
THE FEDERAL COURTS LAW REVIEW

Volume 8, Issue 3

2015

SAFEGUARDING THE SEED SET: WHY SEED SET DOCUMENTS MAY BE ENTITLED TO WORK PRODUCT PROTECTION

The Honorable John M. Facciola and Philip J. Favro

TABLE OF CONTENTS

I.	ABSTRACT	2
II.	INTRODUCTION	3
III.	THE ATTORNEY WORK PRODUCT DOCTRINE	9
	A. The Purposes and Policy of the Work Product Doctrine	9
	B. The Scope of the Work Product Doctrine.....	11
IV.	PREDICTIVE CODING AND SEED SETS	12
	A. The Role of Seed Sets in Predictive Coding	13
	1. Judgmental Sampling	14
	2. Random Sampling	15
	3. The Work Product Doctrine Would Likely Apply Only To Judgmental Samples	16
	B. Jurisprudence Regarding the Protection of Seed Set Documents	17
	1. Cases Favoring Disclosure of Seed Sets	17
	2. Cases Favoring Protection of Seed Sets.....	20
V.	PROTECTING SEED SET DOCUMENTS AS WORK PRODUCT UNDER ANALOGOUS SCENARIOS INVOLVING SELECTIONS OF DOCUMENTS.....	22
	A. Selections of Documents: When Are They Protected as Work Product?	22
	1. Document Selections Protected as Opinion Work Product under <i>Sporck</i>	23
	2. Refining the Scope of the <i>Sporck</i> Rule	24
	3. Application of <i>Sporck</i> in Factual Scenarios Involving Selections of Documents.....	26

2	<i>Safeguarding the Seed Set</i>	[Vol. 8
	a. Preparing Witnesses for Deposition	27
	b. Document Inspections	27
	c. Responses to Contention Interrogatories	29
	B. Application of the <i>Sporck</i> Rule to Predictive Coding Seed Sets.....	30
VI.	CONCLUSION	32

*Hon. John M. Facciola and Philip J. Favro*¹

I. ABSTRACT

Some courts and commentators have been urging that lawyers and litigants only use predictive coding when they have entered into a fully disclosed and cooperative use protocol. This notion – that parties must cooperate in a transparent fashion in order to use predictive coding – is particularly apparent in the development of the training or seed set of documents. Those advocating this position argue that cooperative and transparent seed set creation will more readily ensure that the predictive coding process proceeds in an orderly fashion. The byproduct of doing so, proponents argue, will be document productions that satisfy standards of reasonableness and proportionality.

While it is impossible to argue against cooperation and transparency in discovery, we assert that the allure of this position fails to recognize that a seed set generated through counsel’s exercise of skill, judgment, and reasoning may reflect its perceptions of relevance, litigation tactics, or even its trial strategy. In this article, we argue that these conclusions regarding key strategic issues – memorialized in counsel’s selection of documents – are entitled to protection from disclosure under the attorney work product doctrine since they may reveal counsel’s mental processes and legal theories.

Our arguments on the issues are grounded in the holding and policies underlying the Supreme Court’s venerable opinion of *Hickman v. Taylor*. They also find support in well-established work product jurisprudence that generally protects a lawyer’s selection of documents from discovery in analogous circumstances. While there are limitations to this rule and though

1. Judge Facciola, United States Magistrate Judge for the United States District Court for the District of Columbia (Ret.), AB. College of the Holy Cross, J.D. Georgetown Law. Philip Favro is Senior Discovery Counsel, Recommind, Inc.; J.D., Santa Clara University School of Law, 1999; B.A., Political Science, Brigham Young University, 1994. The authors wish to recognize the insightful work of Adam Kuhn, J.D./M.B.A., University of San Francisco, 2013, in connection with the preparation of this article.

there are obvious potential benefits to cooperation and transparency, we contend that the convenience created by these notions should not be used to coerce a lawyer into revealing its work product as reflected in the development of a predictive coding seed set.

II. INTRODUCTION

The civil discovery process is difficult and complex.² Between the challenges of locating key evidence to the problem of dealing with truculent adversaries, lawyers have frequently struggled with this all-important yet seldom enjoyable aspect of litigation.³ The quest to address these difficulties and simplify discovery has been ongoing now for over 75 years.⁴ Since the advent of the Federal Rules of Civil Procedure (“Rules” or “Rule”) in 1938 and their declared intent to resolve matters in a “just, speedy, and inexpensive” manner,⁵ federal rule makers have labored to help counsel, clients, and courts simplify the discovery process.⁶

Nevertheless, with each Rule change that was designed to tackle a problem, other issues – sometimes more complex than those just addressed – began to surface.⁷ That was certainly the case after the amendment cycles that ended in 1993⁸ and 2006.⁹ In each instance, the sheer growth of electronically stored information (“ESI”), among many other things, threatened to overwhelm the legal system with potentially discoverable

2. See Philip J. Favro & Derek P. Pullan, *New Utah Rule 26: A Blueprint for Proportionality Under the Federal Rules of Civil Procedure*, 2012 MICH. ST. L. REV. 933, 936-38 (2012) (commenting on the increasing costs and related challenges plaguing the discovery process).

3. John S. Beckerman, *Confronting Civil Discovery’s Fatal Flaws*, 84 MINN. L. REV. 505, 517 (2000) (observing that it was “naive to expect ‘that adversarial tigers would behave like accommodating pussycats throughout the discovery period, saving their combative energies for trial.’”).

4. See generally Martin H. Redish, *Electronic Discovery and the Discovery Matrix*, 51 DUKE L.J. 561, 600-03 (2001) (discussing generally the achievements and failures of the Federal Rules of Civil Procedure since their implementation in 1938).

5. FED. R. CIV. P. 1.

6. W. Bradley Wendel, *Regulation of Lawyers Without the Code, the Rules, or the Restatement: Or, What Do Honor and Shame Have to do with Civil Discovery Practice?*, 71 FORDHAM L. REV. 1567, 1573-75 (2003) (explaining generally the “great deal of attention” that rules makers have dedicated to ameliorating challenges associated with discovery practice).

7. See Beckerman, *supra* note 3, at 518-20, 530-40.

8. See JUDICIAL CONFERENCE OF THE UNITED STATES, REPORT OF ADVISORY COMMITTEE ON CIVIL RULES 83-84 (May 2, 2014) (REPORT) (discussing the impact of the “information explosion” on the 1993 changes to Rule 26).

9. See generally Philip J. Favro, *A New Frontier in Electronic Discovery: Preserving and Obtaining Metadata*, 13 B.U. J. SCI. & TECH. L. 1 (2007) (describing the impact and challenges of data growth on discovery practice and the role of the 2006 amendments in addressing the issues).

materials.¹⁰

The latest round of proposed Rule amendments is similarly designed to address issues associated with data growth.¹¹ Slated for implementation in late 2015, the Rule changes emphasize proportionality, cooperation, increased judicial involvement in case management, and a streamlined sanctions analysis for ESI preservation failures as solutions to those issues.¹² As conceived, the Rule proposals are designed to simplify discovery by making it more efficient and cost effective, thereby allowing matters to be litigated on the merits instead of in costly satellite litigation.¹³

While it remains to be seen whether the proposed amendments can simplify discovery and tackle data growth,¹⁴ the recent introduction of predictive coding has the potential to do so. Predictive coding – a computerized process for selecting and ranking a collection of documents¹⁵ – has found welcome recipients in clients, counsel, and the courts.¹⁶ All parties to the litigation process have generally been drawn to predictive coding given its potential to expedite the ESI search and review process.¹⁷

10. REPORT, *supra* note 8, at 84 (“The 1993 Committee Note further observed that “[t]he information explosion of recent decades has greatly increased both the potential cost of wide-ranging discovery and the potential for discovery to be used as an instrument for delay or oppression.” What seemed an explosion in 1993 has been exacerbated by the advent of e-discovery.”); Richard L. Marcus, *Confronting the Future: Coping with Discovery of Electronic Material*, 64 LAW & CONTEMP. PROBS. 253, 281 (2001) (observing that “sensible behavior by lawyers and judges may be much more useful” than additional changes to the Rules in addressing the problems associated with ESI).

11. See Craig B. Shaffer & Ryan T. Shaffer, *Looking Past The Debate: Proposed Revisions to the Federal Rules of Civil Procedure*, 7 FED. CTS. L. REV. 178 (2013).

12. *Id.* at 178-79.

13. Philip J. Favro, *Getting Serious: Why Companies Must Adopt Information Governance Measures to Prepare for the Upcoming Changes to the Federal Rules of Civil Procedure*, 20 RICH. J. L. & TECH. 5, ¶ 1 (2014).

14. If approved by the Supreme Court before May 1, 2015 and unless Congress acts to modify or reject them, the proposed Rules would be enacted on December 1, 2015. Thomas Y. Allman, *The Civil Rules Package As Approved By the Judicial Conference* (Sept. 2014), <http://www.theediscoveryblog.com/wp-content/uploads/2014/10/2014CommentsOnRulePackage.pdf>.

15. See Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review, with a Foreword by John M. Facciola, U.S. Magistrate Judge*, 7 FED. CTS. L. REV. 1, 26 (2013); Bennett Borden & Jason R. Baron, *Finding the Signal in the Noise: Information Governance, Analytics, and the Future of Legal Practice*, 20 RICH. J.L. & TECH. 7, ¶ 10 (2014).

16. See Andrew Peck, *Search, Forward: Will manual document review and keyword searches be replaced by computer-assisted coding?*, L. TECH. NEWS (Oct. 2011) (generally discussing the benefits of predictive coding) available at https://law.duke.edu/sites/default/files/centers/judicialstudies/TAR_conference/Panel_1-Background_Paper.pdf.

17. See Nat'l Day Laborer Org. Network v. U.S. Immigration & Customs Enforcement Agency, 877 F. Supp. 2d 87, 109 (S.D.N.Y. 2012) (“[P]arties can (and frequently should) rely on .

Lawyers and litigants have additionally gravitated toward predictive coding due to its utility in identifying the key documents required to establish their claims or defenses.¹⁸ These byproducts of predictive coding – simplifying the process and identifying strategic information – make this methodology a particularly attractive option for conducting discovery.¹⁹

Despite the potential that predictive coding holds,²⁰ its introduction to the discovery process has not been universally embraced or free from controversy.²¹ There have been disagreements regarding what is predictive coding,²² when it should be used,²³ and the process for how to successfully implement it into a discovery workflow.²⁴ Moreover, the few judicial opinions on predictive coding are based on specific fact patterns that make general application for practitioners difficult.²⁵ These factors have led to uncertainty regarding the manner in which predictive coding may be used and threaten to impede its proliferation.²⁶

Nowhere is this uncertainty more apparent than in the debate over the application of the attorney work product doctrine to the documents that

. . . machine learning tools to find responsive documents.”); Hon. Patrick J. Walsh, *Rethinking Civil Litigation in Federal District Court*, 40 LITIG. 6, 7 (2013) (urging lawyers to use “21st-century computer technology” including predictive coding to address digital age discovery issues and to stop relying on legacy discovery technologies and methods).

