set (e.g., if a form of technology-assisted review has been used for the majority of the Review Set, but manual review has been used that part of the Review Set that consists of images with no usable text), the Responding Party should aggregate the results of those review processes, creating two subsets: documents deemed responsive (regardless of which specific process was the source of that designation) and documents not deemed responsive. After re-integrating document families, these two subsets become the Positive Set and the Negative Set used in the validation exercise.

True Positives. For further discussion of this term (and the related terms *False Positive*, *True Negative*, *False Negative*), see the glossary provided in Chapter 3 of the Guidelines.

Simple random sample. See earlier comment on *Simple random sample*.

---

[52] To be specific: 95% of the time.

[53] Of course, as provided for in the Protocol, quantitative measures need to be supplemented by qualitative analysis to obtain a complete view of the impact of a given exclusion.

[54] For other definitions of recall, see Grossman & Cormack 2013: "The fraction of Relevant Documents that are identified as Relevant by a search or review effort" (p. 27); see also Sedona 2020: "When describing search results, recall is the number of documents retrieved from a search divided by all of the responsive documents in a collection. For example, in a search for documents relevant to a document request, it is the percentage of documents returned from a search compared against all documents that should have been returned and exist in the data set" (p. 360*f.*).

[55] As required, for example, in *In re Diisocyanates Antitrust Litig*., MDL No. 2862, 2021 WL 4295729 (W.D. Pa. Aug. 23, 2021).

400 documents. A sample of this size will allow the Parties to obtain an estimate of the prevalence of responsive documents in the Positive Set with a margin of error (calculated at a 95% confidence level) that does not exceed +/- 5%. This level of constraint on sampling error will suffice in most cases to ensure that this component of the recall estimate is not a practically significant contributor to the margin of error associated with that estimate (and so does not prevent the Parties from obtaining a meaningful gauge of the recall achieved by the Review Process.[56]

In connection with the Positive Sample, it may also be observed that an estimate of the percentage of responsive documents in the Positive Set is in fact an estimate of the *precision* achieved by the review. Precision is a measure of how accurate or "on-target" a review effort has been. It answers the question: *Out of all the documents that the review process classified as responsive, what percentage were actually responsive?* The higher the percentage (i.e., the higher the level of precision), the more successful the review has been at avoiding false positives (non-responsive documents incorrectly classified as responsive); the lower the percentage (i.e., the lower the level of precision), the less successful the review has been at avoiding false positives.[57]

Practitioners, in discussing validation protocols, sometimes ask why it is necessary to obtain a precision estimate when the primary focus, in the context of legal discovery, is on the completeness of the review (i.e., on recall). The answer is that precision is estimated not as an end in itself (although the information it provides may be found useful by both parties), but as a necessary step in obtaining an estimate of recall: we need an estimate of the number of documents captured by the Review Process because that estimate will figure in both the numerator and denominator of our eventual recall estimate.

False Negatives. For further discussion of this term (and the related terms *True Positive*, *False Positive*, *True Negative*), see the glossary provided in Chapter 3 of the Guidelines.

Simple random sample. See earlier comment on *Simple random sample*.

3,400 documents. It is not possible to specify a single size for the Negative Sample that will be suitable in all circumstances; there are just too many variables, both statistical and non-statistical, that affect the sample size that would be appropriate to the conditions in a given matter. That said, if we adopt a sound and practical approach to the analysis of the power of different sample sizes, and if that approach makes allowance for the differences in power caused by differences in prevalence, we can arrive at a reasonable set of prevalence-conditioned default sizes for the Negative Sample as well as a reasonable default size for use in cases in which we lack actionable information about prevalence. We discuss the methodology and results of such an approach in some detail in Chapter 2 of the Guidelines accompanying this document. In this comment, we summarize the results of the analysis provided there; readers interested in a closer look at the reasoning and data that leads to these results will find the more extensive discussion in the Guidelines helpful.

---

[56] It should again be noted that a sample of this size will have the specified characteristics *regardless of the size of the source population* (i.e., in this case, regardless of the size of the Positive Set). For more on the relation of population size to the power of a given sample size, see Chapter 2 of the Guidelines accompanying this protocol.

