

2015

Flying The Flag

Aaron S Kirschenfeld, *University of North Carolina at Chapel Hill*

Abstract

This paper analyzes the accuracy with which descriptions of subsequent negative treatment are applied by an online citator system that employs a hierarchical controlled vocabulary—Shepard's Citations—as opposed to one that does not—KeyCite. After a contextual review of the citator's history, a framework for assessment is proposed and employed to test the hypothesis that a citator employing a hierarchical controlled vocabulary would produce more accurate descriptions. The study's results suggest that a system making use of a hierarchical controlled vocabulary does apply descriptions of subsequent negative treatment in a marginally more accurate way. A discussion of the citator's place in legal research follows, including the suggestion that legal research instructors and researchers themselves, namely lawyers, should reconceptualize the role citators occupy in the legal research process.

Table of Contents

Flying the Flag: An Introduction	2
Looking Backward: Reviewing the Relevant Literature	5
Methodology	13
Results	21
Looking Forward: A Discussion of the Citator's Place	31
Conclusion	42

Flying the Flag: An Introduction

The system of common law¹ as practiced in the United States depends on the doctrine of *stare decisis*, or in adhering to precedent established in earlier cases. Judges in state appellate courts rely on precedent from their jurisdiction and are sometimes bound by it, but also consider precedent from other states, which can be seen as persuasive authority. And on constitutional issues, state judges must defer to pronouncements of the United States Supreme Court. Judges in federal district court look to precedent from the appellate court from the geographic circuit they practice within when interpreting federal law, and to the Supreme Court as well. Precedent from other circuits may also be persuasive. And state law precedent binds federal courts in diversity of jurisdiction actions.² In this system of overlapping, often confusing authority, it is the lawyer's job to research the law and make arguments based on the nature of judicial precedent and its application to the facts of a particular client's situation. One tool in the lawyer's arsenal for accomplishing this—specifically in determining the validity of points of law a particular case—is the citation index, or citator.

Descriptive information about subsequent negative treatment to a case in the major commercial citators comes in one general form: an introductory

¹ The term “common law” is used in an informal sense, given that so much law in the United States is statutory or regulatory in nature. However, judges do still make law by interpreting legislative actions and other cases, and their actions can still be said to be meaningfully different from countries practicing civil law.

² *Erie Railroad Co. v. Tompkins*, 304 U.S. 64 (1938).

symbol—such as a colored flag or a colored shape—followed by an explanatory phrase. The nature of both of these elements varies greatly among the major citators, namely WestlawNext’s KeyCite and Shepard’s from LexisNexis. Some minor citators, such as Fastcase’s Bad Law Bot, do not include both elements. These products rarely agree, for reasons discussed below, on the particular meaning of these explanations of subsequent negative treatment. But do these differences in characterization of a case’s negative treatment affect how both citators themselves and the lawyers who use them evaluate the case’s significance in common-law precedent?

I propose that when a standardized and hierarchical controlled vocabulary is applied to a legal citation indexing system, evaluation of a case’s subsequent negative treatment is more accurate. Knowing whether this is in fact the case will help inform lawyers, law librarians, and teachers of legal research in addition to legal information vendors. The LexisNexis Shepard’s product is an example of a hierarchical controlled vocabulary of subsequent negative treatment while KeyCite uses language from cases to produce explanatory phrases describing subsequent negative treatment.

After review of a set of cases, described in the methodology below, the results suggest that, absent the action of other variables, there is value in using a hierarchical controlled vocabulary system to describe subsequent negative history. Shepard’s performs better than KeyCite by applying explanatory phrases more

accurately in the limited sense tested. It also performs better in applying introductory symbols more accurately. And so while it cannot be shown that the presence of a hierarchical controlled vocabulary for describing precedent causes more accurate description, the Shepard's system does apply descriptions of precedent more accurately than the KeyCite system, which is a useful finding in its own right. But beyond that, this paper introduces an operationalized definition of "accuracy" and a framework or method for determining if descriptions of negative precedent have been applied accurately.