18. Walsh, *supra* note 17, at 7 (“Their goal should be to have the computer sift through the millions of documents and distill and organize the hundreds or thousands of documents that are critical to the case”); Charles Yablon & Nick Landsman-Roos, *Predictive Coding: Emerging Questions and Concerns*, 64 S.C. L. REV. 633, 644 (2013) (describing seed set development and its impact on the need “to identify those documents that are most relevant”).

19. *In re Domestic Drywall Antitrust Litig.*, 300 F.R.D. 228, 233 (E.D. Pa. 2014) (touting the benefits of predictive coding as a new and “sophisticated” discovery search methodology)

20. *Id.*

21. See, e.g., Dana Remus, *The Uncertain Promise of Predictive Coding*, 99 IOWA. L. REV. 1691, 1695 (2014) (suggesting that lawyers “proceed with deliberate care in the use and adoption of predictive-coding technologies”).

22. *Id.* at 1706-07 (observing disapprovingly that “the litigation community is uncritically embracing predictive coding as if its definition is unitary and clear, its accuracy and efficacy well-established”).

23. Borden, *supra* note 15, at ¶ 17 (“[W]e bow to the reality that in a large class of cases the use of predictive coding is currently infeasible or unwarranted, especially as a matter of cost.”).

24. Maura R. Grossman & Gordon V. Cormack, *Comments on “The Implications of Rule 26(g) on the Use of Technology-Assisted Review,”* 7 FED. CTS. L. REV. 285 (2014) (discussing various disagreements between the authors and other thought leaders on the proper use of and objectives surrounding predictive coding).

25. Compare *Progressive Cas. Ins. Co. v. Delaney*, No. 2:11-cv-00678-LRH-PAL, 2014 U.S. Dist. LEXIS 69166, at *27-33 (D. Nev. July 18, 2014) (interpreting case management order provision to prevent the use of predictive coding) with *Bridgestone Americas, Inc. v. Int’l Bus. Mach. Corp. (Bridgestone)*, No. 3:13-1196, 2014 U.S. Dist. LEXIS 142525, at *3 (M.D. Tenn. July 22, 2014) (allowing the use of predictive coding despite an arguably contrary case management order).

26. See generally Remus, *supra* note 21.

counsel selects to train the predictive coding process.²⁷ There is little dispute over the significance that those documents – commonly referred to as training or seed sets – play in a properly functioning predictive coding workflow.²⁸ Seed set documents are essential for training a predictive coding algorithm to identify the documents that “are the most relevant to the case and most representative of those for which each side is looking.”²⁹

Cognoscenti and courts disagree whether the identification of seed set documents is work product and entitled to protection from discovery.³⁰ Some courts and commentators assert that counsel should identify seed documents for its litigation adversaries³¹ despite the potential for disclosing work product. Those who support the disclosure of seed information justify their position on the need for greater certainty in the predictive coding process.³² Disclosure, they argue, will reduce motion practice over the process the party used to search for, review, and produce responsive information.³³

Indeed, sharing seed set documents is seen as an important convenience so that “opposing counsel (and the Court) [are] more comfortable with” predictive coding.³⁴ Courts following this line of analysis have observed that such a step is necessary to alleviate “fears about the so-called ‘black box’ of [predictive coding] technology.”³⁵ Moreover, by insisting on the cooperative development of seed sets, courts are seeking to ensure that discovery is reasonable, proportional, and proceeds in an orderly

27. Richard H. Lowe, James G. Welch & Kimberly G. Lippman, *Disclosure of Seed Sets: Required to Cooperate or Protected as Attorney Work Product?*, THE LEGAL INTELLIGENCER (Feb. 18, 2014), available at http://www.duanemorris.com/articles/disclosure_seed_sets_required_cooperate_protected_attorney_work_product_5140.html (describing how the uncertainty surrounding the issue of whether the work product doctrine applies to seed sets “may serve to stymie the use of predictive coding”).

28. Yablon, *supra* note 18, at 638-39, 642-44 (describing generally the role of seed set documents in training a predictive coding algorithm).

29. *Id.* at 644.

30. Lowe, *supra* note 27; Karl Schieneman & Thomas Gricks, *The Implications of Rule 26(g) on the Use of Technology-Assisted Review*, 7 FED. CTS. L. REV. 239, 262 n.92 (2013) (questioning whether seed sets are entitled to work product protection).

31. See discussion *infra* Part III.B.1.

32. See, e.g., Elle Byram, *The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery*, 29 SANTA CLARA COMPUTER & HIGH TECH. L.J. 675, 699 (“Courts will look more favorably upon a party who discloses its key custodians and how it will [search] for the requested documents. Where a party is transparent, opposing counsel and the Court are more apt to agree to your approach . . .”).

33. *Id.* at 698-699; Schieneman, *supra* note 30, at 261-63 (advocating that seed documents be identified without waiving any applicable work product protection).

34. *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182, 192 (S.D.N.Y. 2012).

35. *Id.*

fashion.³⁶ Eliminating the possibility of collateral litigation should make discovery less costly and more focused on disclosing information so matters can be resolved on the merits.³⁷

The allure of this position, however, fails to recognize that a seed set may reflect a lawyer's perceptions of relevance, litigation tactics, or even its trial strategy.³⁸ These conclusions regarding key strategic issues – memorialized in counsel's selection of documents – have frequently been protected in analogous circumstances³⁹ as work product since they may reveal counsel's "mental impressions, conclusions, opinions, or legal theories."⁴⁰

For example, the specific documents that a lawyer segregated and used to prepare a witness for deposition often merit work product protection.⁴¹ The work product doctrine has also protected the subset of materials that a lawyer identifies during a document inspection.⁴² In addition, litigants have enjoyed immunity from interrogatories that seek the identity of documents supporting their legal theories.⁴³ Underlying each of these lines of authority are two key policies derived from *Hickman v. Taylor*, the Supreme Court's seminal opinion on the work product doctrine.⁴⁴ The first is that counsel is entitled to a zone of privacy to prepare its case for trial.⁴⁵ The second is a logical corollary to the first: that a

36. *Id.* See also Transcript of Record at 13-15, Fed. Hous. Fin. Agency v. J.P. Morgan Chase & Co., Inc., 11-cv-06188-DLC (S.D.N.Y. July 24, 2012) ECF No. 128 (insisting the parties jointly develop the predictive coding workflow, including seed documents).

37. *See id.*

38. Remus, *supra* note 21, at 1716 ("requiring seed-set transparency threatens core protections for attorney work product"); Yablon, *supra* note 18, at 644 ("If . . . the seed set is made up of documents selected or coded by a producing party as relevant, production of that seed set has a much higher probability of disclosing attorney impressions of the case.").

39. See discussion *infra* Part IV.A.3.

40. FED. R. CIV. P. 26(b)(3)(B); United States, *ex rel.* Bagley v. TRW, Inc. (Bagley II), 212 F.R.D. 554 (C.D. Cal. 2003) (holding that the relator's disclosure statement reflecting its counsel's compilation of documents was opinion work product).

41. Sporck v. Peil, 759 F.2d 312 (3rd Cir. 1985) (holding that counsel's selection of certain documents to prepare a client for deposition was protected as opinion work product). See discussion *infra* Part IV.A.1.

42. Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139, 141-44 (D.D.C. 2007) (protecting counsel's selection of materials from a third party document inspection as fact work product); United States, *ex rel.* Bagley v. TRW, Inc. (Bagley I), No. CV94-7755-RAP(AJWx), 1998 U.S. Dist. LEXIS 23585 (C.D. Cal. Dec. 11, 1998) (protecting a lawyer's selection of documents from an opposing party's production as opinion work product). See discussion *infra* Part IV.A.3.

43. See, e.g., Kodak Graphic Commc'ns. Can. Co. v. E. I. du Pont de Nemours & Co., 08-CV-6553T, 2012 U.S. Dist. LEXIS 15752 (W.D.N.Y. Feb. 8, 2012). See discussion *infra* Part IV.A.3.

44. *Hickman v. Taylor*, 329 U.S. 495 (1947). See discussion *infra* Part II.A.

45. *Id.* at 510-11 ("[I]t is essential that a lawyer work with a certain degree of privacy,

litigation adversary should not receive “a free ride on the effort and investment of [] counsel in reviewing and selecting documents and in preparing [its claims] or defense[s].”⁴⁶

In this article, we assert that this rationale is equally applicable to seed set protection.⁴⁷ Counsel should be provided with a “quiet and secluded corner”⁴⁸ to determine which documents are pertinent to the predictive coding process.⁴⁹ This is particularly the case for those lawyers who employ predictive coding to identify and isolate highly relevant data.⁵⁰ Because those documents could very well reflect the manner in which counsel is pursuing discovery and how counsel is establishing what is relevant to the client’s claim or defense, the seed set will disclose counsel’s thought processes, certain conclusions made on the claims and defenses at issue, and/or its strategy for seeking to dispose of the case.⁵¹

Where seed sets actually reflect such information, opposing counsel should not be given a “free ride”⁵² “on wits borrowed from the adversary.”⁵³ Allowing the opposition to understand counsel’s selection process regarding its predictive coding seeds could provide it with access to counsel’s “legal strategy, his intended lines of proof, his evaluation of the strengths and weaknesses of his case,” and prepare its case accordingly.⁵⁴ Under these cases and other authorities,⁵⁵ the identity of those selected documents should merit fact or even opinion work product protection.⁵⁶

While there are limitations to this rule⁵⁷ and though there are obvious potential benefits to disclosure,⁵⁸ we contend that the convenience created

free from unnecessary intrusion by opposing parties and their counsel.”).

46. *Bagley I*, at *4; *Hickman*, 329 U.S. at 516 (Jackson, J., concurring) (“Discovery was hardly intended to enable a learned profession to perform its functions either without wits or on wits borrowed from the adversary.”).

47. See discussion *infra* Parts III.A.3, IV.B.

48. *In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007, 1018-19 (1st Cir. 1988) (finding that counsel’s deposition exhibit list was fact work product, but nonetheless ordered its production).