[57] For other definitions of precision, see Grossman & Cormack 2013: "The fraction of Documents identified as Relevant by a search or review effort, that are in fact Relevant" (p. 25); see also Sedona 2020: "When describing search results, precision is the number of documents retrieved from a search divided by the total number of documents returned. For example, in a search for documents relevant to a document request, it is the percentage of documents returned from a search that are actually relevant to the request" (p. 354).

The approach taken to arriving at default sample sizes is a simple one. In order to control for the factor of prevalence, we distinguish seven prevalence bands: 0% to 1%, 1% to 2%, 2% to 3%, 3% to 5%, 5% to 7%, 7% to 10%, and greater than or equal to 10%. Within each band, we set a criterion that the distribution of the margins of error (for recall) that could result from using a Negative Sample of a given size[58] must meet. We then find the smallest size that meets the criterion for each prevalence band. The results are shown in Table 1.

Table 1: Prevalence Bands, Sample Size Selection Criteria, and Candidate Sample Sizes

Prevalence Band	Criterion	Sample Size
10% ≤ Prevalence	**100%** of margins of error within ± **5%**	**2,230** documents
7% ≤ Prevalence < **10%**	**95%** of margins of error within ± **5%**	**3,230** documents
5% ≤ Prevalence < **7%**	**95%** of margins of error within ± **6%**	**3,400** documents
3% ≤ Prevalence < **5%**	**80%** of margins of error within ± **6%**	**5,080** documents
2% ≤ Prevalence < **3%**	**80%** of margins of error within ± **7%**	**7,260** documents
1% ≤ Prevalence < **2%**	**70%** of margins of error within ± **8%**	**9,570** documents
Prevalence < **1%**	**50%** of margins of error within ± **10%**	**12,050** documents

This analysis provides us with a useful set of candidate sample sizes. To confirm the advisability of the sample sizes, we take a closer look at the distributions of (recall) margins of error they will occasion. Table 2, by providing a "five-number summary"[59] of the distributions, enables such a closer look.

Table 2: Five-Number Summaries for Margins of Error Yielded by Candidate Sample Sizes

Prevalence Band	Sample Size	Distribution of Margins of Error				
		Min	Q1	Med	Q3	Max
10% ≤ Prevalence	2,230	± 0.8%	± 3.9%	**± 4.2%**	**± 4.4%**	± 5.0%
7% ≤ Prevalence < **10%**	3,230	± 0.6%	± 3.2%	**± 4.1%**	**± 4.5%**	± 5.4%
5% ≤ Prevalence < **7%**	3,400	± 0.7%	± 3.8%	**± 4.9%**	**± 5.4%**	± 6.4%
3% ≤ Prevalence < **5%**	5,080	± 0.7%	± 3.9%	**± 5.1%**	**± 5.8%**	± 7.5%
2% ≤ Prevalence < **3%**	7,260	± 0.8%	± 4.3%	**± 5.8%**	**± 6.8%**	± 8.5%
1% ≤ Prevalence < **2%**	9,570	± 0.9%	± 4.9%	**± 6.9%**	**± 8.2%**	± 11.8%

[58] For purposes of obtaining a distribution of margins of error for a given size of Negative Sample, we consider a scenario in which our review has resulted in Positive Set of 200,000 documents and a Negative Set of 1,800,000 documents. From the Positive Set we have drawn a Positive Sample of the Protocol-specified 400 documents; from the Negative Set we have drawn a Negative Sample of the size under evaluation. To obtain the distribution of margins of error, we then calculate the (recall) margins of error that would result from all possible combinations of outcomes (number of responsive documents observed) for both the Positive and Negative Samples, provided that the outcome generated a point estimate for recall of at least 60% (the latter constraint is introduced to narrow the focus to cases that would matter in real-world circumstances). More on these parameters can be found in the Guidelines.

[59] I.e., for each distribution, the minimum, first quartile, median, third quartile, and maximum values. For more on these terms, see the appropriate entry in the glossary provided in Chapter 3 of the Guidelines.

Prevalence < **1%**	12,050	± 1.5%	± 7.0%	**± 10.0%**	**± 12.4%**	± 56.1%

The closer look confirms the advisability of the candidate sample sizes and puts us in a position to summarize recommendations for default sizes for the Negative Sample. We do so, first, under the assumption that we have some actionable information as to the overall prevalence of responsive material in the review set, then, second, under the assumption that we do not have actionable information as to the overall prevalence of responsive material in the review set.