49. See *In re Biomet M2a Magnum Hip Implant Products Liability Litig. (Biomet II)*, No. 3:12-MD-2391, 2013 U.S. Dist. LEXIS 172570, at *4 (N.D. Ind. Aug. 21, 2013) (“The Steering Committee wants to know, not whether a document exists or where it is, but rather how Biomet used certain documents before disclosing them. Rule 26(b)(1) doesn’t make such information disclosable.”).

50. Yablon, *supra* note 18, at 643-44.

51. See discussion *infra* Part III.A.

52. *Bagley I*, at *4.

53. *Hickman*, 329 U.S. at 516 (Jackson, J., concurring).

54. *Sporck v. Peil*, 759 F.2d 312, 316 (3rd Cir. 1985).

55. See discussion *infra* Part IV.

56. *Id.*

57. See discussion *infra*, Part IV.A.2.

58. *Guidelines Regarding the Use of Predictive Coding*, COALITION OF TECHNOLOGY

by cooperation should not be used to coerce a lawyer into revealing its work product as reflected in the development of a predictive coding seed set.⁵⁹ Instead, where warranted by the circumstances, courts should find that counsel's identification of seed documents are safeguarded as work product.⁶⁰

This article will consider these subjects. We first provide a general overview of the law on attorney work product in Part II. We next discuss in Part III the importance of the seed set to establishing a defensible predictive coding process and the steps counsel typically undertakes to develop that set. This includes a discussion of the differences between judgmental sampling and random sampling and why only those seed sets developed through judgmental sampling should merit consideration for work product protection. We also analyze the unsettled state of predictive coding jurisprudence regarding the protection of seed documents. In Part IV, we describe the general requirements that parties must satisfy for work product protection to apply to a lawyer's selection of documents. We then delineate the circumstances when seed set documents would warrant that protection. We conclude by discussing a few of the instances when fact work product might have to yield in the face of compelling litigation interests.

III. THE ATTORNEY WORK PRODUCT DOCTRINE

Before delving into the details regarding the character and development of predictive coding seed sets, it is essential that we examine the basic notions of the work product doctrine to help frame the discussion. In this Part, we analyze the purposes, policy, and scope of the work product doctrine. This includes a discussion of the elements required to establish a bona fide work product claim and the differences between fact and opinion work product.

A. *The Purposes and Policy of the Work Product Doctrine*

The work product doctrine is one of the cornerstones of modern

RESOURCES FOR LAWYERS 8-10, 14-15 (2014), <http://www.ctrinitiative.com/home/protocol/> (describing some benefits to disclosing seed set documents and other aspects of a predictive coding workflow).

59. *Biomet II*, at *5 (“Biomet, the Steering Committee says, isn’t proceeding in the cooperative spirit endorsed by the Sedona Conference and the corresponding Seventh Circuit project. But neither the Sedona Conference nor the Seventh Circuit project expands a federal district court’s powers, so they can’t provide me with authority to compel discovery of information not made discoverable by the Federal Rules.”).

60. See discussion *infra* Parts III.A.3, IV.B.

litigation practice in the United States.⁶¹ Though perhaps lacking the glamour and prominence of the lawyer-client privilege,⁶² work product plays an important role in ensuring a properly balanced and functioning adversary system.⁶³ This fact was established in 1947 when the Supreme Court issued its celebrated opinion in *Hickman v. Taylor*, which formally recognized the concept of attorney work product.⁶⁴

The work product doctrine is designed to provide a lawyer with the professional space needed to prepare a client's case for trial.⁶⁵ As the Supreme Court observed in *Hickman*, counsel must have "a certain degree of privacy, free from unnecessary intrusion by opposing parties and their counsel."⁶⁶ The work product umbrella does so by enabling counsel to "assemble information, sift what he considers to be the relevant from the irrelevant facts, prepare his legal theories and plan his strategy without undue and needless interference."⁶⁷ The *Hickman* case is particularly instructive on this issue.

In *Hickman*, the court frowned on a party's obtaining through discovery witness statements or legal memoranda reflecting witness discussions prepared by its adversary. The information they possessed was equally accessible to both parties⁶⁸ because the witnesses in question were available for interview by either party.⁶⁹ Given this circumstance, the court reasoned that a litigant should not be permitted to develop its case "on wits borrowed from the adversary."⁷⁰ A contrary result would be "demoralizing" to lawyers,⁷¹ ultimately leading to "[i]nefficiency, unfairness and sharp

61. *United States v. Nobles*, 422 U.S. 225, 238 (1975) (discussing the essential nature of the work product doctrine to a properly functioning adversarial system).

62. Philip Favro, *Inviting Scrutiny: How Technologies are Eroding the Attorney-Client Privilege*, 20 RICH. J.L. & TECH. 2, ¶ 16 (2014) (observing how the lawyer-client privilege has been glamorized in popular culture).

63. *Nobles*, 422 U.S. at 238.

64. *Hickman*, 329 U.S. at 510-12.

65. *Id.* at 511-12.

66. *Id.* at 510.

67. *Id.* at 511.

68. *Id.* at 508, 513-14.

69. *Id.* at 513.

70. *Id.* at 516 (Jackson, J., concurring).

71. *Id.* at 511 ("The effect on the legal profession would be demoralizing. And the interests of the clients and the cause of justice would be poorly served."). The concurrence, authored by Associate Justice Robert Jackson – an accomplished trial lawyer, a former Solicitor General and Attorney General of the United States, and the lead prosecutor for the United States at the Nuremberg war trials, also anticipated this same effect. *Id.* at 516 (Jackson, J., concurring); GAIL JARROW, ROBERT H. JACKSON: NEW DEAL LAWYER, SUPREME COURT JUSTICE, NUREMBERG PROSECUTOR (2008). Justice Jackson observed: "I can conceive of no practice more demoralizing to the Bar than to require a lawyer to write out and deliver to his adversary an account of what witnesses have told him." *Hickman*, 329 U.S. at 516 (Jackson, J., concurring).

practices.”⁷² The work product doctrine, as envisioned by the Supreme Court, was critical to preventing such developments and “essential to an orderly working of our system of legal procedure.”⁷³

B. The Scope of the Work Product Doctrine

Since the Supreme Court’s decision in *Hickman*, the basic thrust of the work product doctrine has been codified in the Rule 26(b)(3)⁷⁴ and is also reflected as well established in other rules⁷⁵ and cases.⁷⁶ Those authorities memorialize the notion that a party generally cannot discover documents, other tangible items, or intangible materials “that are prepared in anticipation of litigation or for trial by or for another party or its representative.”⁷⁷ Significant to the issue of predictive coding seed sets, the work product doctrine may also safeguard selections or compilations of documents made by counsel.⁷⁸

Notwithstanding the importance that the work product doctrine plays in fostering a zone of privacy for lawyers to prepare their cases, the privilege that it affords is often only a qualified one.⁷⁹ The Supreme Court noted as much in *Hickman*, explaining that signed witness statements might nonetheless be subject to discovery for purposes of impeachment when witnesses are unavailable, uncooperative, or where the interests of justice so require.⁸⁰ Consistent with *Hickman*, Rule 26(b)(3) provides that a party may obtain an adversary’s work product if it establishes that the materials “are otherwise discoverable under Rule 26(b)(1),” demonstrates a “substantial need” for their production, and shows that they cannot otherwise be obtained “without undue hardship.”⁸¹

On the other hand, “opinion work product” – those materials that reflect a lawyer’s “legal strategy, his intended lines of proof, his evaluation

72. *Id.* at 511.

73. *Id.* at 512.

74. FED. R. CIV. P. 26(b)(3).

75. *See, e.g.*, FED. R. EVID. 502(g).

76. *See, e.g.*, *Upjohn Co. v. United States*, 449 U.S. 383, 397-98 (1981) (reaffirming *Hickman* and the nature of the work product doctrine).

77. FED. R. CIV. P. 26(b)(3)(A); FED. R. EVID. 502(g). This includes written witness statements, legal memoranda, letters, and “oral expressions of an attorney’s mental impressions, legal theories and subjective evaluations.” *Lockheed Martin Corp. v. L-3 Commc’ns Corp.*, No. 6:05-cv-1580-Orl-31KRS, 2007 U.S. Dist. LEXIS 54606, at *26 (M.D. Fla. Jul. 29, 2007).

78. *See* discussion *infra* Part IV.

79. FED. R. CIV. P. 26(b)(3)(A).

80. *Hickman*, 329 U.S. at 519 (Jackson, J., concurring).

81. FED. R. CIV. P. 26(b)(3)(A). *See In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007 (1st Cir. 1988) (permitting the discovery of a deposition exhibit list that was fact work product).

of the strengths and weaknesses of his case” – enjoy near absolute immunity from discovery.⁸² *Hickman* recognized the primacy of opinion work product, reasoning that “not even the most liberal of discovery theories can justify unwarranted inquiries into the files and the mental impressions of an attorney.”⁸³ Rule 26(b)(3) has captured the importance that *Hickman* placed on safeguarding opinion work product, directing courts to “protect against disclosure of the mental impressions, conclusions, opinions, or legal theories of a party’s attorney or other representative concerning the litigation”⁸⁴ even when they find it discoverable under the Rule 26(b)(3) exception.

In many of the cases involving opinion work product, the mental processes of counsel that courts have sought to protect have been reflected in their selections of documents.⁸⁵ In those matters, courts have proscribed the identification of the selected documents from discovery since they could reveal counsel’s mental processes, conclusions regarding the merits of claims or defenses, or even its strategy for pursuing the litigation.⁸⁶ As set forth in Part III, whether those protections should apply to predictive coding seed sets depends in large part on the underlying process used to compile the seed documents.

IV. PREDICTIVE CODING AND SEED SETS

To grasp why seed set documents may merit work product protection, it is important to understand the role they play in developing an effective predictive coding process. Without seed set documents, it would be difficult to train the predictive coding algorithm.⁸⁷ Indeed, seed documents are essential for teaching the algorithm what materials it should search for

82. Kodak Graphic Commc’ns. Can. Co. v. E. I. du Pont de Nemours & Co., No. 08-CV-6553T, 2012 U.S. Dist. LEXIS 15752, at *13 (W.D.N.Y. Feb. 8, 2012) (quoting *Sporck v. Peil*, 759 F.2d 312, 316 (1985)).

83. *Hickman*, 329 U.S. at 513 (observing that it would indeed be “a rare situation justifying production of these matters.”). The Supreme Court has confirmed this precedent multiple times, most notably in *Upjohn*. *Upjohn Co. v. United States*, 449 U.S. 383, 397-98 (1981).

84. FED. R. CIV. P. 26(b)(3)(B). Various courts have embraced and even extended the protections from *Hickman* and Rule 26(b)(3), declaring that opinion work product is “absolutely privileged.” *Caremark, Inc. v. Affiliated Computer Servs., Inc.*, 195 F.R.D. 610, 616 (N.D. Ill. 2000). Those courts which have not done so have nonetheless held that production of this category of work product should be limited to “rare and extraordinary circumstances.” *Upjohn*, 449 U.S. at 399-402; *Bagley II*, at 559.