If a responding party has, in the course of the review, gathered enough empirical evidence to obtain a rough gauge of the prevalence of responsive material in the review set, it should select a Negative Sample in accordance with the following table.

Table 3: Recommended Sizes for the Negative Sample Used in the Estimation of Recall

Prevalence Band	Recommended Sample Size
10% ≤ Prevalence	**2,230** documents
7% ≤ Prevalence < **10%**	**3,230** documents
5% ≤ Prevalence < **7%**	**3,400** documents
3% ≤ Prevalence < **5%**	**5,080** documents
2% ≤ Prevalence < **3%**	**7,260** documents
1% ≤ Prevalence < **2%**	**9,570** documents
Prevalence < **1%**	**12,050** documents

If a responding party has not gathered enough empirical evidence to obtain a rough gauge of the prevalence of responsive material in the review set, it should select a Negative Sample of **3,400** documents. This is the size of sample recommended for cases in which we have reason to believe that the prevalence is between 5% and 7%, so it represents a reasonable middle ground in the absence of any directional information about prevalence. As a practical matter, moreover, drawing and reviewing a Negative Sample of 3,400 documents is not an unduly burdensome task to ask the Responding Party to perform in the interest of fostering a well-grounded trust in the results of a review. Further, as seen in Table 4, when viewed in terms of its power at constraining sampling error, a Negative Sample of 3,400 documents is viable as a basis for meaningfully precise estimates of recall, both when considered with no constraint on prevalence and when considered with prevalence being limited to the 5% to 7% range (a range of prevalence values not uncommonly seen in real-world practice).

Table 4: Five-Number Summary: Negative Sample of 3,400 documents, No Control on Prevalence

Prevalence Band	Sample Size	Min	Q1	Med	Q3	Max
0% ≤ Prevalence < **100%**	3,400	± 0.5%	± 3.3%	**± 3.7%**	**± 4.6%**	± 54.3%
5% ≤ Prevalence < **7%**	3,400	± 0.7%	± 3.8%	**± 4.9%**	**± 5.4%**	± 6.4%

Finally, it should be emphasized here, as it is elsewhere in the Protocol (and associated Commentary), that the sample sizes specified in the Protocol are provided as default sizes that may be adjusted as circumstances warrant. As already noted, it is impossible to specify, in the abstract, sample sizes that will be suitable in every particular circumstance. It is expected that practitioners will take into account the factors that are operative in the case of their specific matter and review process, apply the principle of proportionality[60] as well as the reasoning described in this chapter, and then, when those considerations so warrant, adjust the default sample sizes to suit their circumstances. If, for example, a party has reason to believe that the lower limit for prevalence in the Review Set is higher than 10% (i.e., higher than the lower limit for the top prevalence band defined in Table 3), it may be able to use a smaller sample even than that specified for the top band in the table.[61] Also, in cases involving small collections of documents (i.e., small Review Sets), it could be reasonable to adjust the sample sizes recommended in the Protocol downward.[62] In some instances (e.g., cases with very small Review Sets), parties might make that adjustment simply based on what is reasonable and proportionate; in other instances, parties might wish to put the adjustment on sounder statistical footing and so seek expert support in making that adjustment.[63]

Combine. For the reason this step is taken, see earlier comment on *Combine*.

Manual review. See earlier comment on *Manual review*.

Estimate of the recall. See earlier comment on *Estimated number*.

A reasonable *prima facie* threshold. There is no consensus around a specific minimum value for recall that a review effort must meet to qualify as "reasonably complete." That is as it should be: there are simply too many circumstance-specific variables that affect both what should be considered "reasonable" and what should be considered "complete" to arrive at a single number that will be appropriate in all circumstances. That said, as a practical matter, it is useful for practitioners to have the guidance provided by a *prima facie* threshold that can serve as an actionable target until specific circumstances suggest otherwise. A threshold value of 75% recall serves this purpose well, both on grounds of reasonableness and on grounds of completeness.

With regard to *reasonableness*, the achievement of 75% recall is a challenge, but one that can be met with the competent operation of advanced review technologies. In the studies conducted in the TREC Legal

---

[60] I.e., "considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit" (Fed R. Civ. P. 26(b)(1)).

[61] For example, if a responding party has reasonable grounds for believing that the prevalence in the Review Set is at least 15%, it could use a Negative Sample of 1,290 documents and still meet the criterion of 100% of possible margins of error being within ± 5%. In making these sorts of adjustments, a party would likely benefit from consultation with an individual with the appropriate statistical expertise.