85. *Sporck v. Peil*, 759 F.2d 312, 316 (3rd Cir. 1985).

86. *Id.*; see discussion *infra* Part IV.

87. Yablon, *supra* note 18, at 638-39.

across the universe of potentially responsive information.⁸⁸

Although there is little dispute about the indispensable nature of seed sets, there are different schools of thought regarding how they should be developed.⁸⁹ Some have taken the position that a seed set should be created through a random sampling of the entire universe of potentially responsive information.⁹⁰ Others assert that seed sets should instead be customized with documents that lawyers specifically identify through judgment calls on the relevance and importance of a particular record.⁹¹ Still others maintain that a seed set could include a hybrid approach, incorporating elements from both the sampling and judgmental schools.⁹²

While courts have not mandated that one specific approach be used in lieu of the others, the nature of how a seed set is created will certainly impact the application of the work product doctrine.⁹³ In this Part, we delineate the general differences between judgmental and random sampling and consider why only those seed sets that incorporate a lawyer's specific judgment calls on documents merit consideration for work product protection. We also discuss the body of case law that exists regarding the protection of seed set documents. That discussion is critical since the jurisprudence on this issue is far from settled and has yet to directly address the application of the work product doctrine to seed set development.

A. *The Role of Seed Sets in Predictive Coding*

There are various issues that must be considered to establish a properly functioning and defensible predictive coding process.⁹⁴ Issues such as accurately determining the prevalence of responsive information within the universe of documents, engaging in an iterative training process, and validating the final production results through different forms of testing are all essential to the process.⁹⁵ Despite the magnitude of these issues, the step

88. *Id.*

89. *Id.* at 642-44.

90. Schieneman, *supra* note 30, at 259-61 (arguing that random sampling is preferable to judgmental sampling).

91. Grossman, *supra* note 24, at 288, 298 (touting certain benefits of judgmental sampling over pure random sampling).

92. *Id.*; Ralph Losey, *Predictive Coding and The Proportionality Doctrine: A Marriage Made In Big Data*, 26 REGENT U. L. REV. 7, 22 (2013) (describing different sampling methods and the author's preference for using a combined approach).

93. Yablon, *supra* note 18, at 644.

94. See Nicholas Barry, Note, *Man Versus Machine Review: The Showdown Between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology*, 15 VAND. J. ENT. & TECH. L. 343, 354-55 (2013) (describing a predictive coding workflow and related issues associated with its implementation).

95. *Guidelines Regarding the Use of Predictive Coding*, COALITION OF TECHNOLOGY

that is perhaps the most crucial in this process is the proper development of a seed set.⁹⁶ Indeed, the effectiveness of a predictive coding workflow and its ability to satisfy the traditional discovery touchstones of relevance, proportionality, and reasonableness⁹⁷ will likely turn on what documents counsel uses to train the algorithm.⁹⁸

Simply put, a seed set is a proportionately small subset of data that contains examples of the categories of information being sought.⁹⁹ The predictive coding algorithms use the characteristics of the seed set to find similar documents.¹⁰⁰ Those seed documents are then run through the predictive coding process to train the algorithm to obtain documents for production.¹⁰¹

1. Judgmental Sampling

There are two general approaches for developing a seed set.¹⁰² The first is typically referred to as judgmental sampling.¹⁰³ Judgmental sampling involves customizing a seed set with specific documents or classes of information that, when submitted into a predictive coding workflow, are targeted to uncover additional documents whose content is similar to that of

RESOURCES FOR LAWYERS 5-7 (2014), <http://www.ctrlinitiative.com/home/protocol/>.

96. See Daniel Martin Katz, *Quantitative Legal Prediction - Or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 946 (2013) (explaining that predictive coding “approaches are inductive and typically involve the seeding of the algorithm with training (or labeled) data from which the machine infers the ‘true’ function for assigning a document to a particular group (i.e., relevant versus not relevant).”).

97. *Bridgestone*, at *2-3 (“In the final analysis, the uses of predictive coding is [sic] a judgment call, hopefully keeping in mind the exhortation of Rule 26 that discovery be tailored by the court to be as efficient and cost-effective as possible.”).

98. See Katz, *supra* note 96, at 946.

99. Grossman, *supra* note 15, at 29.

100. Yablon, *supra* note 18, at 638-39, 643-44.

101. Transcript of Record at 114, *Fed. Hous. Fin. Agency v. UBS Americas*, 11-cv-06188 (S.D.N.Y. July 31, 2012) ECF No. 134; Yablon, *supra* note 18, at 639; see generally *In re Biomet M2a Magnum Hip Implant Products Liability Litig. (Biomet I)*, No. 3:12-MD-2391, 2013 U.S. Dist. LEXIS 84440, *3 (N.D. Ind. Apr. 18, 2013) (describing the inextricably intertwined process used for seeding and training the predictive coding workflow).

102. Grossman, *supra* note 15, at 29. There are other methods for creating a seed set, the most prominent of which is referred to as “Active Learning” or “Continuous Active Learning.” *Id.* at 8 (“An Iterative Training regimen in which the Training Set is repeatedly augmented by additional Documents chosen by the Machine Learning Algorithm, and coded by one or more Subject Matter Expert(s).”); see also Grossman, *supra* note 24, at 289-90; Ralph Losey, *Latest Grossman and Cormack Study Proves Folly of Using Random Search For Machine Training – Part Three*, E-DISCOVERY TEAM (July 27, 2014), <http://e-discoveryteam.com/2014/07/27/latest-grossman-and-cormack-study-proves-folly-of-using-random-search-for-machine-training-part-three/>.

103. Grossman, *supra* note 15, at 29.

the seed materials.¹⁰⁴ Such a seed set could be comprised of highly relevant information, marginally responsive materials, privileged communications, non-responsive data, or a combination of these or even other items.¹⁰⁵ The specific compilation of a seed set, the number of documents in a seed set, and the number of seed sets created will depend on the facts and circumstances of a particular matter.¹⁰⁶

Judgmental seed documents could be identified through any number of methods. They could include emails, contracts, memoranda, tweets, or other records that a client first isolated and then provided to its counsel.¹⁰⁷ A lawyer might also uncover seed documents by searching through client information or an adversary's production with the use of keywords, concept search, data clustering, manual review, or other search methodologies.¹⁰⁸

2. Random Sampling

In contrast to judgmental sampling, seed set documents obtained through random sampling do not generally involve the exercise of independent legal judgment.¹⁰⁹ Instead, as the term suggests, the seed set is generated by taking a statistically valid sample from the universe of potentially responsive information.¹¹⁰ That sample, typically created through the functionality of a particular predictive coding technology, is designed to ensure that the seed set reflects the characteristics of the entire document population.¹¹¹

Once the randomly comprised seed set is reviewed and coded for responsiveness, that remaining subset of information is run against the universe of documents.¹¹² That subset may or may not be increased with additional samples to ensure that it is truly representative of the overall percentage of responsive information within the universe of documents.¹¹³ The additional samples could be derived from further rounds of random sampling or from judgmental samples that a lawyer obtained through its

104. Yablon, *supra* note 18, at 639, 642-43.

105. Nicholas M. Pace & Laura Zakaras, *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*, RAND INSTITUTE FOR CIVIL JUSTICE 59 (2012), available at <http://www.rand.org/pubs/monographs/MG1208.html>.

106. Losey, *supra* Note 92, at 22.

107. *See id.*

108. Pace, *supra* note 105, at 60; Losey, *supra* note 92, at 22; *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182, 200-01 (S.D.N.Y. 2012).

109. Grossman, *supra* note 15, at 27; Losey, *supra* note 92, at 22.

110. Yablon, *supra* note 18, at 639, 643.

111. Schieneman, *supra* note 30, 260-61.

112. Yablon, *supra* note 18, at 639-40.

113. *Id.* at 643.

exercise of legal reasoning.¹¹⁴

3. The Work Product Doctrine Would Likely Apply Only To Judgmental Samples

Whether the work product doctrine applies to seed documents would depend on a variety of factors.¹¹⁵ However, it seems like a fairly straightforward proposition that a seed set derived exclusively from random sampling methods would not qualify as work product. This is because the grouping of documents would have been generated through computer software and not by counsel's exercise of skill, judgment, and reasoning.¹¹⁶ Indeed, it seems counterintuitive that a seed set created through a mere keystroke of computer functionality would implicate the policy concerns articulated in *Hickman* and its progeny.¹¹⁷ Such functionality does not encompass a lawyer's legal strategy, its intended lines of proof, or its evaluation of the strengths and weaknesses of the case,¹¹⁸ much less the need for a "quiet and secluded corner" to prepare for discovery.¹¹⁹ The plain language of Rule 26(b)(3) and the underlying policy from *Hickman* do not permit such a finding.¹²⁰

On the other hand, seed sets that incorporate elements of judgmental sampling may be entitled to some form of work product protection. While the methods may vary for identifying judgmental seeds, the distinct and overriding characteristic of judgmental seeds is that a lawyer meticulously segregated those materials based on the exercise of legal judgment.¹²¹ At some level, counsel made a determination that a specific seed document was pertinent to the exercise of discovery and the overall direction of the litigation.¹²² For those attorneys who specifically use predictive coding to segregate highly relevant data, judgmental seeds may reflect information

114. Losey, *supra* note 92, at 22.

115. See discussion *infra* Part IV.

116. Yablon, *supra* note 18, at 644 ("[i]f the seed set is merely a random sample of the entire document population and is produced without coding as to whether documents are deemed responsive or not, the production is unlikely to concern a producing party."). While a randomly generated seed set generally lacks the indicia that would warrant work product protection, that may not be the case if counsel subsequently codes the randomly selected documents. Pace, *supra* note 105, at 60. The coding process would likely entail the same exercise of legal reasoning and acumen involved in creating judgmental samples.

117. See discussion *supra* Part II.

118. *Hickman*, 329 U.S. at 510; *Sporck*, 759 F.2d. at 316

119. *In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007, 1018-19 (1st Cir. 1988).

120. FED. R. CIV. P. 26(b)(3); *Hickman*, 329 U.S. 495.

121. Yablon, *supra* note 18, at 643-44.

122. Losey, *supra* note 92, at 22.

beyond mere perceptions of relevance.¹²³ It could memorialize their thought processes, certain conclusions they have made on the claims and defenses at issue, and/or their strategy for seeking to dispose of the case.¹²⁴

Given that judgmental samples are prepared by a lawyer, either in preparation for or in connection with litigation, and that the identity of those samples potentially reflects its “mental impressions, conclusions, opinions, or legal theories,” the identity of those documents could merit work product protection.¹²⁵ The courts, however, have yet to issue a holding that directly addresses this question.¹²⁶ Contrary to the notion that seed documents may be protected from discovery, some of the courts that have focused on seed set development have urged litigants to disclose or cooperatively develop the seeds.¹²⁷ Such decisions seem to gloss over the possibility that the responding party’s seed set could constitute work product.

B. Jurisprudence Regarding the Protection of Seed Set Documents

Nearly three years have transpired since the issuance of the first predictive coding related opinion.¹²⁸ While no court has directly applied the work product doctrine to seed set documents, there are several cases that address issues relating to seed set development.¹²⁹ In most of these cases, seed documents have been identified to litigation adversaries.¹³⁰ In contrast, only the court in *Biomet II* has unequivocally protected the identity of seed documents from discovery.¹³¹ In this Section, we discuss the key jurisprudence on this issue to demonstrate that courts typically favor identification of seed sets to advance the notions of cooperation and transparency.

1. Cases Favoring Disclosure of Seed Sets

The cases involving disclosure of seed sets are generally divided between those instances where the parties voluntarily shared seed documents with the opposition and others where courts ordered the

123. Remus, *supra* note 21, at 1716.

124. *See id.*; Yablon, *supra* note 18, at 644. This analysis should be equally applicable to seed sets developed through the “Active Learning” approach since it includes elements of judgmental sampling. *See* Grossman, *supra* note 24, at 298.

125. *See* discussion *infra* Part IV.B.

126. *See* discussion *infra* Part III.B.

127. *Id.*

128. *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012).

129. *See, e.g., Bridgestone; Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc.*, 11-cv-6189, 2014 U.S. Dist. LEXIS 19156 (S.D.N.Y. Feb. 14, 2014); *Biomet II*.

130. *Da Silva Moore*, 287 F.R.D. at 200-01; *Bridgestone*, at *2.

131. *Biomet II*, at *4-5.

identification of such information.¹³² In either scenario, disclosure occurred to further the cognate aims of cooperation and transparency in discovery.¹³³ Courts strongly believe that adversarial cooperation and openness can decrease discovery costs and delays.¹³⁴ Indeed, even the *Biomet II* decision – which declined to order the production of seed information, touted the “significant, salutary, and persuasive impact on federal discovery practice” of cooperative advocacy.¹³⁵ Thus, in the name of cooperation, courts have encouraged the disclosure or joint creation of seed sets.¹³⁶

Perhaps the most noteworthy example of this occurred in *Da Silva Moore v. Publicis Groupe*, the first predictive coding related opinion.¹³⁷ In that case, the parties entered into a stipulated protocol to use predictive coding.¹³⁸ According to the terms of the protocol, the defendant agreed to provide the plaintiffs with all of the non-privileged seed documents used to train the predictive coding process, together with the defendant’s document coding designations.¹³⁹ In approving this arrangement, the court observed that such “transparency” with respect to seed sets was proper and important.¹⁴⁰ Given the convenience that disclosure provided, the court recommended that this practice should generally be adopted:

Such transparency allows the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called “black box” of the technology. This Court highly recommends that counsel in future cases be willing to at least discuss, if not agree to, such transparency in the computer-assisted review process.¹⁴¹

132. *Compare* In re Actos (Pioglitazone) Products Liab. Litig., 6:11-MD-2299, 2012 U.S. Dist. LEXIS 187519 (W.D. La. July 27, 2012) (reflecting the parties’ agreement to jointly develop the predictive coding seed set) with *Bridgestone*, at *2 (ordering the plaintiff to share its seed set with the defendant).

133. *See, e.g.*, In re Actos (Pioglitazone) Products Liab. Litig., 6:11-MD-2299 (W.D. La. July 13, 2012) ECF No. 1413 (noting the “impressive levels of cooperation, coordination, and negotiation that have taken place among counsel for the plaintiffs and the defense”).

134. *See e.g., Biomet II*, at *5; *Da Silva Moore*, 287 F.R.D. at 192.

135. *Biomet II*, at *5.

136. *Da Silva Moore*, 287 F.R.D. at 192; Transcript of Record at , at 9, 13-15, 24, Fed. Hous. Fin. Agency v. JPMorgan Chase & Co., 11-cv-05201 (S.D.N.Y. July 24, 2012) ECF No. 128.

137. *Da Silva Moore*, 287 F.R.D. 182.

138. *Id.* at 193.

139. *Id.* at 186-87.

140. *Id.* at 192.

141. *Id.* Despite the parties’ agreement to use predictive coding, there is considerable debate regarding whether the defendant’s cooperation and transparency in this regard achieved lower discovery costs. The parties only reached a stipulated protocol after substantial prodding from the court and with the plaintiffs taking the unusual step of “objecting” to the agreement. *Id.* at 187, n.6. The relationship between the parties subsequently deteriorated into protracted motion practice over minor issues of relevance on individual documents. *See, e.g.*, Plaintiffs’ Rule 72(a)

Other decisions have also picked up on the convenience theme from *Da Silva Moore* and insisted that parties who wish to use predictive coding should identify their seed set documents.¹⁴² In *Federal Housing Finance Agency v. UBS Americas*, defendant JPMorgan Chase obtained a court order approving its use of predictive coding over the objections of the plaintiff.¹⁴³ Faced with a universe of over 2.5 million potentially relevant documents, a production deadline approaching within approximately two months, and a skeptical plaintiff who was throwing up roadblocks to the use of predictive coding, JPMorgan felt compelled to seek a court order approving its predictive coding process.¹⁴⁴

While the court approved the use of predictive coding, it required JPMorgan to involve the plaintiff in every step of the predictive coding workflow.¹⁴⁵ This included joint development and full disclosure of the seed set.¹⁴⁶ Drawing on the dicta from *Da Silva Moore*, the court observed that such “transparency and cooperation” was necessary for the overall success of the predictive coding process.¹⁴⁷

A third example of this trend is found in *Bridgestone Americas v. International Business Machines*.¹⁴⁸ In *Bridgestone*, the court approved the plaintiff’s use of predictive coding over the defendant’s objection.¹⁴⁹ The defendant had argued that predictive coding was not appropriate since the parties had previously agreed to a case management order requiring “attorney review.”¹⁵⁰ The court rejected the defendant’s argument and allowed the plaintiff to modify its choice of search methodologies.¹⁵¹ However, because the court permitted the plaintiff to “switch horses in midstream,” it insisted on “openness and transparency” in the predictive

Objection to the Magistrate’s April 25, 2012 Discovery Rulings, *Da Silva Moore v. Publicis Groupe*, No. 11-cv-1279 (S.D.N.Y. May 9, 2012), ECF No. 190 (seeking reconsideration of various court rulings on discovery issues relating to the defendant’s use of predictive coding).

142. *Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc.*, 11-cv-6189, 2014 U.S. Dist. LEXIS 19156 (S.D.N.Y. Feb. 14, 2014).

143. *Id.* at *3.

144. Transcript of Record at 10-15, *Fed. Hous. Fin. Agency v. JPMorgan Chase & Co.*, No. 11-cv-05201 (S.D.N.Y. July 24, 2012) ECF No. 128.

145. *Id.* at 9, 13-15, 24.

146. *Id.* at 24 (excepting privileged materials, the court expected “that the plaintiff will be given full access to . . . documents within the seed set”).

147. *Id.* at 9 (“And for this entire process to work, I think it needs transparency and cooperation of counsel.”).

148. *Bridgestone*.

149. *Id.* at *2-3.

150. See Case Management Order, *Bridgestone Americas, Inc. v. Int’l Bus. Mach. Corp.*, No. 3:13-cv-1196 (M.D. Tenn. May 20, 2014) ECF No. 76; see also IBM’s Response to Bridgestone’s Request to Add Predictive Coding to the ESI Protocol, *Bridgestone Americas, Inc. v. Int’l Bus. Mach. Corp.*, No. 3:13-cv-1196 (M.D. Tenn. July 14, 2014) ECF No. 85.

151. *Bridgestone*, at *2.

coding process.¹⁵² Indeed, in response to the plaintiff's offer to share its seed documents with the defendant, the court essentially incorporated that proposal into its order, reiterating its expectation of "full openness" between the litigants and their lawyers.¹⁵³

Da Silva Moore, *Federal Housing*, *Bridgestone* and other cases¹⁵⁴ exemplify the judiciary's general expectation that parties identify their seed set documents for convenience and expediency in discovery. That trend, though, has not been universally followed by the courts.

2. Cases Favoring Protection of Seed Sets

In contrast to the decisions favoring disclosure stand those instances where courts have protected seed sets from discovery.¹⁵⁵ The most notable decision in this regard and the lone opinion to directly address this issue is *Biomet II*.¹⁵⁶ The *Biomet II* court denied the plaintiffs' request that the defendant pharmaceutical company identify the seed set documents that were used to train its predictive coding algorithm.¹⁵⁷ The plaintiffs were seeking to develop search terms to run against the subset of documents that the defendant's predictive coding process excluded from its production.¹⁵⁸ The plaintiffs asserted that identification of the seed documents¹⁵⁹ was necessary so they could prepare their own search terms.¹⁶⁰ Unless they were given an understanding of the defendant's seed documents, argued the plaintiffs, they could not effectively prepare search terms.¹⁶¹

The court rejected that argument and held that the plaintiffs' request for seed information was beyond the scope of permissible discovery.¹⁶²

152. *Id.*

153. *Id.* at *3.

154. *See, e.g.*, *Indep. Living Ctr. of S. Cal. v. City of L.A.*, 2:12-cv-00551-FMO-PJW, at *1 (C.D. Cal. June 13, 2014) ECF No. 371 (ordering the defendant "to use a predictive coding system for identifying the 10,000 most relevant documents in its databases and, after reviewing them for privilege, etc., produc[e] them to Plaintiffs.").

155. *Biomet II*; *Gordon v. Kaleida Health*, No. 08-CV-378S(F), 2013 U.S. Dist. LEXIS 73330, *6-7 (W.D.N.Y. May 21, 2013) (rejecting plaintiffs' motion to compel the production of "important information regarding Defendants' selection of so-called 'seed set documents' which are used to 'train the computer' in the predictive coding search method."); *Hinterberger v. Catholic Health Sys.*, 08-CV-380S(F), 2013 U.S. Dist. LEXIS 73141 (W.D.N.Y. May 21, 2013) (reaching the same holding as in the related *Gordon* case).

156. *Biomet II*.

157. *Id.* at *3-5.