[62] Note that the margin-of-error calculations discussed in the accompanying Guidelines do contain a term that accounts for the size of the population relative to the size of sample, so, for any given sample size, the size of the margin of error will decrease as the size of the sample approaches the size of the population (assuming prevalence levels (and other relevant inputs) are held constant). When Review Sets are large (and so even samples on the order of 20,000 documents would be small relative to the population), this term has no meaningful impact; when, however, Review Sets are small, the term could have an impact on the margin of error (thus allowing smaller sample sizes).

[63] A competent expert should be able to provide the required guidance at reasonable cost. Of course, one motivation for creating the Protocol (and supporting materials) is to promote **a wider distribution of the required skills (or at least a wider distribution of an understanding of what skills are required)**. To the extent that it is successful at meeting that objective, it should increase the supply of individuals with the requisite expertise and therefore bring down the cost of obtaining it.

Track,[64] for example, of the 53 submissions to the Interactive Task from 2008 to 2010, seven were found to have achieved recall of 75% or greater; and, of those seven, five achieved that threshold while also meeting or exceeding 75% precision. Of the 70 submissions to the Track's Learning Task in 2011, none met the 75% recall threshold while also maintaining at least 75% precision (or even while also maintaining at least 50% precision), but several[65] were able to meet the 75% recall threshold at lower levels of precision. Exercises conducted in the TREC Total Recall Track in 2015 and 2016 provide further evidence that 75% recall is an achievable threshold.[66] From the 2015 edition of the track, for example, we find that, if we consider each participant's best run for each of the five test collections featured in the exercise, over half of the runs (21 out of 34) achieved, averaging across the topics featured in each collection, recall of at least 75% while maintaining precision of at least 50%.[67] That a review exercise should meet a minimum threshold of 75% recall, even if meeting that threshold requires the tolerance of a low level of precision, is thus a reasonable expectation.

With regard to *completeness*, it is helpful to keep in mind the redundancy that characterizes the distribution of information across the documents in a collection. In the collections of documents typically subject to legal discovery, the information salient to discovery requests is not distributed such that each document contains its own unique bit of information; rather the information is distributed in a redundant (one-to-many) fashion, such that the same information is often contained in many documents.

What this means is that, on the one hand, it is true that there is a general correlation between retrieving documents and retrieving information: generally speaking, the greater the number of responsive documents we have retrieved, the greater the amount of salient information we will have retrieved. Hence the value of the (document-based) recall metric as a measure of the effectiveness of a review. On the other hand, the correlation is not perfect. Given the redundancy of information, the return on investment (new information gained from documents newly retrieved) may decline as the number of documents already retrieved increases. In most cases, therefore, once a review has achieved a reasonably high level of recall (e.g., 75%), the unretrieved (missed) documents should add little new or important information to that which can already

[64] For a complete archive of resources related to the TREC Legal Track, see: https://trec-legal.umiacs.umd.edu/. For summaries of the results of each edition of the Legal Track, see the track overviews for each year (available at the resource page just noted and at: https://trec.nist.gov/proceedings/proceedings.html).
[65] To be specific, using the cutoffs reported in the Track Overview for 2011, 19 submissions achieved recall of 75% or greater while also maintaining precision of at least 2%. In the Overview, results were reported at different "cutoff" points down the ranked lists of documents submitted by task participants. The achievement of any target level of recall (even 100%) is possible as long as one chooses a cutoff deep enough down the list (although that will often coincide with a very low level of precision).
[66] Grossman et al. 2016; Roegiest et al. 2015.
[67] Roegiest et al. 2015.

be gathered from those that were successfully retrieved.[68] A threshold of 75% recall is thus suitably high to serve as reasonable *prima facie*[69] threshold for completeness in most cases.[70]

Finally, with regard to assessing whether a given quantitative threshold is met, the *point estimate* for recall (as opposed, for example to the lower limit of the range defined by the margin of error), as long as the margin of error associated with that estimate is reasonably small, will suit the purpose. As a practical matter, when assessing completeness, once we are in an acceptable range with our quantitative measures (as indicated by a point estimate of 75% to 80% recall), we learn more from a qualitative analysis of false negatives (missed responsive documents) than we do from further pinpointing our quantitative measures a few percentage points higher or lower.