158. *Id.* at *2

159. It appears that the defendant resorted to a form of judgmental sampling to prepare its seed set. *Biomet I*, at *3.

160. *Biomet II*, at *3.

161. *Id.*

162. *Id.* at *3-5.

Because the defendant had already produced all of the responsive seed set documents to the plaintiffs, the court reasoned that the plaintiffs were not seeking to ascertain “whether a document exists or where it is, but rather *how Biomet used certain documents before disclosing them.*”¹⁶³ That type of information, the court explained, was not within the permissible scope of discovery.¹⁶⁴

Lastly, the court sympathized with but ultimately disregarded the plaintiffs’ complaints that the company was not being cooperative.¹⁶⁵ While touting the virtues of cooperation enshrined in both the Seventh Circuit Pilot Program¹⁶⁶ and The Sedona Conference Cooperation Proclamation,¹⁶⁷ the court observed that “neither the Sedona Conference nor the Seventh Circuit project expands a federal district court’s powers, so they can’t provide me with authority to compel discovery of information not made discoverable by the Federal Rules.”¹⁶⁸ Moreover, even though the company’s lack of cooperation was “troubling” and could possibly lead the court to later conclude it was “hiding something,” the court took no action since it had no “discretion” to order the identification of the seed information.¹⁶⁹

Biomet II stands for the proposition that Rule 26(b)(1) may protect the identity of seed set documents from discovery.¹⁷⁰ While notions of cooperation and transparency are significant to advancing the tripartite aims of Rule 1 in litigation, they cannot supersede the strictures from Rule 26(b)(1). Absent some showing of relevance or a stipulation between the parties, *Biomet II* teaches that the identification of seed set documents should generally be proscribed from discovery.

It is also significant that the court specifically forbade the plaintiffs from discovering the seeds because they sought to understand how the company used those documents in the discovery process.¹⁷¹ Probing the manner in which a party and its counsel selected documents for purposes of satisfying discovery obligations could very well reflect their deliberations on claims and defenses, along with related tactical choices. These items

163. *Id.* at *4 (emphasis added).

164. *Id.*

165. *Id.* at *5.

166. 7TH CIR. ELEC. DISCOVERY COMM., PRINCIPLES RELATING TO THE DISCOVERY OF ELECTRONICALLY STORED INFORMATION, at princs. 1.01-.03 (2010), available at http://www.discoverypilot.com/sites/default/files/Principles8_10.pdf.

167. The Sedona Conference, *The Sedona Conference Cooperation Proclamation*, 10 SEDONA CONF. J. 331 (2009).

168. *Biomet II*, at *5.

169. *Id.* at *5-6.

170. *Id.* at *3-5.

171. *Id.* at *4.

thus fall squarely within the ambit of Rule 26(b)(3)'s qualified prohibition on the discovery of fact work product and its admonition against disclosing opinion work product.¹⁷² Absent a showing of a substantial or compelling need for such information – let alone a basic demonstration of relevance, the *Biomet II* court properly precluded the plaintiffs' inquiry. Its reasoning supports the conclusion that a judgmental seed set must be protected as work product.

Despite what we believe to be the implicit inference from *Biomet II* that seed sets are non-discoverable and work product, the fact remains that no court has specifically analyzed whether the work product doctrine applies to seed sets. Juxtaposed against this absence of authority are *Da Silva Moore* and its progeny, whose holdings in favor of seed set identification sharply contrast with *Biomet II*. Given the conflict in these authorities and the absence of predictive coding decisions on this issue, other case law must be explored for guidance. The jurisprudence that provides the most logical analogue on this issue involves decisions that have addressed the merits of work product protection over a lawyer's selection of documents in other less technical scenarios.

V. PROTECTING SEED SET DOCUMENTS AS WORK PRODUCT UNDER ANALOGOUS SCENARIOS INVOLVING SELECTIONS OF DOCUMENTS

The question of whether work product protection should ultimately apply to judgmental seed sets turns on the line of analogous cases that have protected selections of documents as either opinion or fact work product. In this Part, we detail the circumstances under which a lawyer's selection of documents would qualify as work product. This includes a discussion of *Sporck v. Peil*, the foundational case on this issue.¹⁷³ We then analyze how the rule from *Sporck* has been clarified, refined, and played out in specific case scenarios. Finally, we consider the circumstances when seed documents should be protected in light of those authorities and the circumstances of a case.

A. *Selections of Documents: When Are They Protected as Work Product?*

An enduring question over the past thirty years of work product jurisprudence is whether courts should protect the identity of certain documents that counsel compiles or selects in connection with its case

172. See discussion *supra* Part II.

173. *Sporck*, 759 F.2d at 312.

preparations.¹⁷⁴ Since the issuance of *Sporck v. Peil* by the United States Court of Appeals for the Third Circuit in 1985,¹⁷⁵ the answer has been a definite “yes.”¹⁷⁶ *Sporck* unequivocally held that selections of documents could be protected as opinion work product.¹⁷⁷ Since that time, however, the *Sporck* rule has been refined as subsequent opinions delineated the circumstances under which *Sporck* would apply.¹⁷⁸ In this Section, we review the development of the selection of documents rule and what parties must generally show to satisfy its present requirements.

1. Document Selections Protected as Opinion Work Product under *Sporck*

The seminal case on the work product status of documents chosen from a larger population is the Third Circuit’s *Sporck* decision.¹⁷⁹ In *Sporck*, the court held that a lawyer’s selection of a “few documents out of thousands” constitutes protected work product.¹⁸⁰ The issue arose in the context of a deposition of the defendant.¹⁸¹ During the deposition, plaintiff’s counsel asked the deponent whether he had reviewed any documents in preparation for the deposition.¹⁸² After answering yes, the defendant declined to identify the documents he had reviewed.¹⁸³ Explaining that the documents in question had all been produced in discovery, the defendant’s lawyer asserted that their particular selection in this context – to prepare the client for deposition – was protected as work product.¹⁸⁴ While the lower court agreed with this assertion, it nonetheless ordered the selected documents identified since they did not constitute opinion work product.¹⁸⁵

The Third Circuit reversed this finding and instead held that counsel’s compilation of documents was opinion work product and protected from

174. *Charleswell v. Chase Manhattan Bank, N.A.*, 277 F.R.D. 277, 282-83 (D.V.I. 2011) (confirming that a lawyer’s compilation of documents may be protected as work product).

175. *Sporck*, 759 F.2d at 312.

176. *See, e.g., Shapiro v. United States Dep’t of Justice*, 969 F. Supp. 2d 18, 31-32 (D.D.C. 2013) (citing *Hickman* and holding that “there is a clear policy rationale for protecting the compilation of records” as work product).

177. *Sporck*, 759 F.2d at 315.

178. *See discussion infra* Part IV.A.2.

179. *Sporck*, 759 F.2d 312.

180. *Id.* at 316 (quoting *James Julian, Inc. v. Raytheon Co.*, 93 F.R.D. 138, 144 (D. Del. 1982)).

181. *Id.* at 314.

182. *Id.* at 313-14.

183. *Id.* at 314.

184. *Id.*

185. *Id.*

discovery.¹⁸⁶ The court reasoned that disclosure of this information could “reveal important aspects of [counsel’s] understanding of the case” and thereby divulge his mental impressions.¹⁸⁷ Foreshadowing the need for such work product protection in today’s ESI-driven discovery process, the court declared that safeguarding such document compilations was particularly important in cases “involving extensive document discovery.”¹⁸⁸ Citing *Hickman*, the court also observed that the document selection process was an essential aspect of case preparation.¹⁸⁹ Counsel’s assembling of information, sifting “the relevant from the irrelevant facts,” preparing “legal theories,” and developing “strategy without undue and needless interference” were inherent in this process.¹⁹⁰ Disclosing counsel’s document selection would inappropriately reveal these mental processes to a litigation adversary.¹⁹¹

2. Refining the Scope of the *Sporck* Rule

Sporck stands for the proposition that a lawyer’s selection of documents may qualify as opinion work product.¹⁹² While many courts have followed *Sporck*,¹⁹³ other opinions have refined the scope of its holding to delineate more precisely when work product protection should apply. That jurisprudence has generally resulted in four clarifying principles to the application of the work product doctrine to a lawyer’s selection of documents.

The most significant of these principles is that the party asserting a claim of work product must show that the identification of the selected documents raises a genuine threat of disclosure of counsel’s mental impressions.¹⁹⁴ The purpose of this principle is to distinguish meritorious

186. *Id.* at 316.

187. *Id.* (quoting *James Julian, Inc. v. Raytheon Co.*, 93 F.R.D. 138, 144 (D. Del 1982)).

188. *Id.* (quoting *James Julian, Inc. v. Raytheon Co.*, 93 F.R.D. 138, 144 (D. Del 1982)).

189. *Id.* at 316-17 (citing *Hickman v. Taylor*, 329 U.S. 495 (1947)).

190. *Id.* at 316 (quoting *Hickman v. Taylor*, 329 U.S. 495 (1947)).

191. *Id.* at 317.

192. *Id.*; accord *Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, 242 F.R.D. 139, 142 (D.D.C. 2007).

193. See, e.g., *Shelton v. American Motors Corp.*, 805 F.2d 1323, 1328 (8th Cir. 1986) (holding that a lawyer’s document selection should be protected as work product since she “identified, selected, and compiled documents that were significant to her client’s defenses in this case.”); *In re Allen*, 106 F.3d 582, 608 (4th Cir. 1997) (determining that counsel’s choice and arrangement of certain materials constituted opinion work product).

194. *Gould v. Mitsui Mining & Smelting Co., Ltd.*, 825 F.2d 676, 680 (2nd Cir. 1987) (“the selection and compilation of documents by counsel for litigation purposes is protected opinion work product . . . but its application depends upon the existence of a real, rather than speculative, concern that the thought processes of [] counsel in relation to pending or anticipated litigation would be exposed.”).

work product claims from those that are specious.¹⁹⁵ Memorialized in various cases,¹⁹⁶ this principle has become a standard limitation on the application of the *Sporck* rule.¹⁹⁷

The second clarifying principle is that the *Sporck* rule generally applies to the identification of relevant documents and typically does not prevent the documents themselves from being produced in litigation.¹⁹⁸ The basis for this principle is that withholding relevant, non-privileged documents from discovery generally does not promote the policies supporting the work product doctrine.¹⁹⁹ Only in exceptional circumstances could a party justify keeping the selected documents from discovery.²⁰⁰

The next refining principle is that a lawyer's compilation of documents may only constitute fact work product in certain instances.²⁰¹ Such a restriction on the *Sporck* rule was necessary since "not every item which may reveal some inkling of a lawyer's mental impressions, conclusions, opinions, or legal theories is protected as opinion work product."²⁰² This is particularly the case where counsel has "no justifiable expectation that the mental impressions revealed by the materials will remain private."²⁰³ This principle is exemplified in *In re San Juan Dupont Plaza Hotel Fire Litigation*, which ordered that a lawyer's exhibit list be produced in advance of a deposition.²⁰⁴ Although the list was fact work product, it could not attain opinion work product status because the identified documents would be shared with opposing counsel.²⁰⁵

195. See *In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007, 1015-16 (1st Cir. 1988).

196. See, e.g., *Gould*, 825 F.2d 676 (articulating the principle, but declining to rule on the matter and remanding for determination by the lower court); *Mercator Corp. v. United States* (In re Grand Jury Subpoenas Dated March 19, 2002 & August 2, 2002), 318 F.3d 379, 386-87 (2nd Cir. 2002) (holding that the movant did not satisfy its burden on the issue since it neglect to disclose *ex parte* to the court the "defense strategy" that the selected documents would allegedly reveal).

197. See, e.g., *In re Trasylol Prods. Liab. Litig.*, No. 08-MD-1928, 2009 U.S. Dist. LEXIS 85553 (S.D. Fla., Aug. 12, 2009) (reasoning that "a party asserting an attorney opinion privilege claim must come forward with some evidence that disclosure of the requested documents creates a real, non-speculative danger of revealing counsel's thoughts."); *Disability Rights*, 242 F.R.D. at 143 (noting the limitation of the *Gould* holding on *Sporck*).

198. See, e.g., *Mercator*, 318 F.3d 379.

199. *Id.* at 384-85.

200. See *Gould*, 825 F.2d at 680.

201. *San Juan*, 859 F.2d 1007.

202. *Id.* at 1015.

203. *Id.* at 1016.

204. *Id.* at 1018-19.

205. *Id.* at 1017-19 ("the exhibits are integral to the taking of the deposition and will, by definition, have to be revealed during the session."). As the *San Juan* court acknowledged, however, such a scenario is distinguishable from *Sporck* where the documents used to prepare a witness for a deposition "were never meant to be placed on public display." *Id.* at 1018.

The fourth and final clarifying principle is that courts will only safeguard as fact work product document compilations that are considerable in size.²⁰⁶ This is because courts find it unlikely that an adversary can glean a lawyer's mental processes when a voluminous number of documents comprise the selection.²⁰⁷ While the meaning of "voluminous" will vary depending on the circumstances of a particular matter, various cases have found that selections that number in the tens of thousands will typically only merit fact work product status.²⁰⁸

Despite this clarification, caution should be taken to ensure the rule is applied in a proportional fashion to ensure the achievement of *Sporck*'s overarching objective: protecting counsel's mental impressions in cases involving "extensive document discovery."²⁰⁹ As *Sporck* observed, "the process of selection and distillation [of documents] is often more critical than pure legal research" in cases involving large volumes of documents.²¹⁰ All of which suggests that a seed set numbering in the thousands or more could warrant opinion work product status since today's cases frequently involve millions – not thousands – of documents.²¹¹

3. Application of *Sporck* in Factual Scenarios Involving Selections of Documents

The foregoing principles confirm the continuing vitality of the *Sporck* rule, *i.e.*, that a lawyer's selection of documents can merit work product protection. So long as a party demonstrates an actual threat of disclosure of its counsel's mental processes and only seeks to withhold the identity of the compiled documents, the selection could warrant fact or even opinion work product protection. The context in which courts have applied the *Sporck* rule in specific factual scenarios is instructive on how it should apply to predictive coding seed sets.

206. Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139, 141-44 (D.D.C. 2007).

207. Wollam v. Wright Med. Group, Inc., No. 10-cv-03104-DME-BNB, 2011 U.S. Dist. LEXIS 106768, *5-6 (D. Colo. Sept. 20, 2011) (granting work product protection to the compilation of documents at issue).

208. *Disability Rights*, 242 F.R.D. at 144 (citing various cases).

209. *Sporck*, 759 F.2d at 316

210. *Id.*; accord *Shelton v. American Motors Corp.*, 805 F.2d 1323, 1329 (8th Cir. 1986) ("[i]n cases that involve reams of documents and extensive document discovery, the selection and compilation of documents is often more crucial than legal research.").

211. Such a rule seems logical, particularly given that *Sporck* sought to protect the disclosure of "a few documents out of thousands." *Sporck*, 759 F.2d at 316.

a. Preparing Witnesses for Deposition

One such scenario is the identification and selection of documents in connection with preparing a witness for deposition.²¹² A selection of documents is safeguarded as opinion work product where the documents counsel selects have been produced in litigation and where their disclosure would likely divulge counsel's mental processes.²¹³ This is precisely the fact pattern considered in *Sporck* and which the Third Circuit found to merit opinion work product protection.²¹⁴ Various courts have followed the specific *Sporck* holding,²¹⁵ including those cases which found *Sporck* factually inapposite.²¹⁶

b. Document Inspections

Another protected scenario that has significance for seed sets involves the subset of information that a lawyer identifies during a document inspection.²¹⁷ Courts have frequently forbidden adversaries from learning the precise documents that a lawyer picked during an inspection.²¹⁸ A contrary result would enable opposing counsel to receive a "free ride" on the lawyer's work in choosing documents that pertain to its claims or

212. *Sporck*, 759 F.2d 312; *Briese Lichttechnik Vertriebs GmbH v. Langton*, 272 F.R.D. 369, 376 (S.D.N.Y. 2011) ("the attorney's decision to select specific documents to show the witness embodies the lawyer's mental processes, specifically, her evaluation of the significance of those documents, and hence is presumptively protected work-product.").

213. *See, e.g., In re Yasmin & Yaz (Drospirenone) Mktg., Sales Practices & Prods. Liab. Litig.*, No. 3:09-md-02100, 2011 U.S. Dist. LEXIS 69711 (S.D. Ill. June 29, 2011) ("Forcing Bayer to disclose the compilation would implicitly reveal the thought processes of the attorney who selected the documents and would allow plaintiffs to glean which documents, out of the millions already produced, opposing counsel believes are legally significant."). *Contra New Jersey v. Sprint Corp.*, 258 F.R.D. 421, 436 (D. Kan. 2009) ("Having prepared literally hundreds of witnesses for deposition and trial while in private practice, the undersigned simply believes it is too big a leap to suggest that the mere identification of documents a witness reviews at the direction of counsel improperly provides a roadmap of the attorney's strategies and opinions.").

214. *Sporck*, 759 F.2d at 316.

215. *See, e.g., Briese*, 272 F.R.D. at 376; *Yasmin*, at *6-7.

216. *Mercator Corp. v. United States (In re Grand Jury Subpoenas Dated March 19, 2002 & August 2, 2002)*, 318 F.3d 379, 385 (2nd Cir. 2002) (holding that *Sporck* could safeguard the identity of certain information when the objective of a request was to obtain "the opposing attorney's thinking or strategy"); *In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007, 1018 (1st Cir. 1988) (confirming the viability of *Sporck* in the context of preparing a witness for deposition).

217. *Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, 242 F.R.D. 139, 141-44 (D.D.C. 2007).

218. *Id.* at 144. *Contra In re Shell Oil Refinery*, 125 F.R.D. 132, 134 (E.D. La. 1989) (ordering the production of documents over the plaintiffs' work product objection as it was unlikely that counsel's "theory of the case" could be revealed by "the 65,000 documents out of 660,000 documents . . . selected for copying.").

defenses.²¹⁹ The *Bagley I* and *Disability Rights* cases are particularly instructive on this issue.²²⁰

In *Bagley I*, the court proscribed the government from learning the identity of the documents the defendant manufacturer selected during an inspection of the government's records.²²¹ The government had argued that for the sake of efficiency and convenience, the manufacturer should identify the documents its counsel had selected.²²² While the court sympathized with the government,²²³ it found that the manufacturer's selection of documents would divulge information regarding its lawyer's "mental impressions or strategy" and "therefore constitutes opinion work product."²²⁴

Similarly, the court in *Disability Rights* protected as fact work product the documents that were selected from certain records made available to the litigants during a third party document inspection.²²⁵ The court found the existence of an actual threat that counsel's thought processes could be divulged if the documents were identified.²²⁶ Nevertheless, opinion work product protection could not apply since the documents in the compilation numbered in the tens of thousands.²²⁷

Bagley I and *Disability Rights* demonstrate that counsel's mental processes would be improperly revealed if the precise nature of the documents selected during a document inspection were identified. Whether the selection of documents should attain fact or opinion work product status, however, depends on whether the compiled materials are proportionately voluminous in relation to the overall production set.²²⁸

219. *Bagley I*, at *4; *Bagley II*, at 564.

220. *Bagley I; Disability Rights*, 242 F.R.D. 139.

221. *Bagley I*, at *4-5.

222. *Id.* at *4.

223. *Id.* ("[v]iewed strictly from the standpoint of efficiency, the procedure proposed by the government probably is superior to any of the alternatives The Court is sympathetic with the general philosophy expressed in some of the cases cited by the government, namely, that full disclosure, conservation of resources, and judicial efficiency usually should predominate over tactical partisan concerns.").

224. *Id.* at *3-5. *See Bagley II*, at 564.

225. *Disability Rights*, 242 F.R.D. at 144-45.

226. *Id.* at 142-44.

227. *Id.* at 143-44 (explained that "it would be difficult to conceive that Plaintiffs' trial strategy could be gleaned solely by virtue of Plaintiffs' disclosure of the documents selected."). The defendants had argued that the compilation of documents at issue would not expose counsel's mental process since "the number of documents totals over 40,000 pages." *Id.* at 142.

228. *See discussion supra* Part IV.A.2.

c. Responses to Contention Interrogatories

Another document selections analogy instructive on the seed set front involves the immunity that litigants often receive from preparing responses to contention interrogatories that seek the identity of documents supporting their claims or defenses.²²⁹ The purpose for this immunity is to protect litigants from a “thinly-veiled effort” by their adversaries to learn about how counsel “intends to marshal [sic] the facts, documents and testimony” in support of its positions.²³⁰ The need for such protection is exemplified in the *Kodak Graphic Communications Canada Company v. E. I. du Pont de Nemours and Co.* case.²³¹

In *Kodak*, the court proscribed the defendant from discovering the precise documents supporting one of the plaintiff’s legal theories.²³² Such an inquiry was calculated to obtain the plaintiff’s counsel’s mental processes as reflected by the requested compilation of documents.²³³ This was all the more apparent since all of the relevant documents had been produced in discovery and the requested compilation would be principally populated from materials derived from the defendant’s production.²³⁴ Since the plaintiff agreed to disclose the facts supporting its theory, the defendant could simply review the productions to find the supporting documents.²³⁵

Kodak teaches that selections of documents should not be disclosed to provide an adversary with a shortcut to completing its own trial preparations.²³⁶ While convenient for opposing counsel, such a result would

229. See *Johnson v. Ocean Ships, Inc.*, No. C05-5615RJB, 2006 U.S. Dist. LEXIS 52281 (W.D. Wash. July 31, 2006) (relying on *Sporck* and finding that “the exact documents and witnesses [the defendant] intends to use for each affirmative defense reveals defense counsel’s mental impressions, is work product and so is privileged.”).