Supplemented. The course of action required to address any deficiencies identified via the validation exercise will of course be dependent on the nature of the deficiencies. In some cases (e.g., when the deficiencies are localized in a readily identifiable subset of the Review Set, the remedy may be simply to identify the missed documents by the use of search terms or channel the documents in the subset in question to a manual review process; in other cases, the remedy may be to update the statistical model being used to classify the documents in the Review Set through the provision of additional training data that accounts for the sorts of deficiencies identified; and other solutions might be appropriate depending on circumstance. It will be the Responding Party's decision as to how to address such deficiencies. Consultation with search experts may be useful in this regard.

The possibility that remediation steps may be occasioned by the validation exercise raises the question of whether, post-remediation, another validation exercise must be conducted. The answer is: *Yes, but not necessarily a full-scale re-run of the validation exercise specified in Appendix B of the Protocol*. More specifically:

- If the observed deficiencies are easily **localized** within specific subsets of the Review Set (i.e., occur predominantly, or only, in documents that exhibit specific identifiable characteristics) and if the deficiencies are **homogeneous** in character (i.e., are of a specific type of error or of a small number of specific types of error), then all that is required is a narrow test designed specifically to

[68] For additional discussion of the distinction between document recall and information recall, see Hedin et al. 2016: 412.

[69] The threshold is qualified as being only *prima facie* because, as provided for in the Protocol, the quantitative measure must be supplemented by qualitative analysis. It is not impossible that, when a review has achieved 75% recall, there is still a significant amount of important information yet to be retrieved. Perhaps, for example, the information contained in the 75% of responsive documents that were successfully retrieved provides good coverage of a subset of responsive topics but leaves other topics inadequately addressed (For further discussion of the implications of this scenario, see Grossman & Cormack 2021: 26). This is where qualitative assessment comes in; it enables the Parties to evaluate the *nature* of any responsive documents that have been missed by the Review Process and, more specifically, the nature of the information contained in such documents and whether that information is both genuinely *important* to the issues being litigated and *non-redundant* with information that can be gathered from the set of documents the review has successfully identified as responsive

[70] This statement assumes, of course, that the recall estimate has been obtained by sound methods (an appropriate sampling design, blind review of samples, proper execution of estimation procedures), such as those described in this protocol and the accompanying guidelines.

assess whether the chosen remedy has had the intended effect. **In this circumstance, there is no need to re-run a full-scale validation exercise**.[71]

- If, on the other hand, the observed deficiencies are **pervasive** throughout the Review Set (i.e., are not easily localized within specific subsets) and if the deficiencies are **heterogeneous** in character (i.e., represent many distinct types of error), then it is likely that remediation will require a more thorough and wide-ranging analysis and recalibration of the review effort. **In this circumstance, it will be necessary,[72] in order to arrive at a well-grounded trust in the results, to re-run the validation exercise specified in Appendix B of the Protocol**.[73]

Following information. See comment made in reference to Following information in Appendix A.

Disclosure of privileged material. The text of the first two provisions in this appendix is based on that drafted by Hon. Andrew J. Peck (ret.) as a model order for parties choosing to invoke Federal Rule of Evidence 502 ("Rule 502")[74] and included as an appendix to The Sedona Conference Commentary on Protection of Privileged ESI.[75] With the third provision, the Protocol adds the further protection that whatever is produced by the responding party is deemed to have been inadvertently produced and the responding party is deemed to have taken reasonable steps to prevent that disclosure. This provides the responding party with all the protection that Rule 502 provides. For a discussion of whether the training documents used in a technology-assisted review are entitled to work-product protections, see Facciola & Favro 2015; in that paper, the authors argue that training documents (and other potential interim productions) are subject to work-product protections and so their production to a requesting party could constitute a waiver of that protection (hence the need for the protection afforded by Rule 502).

[71] This will especially be the case when the Responding Party's chosen remedy is such that it could only increase recall (i.e., could only add to the set of documents designated by the Review Process as responsive, not subtract from that set).
[72] Subject, as always, to circumstance-specific proportionality considerations.
[73] This will especially be the case when the Responding Party's chosen remedy is such that it could have the effect of either increasing or decreasing recall (i.e., could both add to and subtract from the set of documents designated by the Review Process as responsive).
[74] Fed. R. Evid. 502.
[75] Sedona 2016: 198.

References

Orders and Opinions

In Re: Actos (Pioglitazone) Products Liability Litigation, MDL No. 6:11-md-2299 (W.D. La., Jul. 2, 2012).

In Re: Blue Cross Blue Shield Antitrust Litigation, MDL No.: 2406, Master File No.: 2:13-CV-20000-RDP (N.D. Ala., Southern Division, Apr. 16, 2015).

In Re Broiler Chicken Antitrust Litigation, Case No. 1:16-cv-08637 (N.D. Ill., Eastern Division, Jan. 3, 2018).

In re Diisocyanates Antitrust Litig., MDL No. 2862, 2021 WL 4295729 (W.D. Pa. Aug. 23, 2021).

D. Md. (2007). *Suggested Protocol for Discovery of Electronically Stored Information*. Available: https://www.mdd.uscourts.gov/sites/mdd/files/ESIProtocol.pdf.

Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182 (S.D.N.Y. 2012), adopted sub nom. Moore v. Publicis Groupe SA, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

Greater New York Taxi Association v. City of New York, No. 13 Civ. 3089 (VSB) (JCF) (S.D.N.Y. Sept. 11, 2017).

Livingston v. City of Chicago, No. 16 CV 10156 (N.D. Ill. Sep. 3, 2020

Nichols v. Noom, Inc., No. 20-CV-3677 (LGS) (KHP) (S.D.N.Y. Mar. 11, 2021).

Rio Tinto Plc v. Vale S.A., 14-Civ. 3042 (RMB)(AJP) (S.D.N.Y. Mar. 2, 2015) (Also: Stipulation and Order re: Revised Validation and Audit Protocols for the Use of Predictive Coding in Discovery, Sep. 8, 2015).

United States v. O'Keefe, 537 F. Supp. 2d 14 (D.D.C. 2008).

In Re: Valsartan, Losartan, and Irbesartan Products Liability Litigation (D.N.J. Dec. 2, 2020).

Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260, 262 (D. Md. 2008).

W.D. Wash. (2020). *Model Agreement re: Discovery of ESI*. Available: https://www.wawd.uscourts.gov/sites/wawd/files/ModelESIAgreement.pdf.

Publications

Ball, C. (2023). The Annotated ESI Protocol. Retrieved from https://craigball.net/2023/01/09/the-annotated-esi-protocol/ (accessed: May 26, 2023).

Baron, J. R., Losey, R. C., & Berman, M. D. (Eds.). (2016). *Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner*. American Bar Association, Section of Litigation.

Bolch Judicial Institute (2019). *Technology Assisted Review (TAR) Guidelines*. Bolch Judicial Institute. Available: https://scholarship.law.duke.edu/bolch/7/.

Coeckelbergh, M. (2012). Can we trust robots?. *Ethics and information technology*, 14(1), 53-60.

Desai, D. R., & Kroll, J. A. (2017). Trust but verify: A guide to algorithms and the law. *Harv. JL & Tech.*, 31, 1.

DOJ (2018). *Predictive Coding Model Agreement*. Available: https://www.justice.gov/atr/page/file/1096101/download.

Economou, N. (2019). Artificial Intelligence and the Law: the ABA's important and timely contribution. *Legal Analytics Committee Newsletter*. ABA (August, 2019).

Facciola, J. M., & Favro, P. (2015). Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled to Work Product Protection. *Federal Courts Law Review*, *8*.

Ferrario, A., Loi, M., & Viganò, E. (2019). In AI we trust incrementally: a multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology*, 1-17.

Freedman, D., Pisani, R., & Purves, R. (2011). *Statistics*, Fourth Edition. W.W. Norton & Company, Inc. New York.

Grossman, M. R. & and Cormack, G. V. (2011), "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," *Richmond Journal of Law and Technology*, vol. 17, no. 3.

Grossman, M. R. & Cormack, G. V. (2013). The Grossman-Cormack Glossary of Technology-Assisted Review, 7 *Fed. Cts. L. Rev.* 1.

Grossman, M. R. & Cormack, G. V. (2021). Vetting and Validation of AI-Enabled Tools for Electronic Discovery, in *Litigating Artificial Intelligence*, J. Presser, J. Beatson, & G. Chan, Ed., Edmond Publishing.

Grossman, M. R., Cormack, G. V., & Roegiest, A. (2016). TREC 2016 Total Recall Track Overview. In *TREC*.