230. *Kodak Graphic Commc’ns. Can. Co. v. E. I. du Pont de Nemours & Co.*, 08-CV-6553T, 2012 U.S. Dist. LEXIS 15752, *13 (W.D.N.Y. Feb. 8, 2012).

231. *Id.* at *13-18.

232. *Id.* at *17-18.

233. *Id.* at *17.

234. *Id.* at *9, *18.

235. *Id.* at *18.

236. *Id.* at *13; accord *FDIC v. Brudnicki*, 291 F.R.D. 669, 679 (N.D. Fla. 2014) (citing *Sporck* and declining on work product grounds to order the defendant to categorize documents in response to “contention requests”). *But see SEC v. Collins & Aikman Corp.*, 256 F.R.D. 403 (S.D.N.Y. 2009) (distinguishing the *Sporck* rule and ordering the plaintiff to produce documents that supported allegations from its complaint). We appreciate that in *SEC v. Collins & Aikman Corp.*, the court compelled over the plaintiff’s work product objection the disclosure of “approximately 175 file folders” that its counsel created to categorize materials supporting the factual allegations in its complaint. We see a marked difference between being compelled to provide one’s evidentiary support for an allegation and demanding that counsel identify what documents it deemed relevant to a legal theory advanced by the client. The former seeks a fact – what documents support the allegation in paragraph 1 – while the latter probes improperly into how counsel has constructed the legal theory it is advancing on behalf of the client. See *Hickman*

run contrary to the policies underlying the work product doctrine articulated in *Hickman* and *Sporck*.²³⁷

B. Application of the Sporck Rule to Predictive Coding Seed Sets

The analogous factual scenarios detailed above, taken together with *Sporck* and its four clarifying principles, demonstrate that seed sets prepared using judgmental sampling methods should be protected either as opinion or fact work product. For example, the rationale from the “witness preparation” line of cases – beginning with *Sporck* – that protects document compilations as opinion work product is equally applicable to judgmental seed documents.

Like the documents that counsel segregates for preparing a witness for deposition, the judgmental seeds that counsel identifies represent a proportionately limited number of documents.²³⁸ In addition, those seed documents – as evidenced by *Biomet II* – are not intended for public dissemination.²³⁹ Instead, counsel prepares judgmental seeds to sift “relevant from the irrelevant facts,” isolate the most relevant information, and thereby “prepare his legal theories and plan his strategy.”²⁴⁰ These factors regarding the purpose and preparation of judgmental seeds demonstrate that disclosure of the seeds raises a genuine threat that counsel’s mental processes would be exposed to its litigation adversaries.

Moreover, protecting judgmental seeds from disclosure also implicates the policy concerns from *Sporck* regarding cases that involve “extensive document discovery.”²⁴¹ While such large scale cases in the 1980s may have involved productions numbering in the tens or hundreds of thousands of pages, discovery now routinely involves millions of

v. Taylor, 329 U.S. 495, 510-11 (1947).

237. *Kodak*, at *9-10 (“The purpose of the attorney work product doctrine is ‘to preserve a zone of privacy in which a lawyer can prepare and develop legal theories and strategy . . . free from unnecessary intrusion by his adversaries.’”) (quoting *United States v. Adlman*, 134 F.3d 1194, 1196 (2nd Cir. 1998)).

238. *Edwards v. Nat’l Milk Producers Fed’n*, No. 3:11-CV-04766-JSW, 3 (N.D. Cal. Apr. 17, 2013), ECF No. 155 (specifically identifying the composition of the seed set as initially including approximately 1,400 documents); Jason Lichter, *Transparency in Predictive Coding: How Much Is Too Much?*, NEW YORK LAW JOURNAL (Mar. 17, 2014) available at http://www.pepperlaw.com/publications_article.aspx?ArticleKey=2887 (explaining that a seed set could range from hundreds to “tens of thousands of documents,” depending on their development).

239. *Biomet II*, at *2.

240. *Hickman*, 329 U.S. at 510-11; see *Biomet II*, at *2-4.

241. *Sporck*, 759 F.2d at 316 (quoting *James Julian, Inc. v. Raytheon Co.*, 93 F.R.D. 138, 144 (D. Del. 1982)).

documents and petabytes of information.²⁴² Measured proportionately under *Sporck*, seed sets involving thousands of documents are properly protected as opinion work product. And should the number of judgmental seeds be deemed too large to reveal counsel's mental processes,²⁴³ a court should nonetheless protect the seed set as fact work product and require the movant to demonstrate "substantial need" and "undue hardship" under Rule 26(b)(3).²⁴⁴

The "document inspections" and "contention interrogatories" jurisprudence also justify the protection of judgmental seeds as opinion or fact work product. Just as the courts reasoned in *Bagley I* and *Kodak*,²⁴⁵ disclosing judgmental seeds could advance the efficiency of litigation and make the pursuit of discovery more convenient for opposing counsel and perhaps the court. However important those objectives are, they cannot supersede the policies that *Hickman* articulated and that are now codified in Rule 26(b)(3).²⁴⁶ Lawyers and their clients must be given a zone of privacy to prepare a properly functioning predictive coding workflow without "needless interference" from their adversaries.²⁴⁷

Nor should the opposition be given "a free ride on the effort and investment" of counsel in meticulously preparing its judgmental seeds.²⁴⁸ A contrary result would make the court a party to an adversary's "thinly-veiled effort" to discover a lawyer's litigation and trial strategy as reflected in that compilation of documents.²⁴⁹ As *Biomet II* confirmed, it is generally improper for opposing counsel to learn how judgmental seeds were used before producing them in discovery.²⁵⁰

In summary, judgmental seeds should generally be protected as opinion work product when they satisfy the *Sporck* rule and its refining principles. Nevertheless, as *San Juan* and its progeny have held, not every seed set will merit this type of protection.²⁵¹ There will be instances when

242. See generally Philip J. Favro, *Sea Change or Status Quo: Has The Rule 37(e) Safe Harbor Advanced Best Practices for Records Management?*, 11 MINN. J.L. SCI. & TECH. 317 (2010) (detailing the data governance challenges that organizations face from the information explosion).

243. See Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth., 242 F.R.D. 139, 142-44 (D.D.C. 2007).

244. *Id.*; FED. R. CIV. P. 26(b)(3)(A).

245. *Bagley I*, at *3-5; *Kodak Graphic Communs. Can. Co. v. E. I. du Pont de Nemours & Co.*, 08-CV-6553T, 2012 U.S. Dist. LEXIS 15752, *13-18 (W.D.N.Y. Feb. 8, 2012).

246. *Hickman*, 329 U.S. at, 510-11; FED. R. CIV. P. 26(b)(3)(B).

247. *Hickman*, 329 U.S. at 511.

248. *Bagley I*, at *4. See also *Hickman*, 329 U.S. at 516 (Jackson, J., concurring).

249. *Kodak*, at *13.

250. *Biomet II*, at *3-4.

251. *In re San Juan Dupont Plaza Hotel Fire Litig.*, 859 F.2d 1007, 1015 (1st Cir. 1988); *Northern Natural Gas Co. v. Approximately 9117.53 Acres*, 289 F.R.D. 644, 647-50 (D. Kan.

the quantity of judgmental seeds is simply too large in proportion to the overall production set and will not require opinion or perhaps even fact work product protection.²⁵² Legitimate case management considerations could also warrant the disclosure of seed documents.²⁵³ Discovery of seed information might also be ordered when reasonable questions are raised regarding the quality and nature of a production or the structure of a predictive coding workflow.²⁵⁴ Finally, courts will likely veto a party's attempt to withhold the production of relevant documents in lieu of their identification under *Hickman* and *Sporck*.²⁵⁵

While the foregoing circumstances are foreseeable, they are exceptions and should not swallow the general rule that judgmental seeds deserve work product status. Requiring litigants to routinely disclose such seed information runs contrary to the well-reasoned policies supporting the work product doctrine and its protection of a lawyer's selection of document

VI. CONCLUSION

Predictive coding represents a wave of innovations that are revolutionizing the discovery process.²⁵⁶ When properly implemented under the careful and painstaking supervision of counsel, predictive coding has the potential to simplify some of discovery's troubling complexities. To ensure the continued proliferation of this technology, we assert that courts must take the lead in creating certainty surrounding the predictive coding process. In particular, this requires courts to protect judgmental seeds as work product. While parties remain free to enter into a predictive coding protocol that involves the joint development of a seed set, orders that force the parties into such a process are not generally desirable. It is impossible to argue against transparency and cooperation in the discovery process and we

2013) (rejecting the assertion of work product where the plaintiff failed to establish an actual threat that its lawyers' mental processes would be revealed).

252. See *In re Shell Oil Refinery*, 125 F.R.D. 132, 134 (E.D. La. 1989).

253. See *San Juan*, 859 F.2d at 1015, 1018-19.

254. See *Formfactor, Inc. v. Micro-Probe, Inc.*, No. C-10-03095 PJH (JCS), 2012 U.S. Dist. LEXIS 62233, at *7, n.4 (N.D. Cal. May 3, 2012) (ordering the production of search terms over the plaintiff's work product objection so as to ascertain whether an adequate search was performed for responsive information).

255. See *Mercator Corp. v. United States* (In re Grand Jury Subpoenas Dated March 19, 2002 & August 2, 2002), 318 F.3d 379, 386-87 (2nd Cir. 2002).

256. Gareth Evans & David Grant, *Tools Let Attorneys Follow the Breadcrumbs: Analytics applications can help litigators identify surface patterns that point to the key evidence*, THE NAT'L LAW JOURNAL (Sept. 1, 2014), available at <http://www.ftitechnology.com/doc/Media-Articles/media-national-law-journal-predictive-coding-090114.pdf> (describing the capabilities of advanced analytics in discovery).

2015]

FEDERAL COURTS LAW REVIEW

33

certainly make no such arguments. Nevertheless, parties must retain a zone of privacy in which to prepare for trial and marshal the evidence in discovery to do so. Protecting their seed sets as work product is a significant step forward in accomplishing this objective.