Hedin, B., Tomlinson, S., Baron, J.R., & Oard, D.W. (2009). Overview of the TREC 2009 Legal Track, in *NIST Special Publication: SP 500-278, THE Eighteenth Text Retrieval Conference Proceedings (TREC 2009)*.

Hedin, B., Brassil, D., and Jones, A. (2016). On the Place of Measurement in E-Discovery, in *Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner*, J. R. Baron, R. C. Losey, and M. D. Berman, Ed. Chicago: American Bar Association.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE.

International Organization for Standardization (2017). *ISO/IEC 27050-3: Information technology — Security techniques — Electronic discovery — Part 3: Code of practice for electronic discovery* (Geneva, 2017).

Lat, D. (2020). The Ethical Implications of Artificial Intelligence, *Above the Law: Law2020*, https://abovethelaw.com/law2020/the-ethical-implications-of-artificial-intelligence/.

Lau, T.T. and Lee, E.G., III (2017). *Technology-Assisted Review for Discovery Requests: A Pocket Guide for Judges*. Federal Judicial Center. Available: https://www.fjc.gov/sites/default/files/2017/Technology-Assisted%20Review%20for%20Discovery%20Requests.pdf.

Losey, R. (2018). Judge goes where angels fear to tread: tells the parties what keyword searches to use [blog post]. Retrieved from https://e-discoveryteam.com/2018/06/24/judge-goes-where-angels-fear-to-tread-tells-the-parties-what-keyword-searches-to-use/ (accessed: May 25, 2023).

Mendenhall, W. and Beaver, R. J. (1991). *Introduction to Probability and Statistics*, Eighth Edition. Boston, MA: PWS-Kent Pub. Co.

O'Halloran, T., McManus, B., Harbison, A., Grossman, M. R., & Cormack, G. V. (2023, August). Technology-Assisted Review for Spreadsheets and Noisy Text. In *Proceedings of the ACM Symposium on Document Engineering 2023* (pp. 1-4).

Payne, C., & Six, M. (2020). A Proposed Technology-Assisted Review Framework. *Law360*, April 27, 2020.

Pizzi, P. J. & Brickell, J. L. (2014). The Ethics of Technology in E-Discovery - An Introduction. *IADC Committee Newsletter* (December 2014).

Roegiest, A., Cormack, G. V., Clarke, C. L., & Grossman, M. R. (2015). TREC 2015 Total Recall Track Overview. In *TREC*.

The Sedona Conference (2014). The Sedona Conference Commentary on Achieving Quality in the E-Discovery Process, 15 *Sedona Conf. J*. (2014).

The Sedona Conference (2016). Commentary on Protection of Privileged ESI, 17 *Sedona Conf. J.* 95 (2016).

The Sedona Conference (2017). The Sedona Conference Commentary on Proportionality in Electronic Discovery, 18 *Sedona Conf. J.* 141 (2017).

The Sedona Conference (2018). The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production. 19 *Sedona Conf. J.* 1 (2018).

The Sedona Conference (2020). The Sedona Conference Glossary: eDiscovery & Digital Information Management, Fifth Edition, 21 *Sedona Conf. J*. 263 (2020).

The Sedona Conference (2023), TAR Case Law Primer, Second Edition, 24 *Sedona Conf. J.* 1 (2023), https://thesedonaconference.org/publication/TAR_Case_Law_Primer.

Spiekermann, S. (2015). *Ethical IT innovation: A value-based system design approach*. CRC Press.

Taddeo, M. (2017). Trusting digital technologies correctly. *Minds and Machines*, 27(4), 565-568.

Taddeo, M. (2010). Trust in technology: A distinctive and a problematic relation. *Knowledge, Technology & Policy*, 23(3-4), 283-286.

Taddeo, M., & Floridi, L. (2011). The case for e-trust. *Ethics and Information Technology*, 13(1), 1-3.

Thompson, S. K. (2002). *Sampling*, Second Edition. John Wiley & Sons, Inc. New York.

Walker, J. Q. (2018). What's Artificial About Intelligence? The Ethical and Practical Considerations When Lawyers Use AI Technology. *Bloomberg Law* (2018).

Acknowledgements

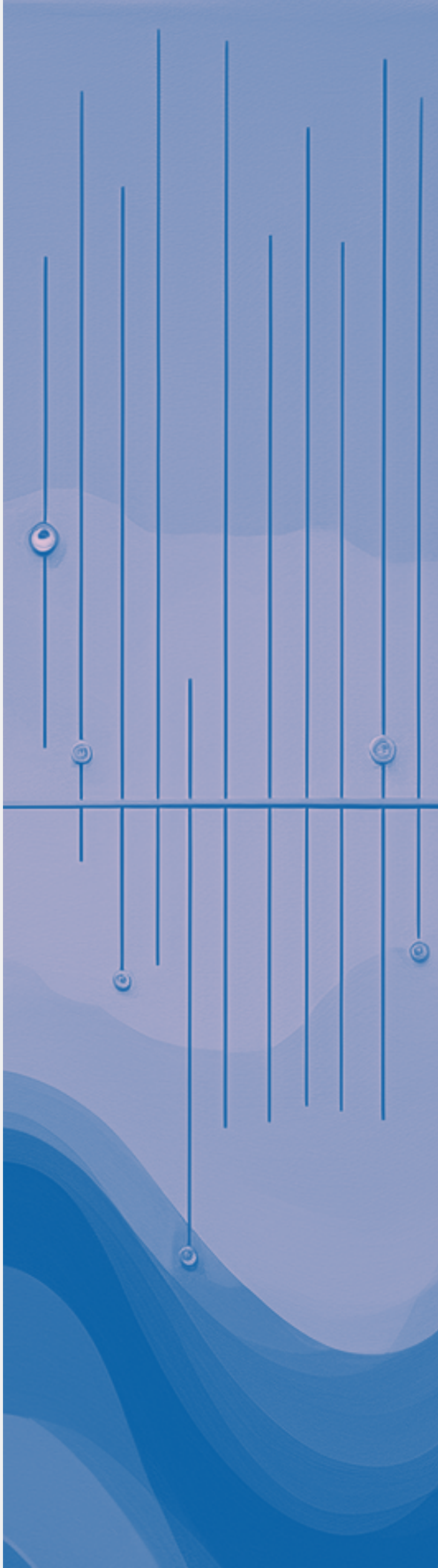
This document has benefited greatly from the many insightful comments provided by David Siffert and his colleagues at NYU's Center on Civil Justice. It has also been greatly improved by suggestions from Samuel Curtis, Nicolas Economou, Niki Iliadis, and Nicolas Miailhe of The Future Society. As observed in the *Notes on Preparation of these Documents*, a Review Group has been very generous with their time in reviewing and commenting on multiple drafts of these documents. The members of the group are as follows.

- Hon. Benes Z. Aldana (Ret.)
  President & CEO
  The National Judicial College
- Jason R. Baron
  Professor of the Practice
  University of Maryland - College Park
- Julia L. Brickell
  Lecturer, Columbia University
  Board Member, The AI Education Project
  Attorney at Law
- Judge Herbert B. Dixon, Jr. (Ret.)
  Superior Court of DC
- Hon. John M. Facciola
  U.S. Magistrate Judge (Ret.)
- Professor Keith R. Fisher
  Distinguished Fellow
  National Judicial College
- Hon. Paul W. Grimm (Ret.)
  David F. Levi Professor of the Practice of Law
  Director of the Bolch Judicial Institute
  Duke Law School
- Maura R. Grossman, J.D., Ph.D.
  Research Professor, University of Waterloo
  Principal, Maura Grossman Law
- Timothy T. Lau, Sc.D., J.D.
  Senior Research Associate
  Federal Judicial Center
  Research Division
- Christine P. Payne
  Shareholder
  Gunster
- Hon. Andrew J. Peck
  United States Magistrate Judge, SDNY (Ret.)
  Senior Counsel, DLA Piper LLP
- Michael Quartararo
  President of ACEDS & Professional Development
  The Association of Certified E-Discovery Specialists

- MK Sapp
  Associate
  Gunster

The perspective and insights provided by this group has been immensely helpful on both general questions and specific points; we are much indebted to the members of the group for their contributions. Contributors participated in their personal capacity and their participation does not necessarily imply an endorsement, by themselves or the organizations with which they are associated, of all particulars of the final product. Finally, we express sincere gratitude for IEEE's support for the project and, in particular, the interest shown by Lloyd Green, Konstantinos Karachalios, and Alpesh Shah.

Of course, responsibility for any errors belongs to the author.



THE FUTURE SOCIETY

867 Boylston Street
Boston, MA 02166