Based on the findings here, I raise questions about the citator's place in legal research and provide suggestions for teaching students and practitioners about how citators work. Among other things, I suggest that it is the responsibility of legal information professionals to question the validity of claims made by legal information vendors about the accuracy of their citator products. I also offer areas of further research that would be useful to developing a more robust, rigorous literature on citators as legal research products. Finally, I present alternatives to traditional citators that might someday subsume their functions, including citation analysis tools based on visualization or on algorithmic extraction and presentation of subsequent negative treatment.

Looking Backward: Reviewing the Relevant Literature

Instructional legal research materials and legal research courses emphasize the importance of updating one's research.³ Updating involves revisiting selected documents and determining their currency at the time of publication or submission to a court or other lawmaking body. (While citators are used to update statutory, regulatory, and case law materials, the focus of this paper is on description of legal precedent found in case law.) Usually, a lesson on using citators involves a variation on a familiar theme in the law – fear of incompetence. Young lawyers who fail to update a case will find themselves in hot water, and fast. This is no doubt true to a degree—judges do not generally look kindly on counsel whose arguments rest on shaky precedent.⁴ Others have raised the question of whether the failure to update a case is a violation of lawyer's professional ethical obligations.⁵

To date, academic research on citators has fallen into one of two general camps: (1) the history and development of citators and (2) comparisons of

³ See, e.g., J. ARMSTRONG & CHRISTOPHER A. KNOTT, WHERE THE LAW IS: AN INTRODUCTION TO ADVANCED LEGAL RESEARCH 143 (4th ed. 2013); SCOTT CHILDS & SARA SAMPSON, NORTH CAROLINA LEGAL RESEARCH 143 (2nd ed. 2014); MORRIS L. COHEN & KENT C. OLSON, LEGAL RESEARCH IN A NUTSHELL 108 (10th ed. 2010); KENT C. OLSON, PRINCIPLES OF LEGAL RESEARCH 288 (2009); and MARK K. OSBECK, IMPECCABLE RESEARCH: A CONCISE GUIDE TO MASTERING LEGAL RESEARCH SKILLS 51 (2010).

⁴ Lisa McElroy, *Motivating Students to Use Citators: Lessons from Litigation*, 18 PERSPECT. TEACH. LEG. RES. WRIT. 140 (2010).

⁵ Carol M. Bast & Susan W. Harrell, *Ethical Obligations: Performing Adequate Legal Research and Legal Writing*, 29 NOVA LAW REV. 49 (2004).

Shepard's Citations and West Publishing's KeyCite.⁶ While interesting in its own right, a thorough review of the literature on the history and development of citators is unwarranted here, though it will be treated in some depth in this paper's discussion section.⁷ More significant are the attempts—made both by law librarians and by employees of the legal information vendors—to differentiate the market-leading citators at different points in time.⁸

Perhaps the leading, albeit dated, comparison of citators (following the introduction of KeyCite in the mid 1990s) is Taylor's 2000 article in *Law Library Journal*.⁹ He compared Shepard's and KeyCite for three factors: completeness, currency, and accuracy. He defined accuracy as "whether the system correctly

⁶ See, e.g., Patti Ogden, *Mastering the Lawless Science of Our Law: A Story of Legal Citation Indexes*, 85 LAW LIBR. J. 1 (1993); Laura C. Dabney, *Citators: Past, Present, and Future*, 27 LEGAL REF. SERV. Q. 165 (2008); and Lynch, Michael J. *Citators in the Early Twentieth Century—Not Just Shepard's*, 16 LEGAL REF. SERV. Q. 5 (1998). For the institutional history, see FRANK SHEPARD COMPANY, A RECORD OF FIFTY YEARS OF SPECIALIZING IN A FIELD THAT IS OF FIRST IMPORTANCE TO THE BENCH AND BAR OF THE UNITED STATES: AN INSIGHT INTO AN ESTABLISHMENT THAT HAS GROWN FROM SMALL BEGINNINGS TO THE FIRST RANK IN THE LAW PUBLISHING (1923). For the intriguing argument that citation indexing began with law, see Fred R. Shapiro, *Origins of Bibliometrics, Citation Indexing, and Citation Analysis: The Neglected Legal Literature*, 43 J. AM. SOC. INF. SCI. 337 (1992).

⁷ Specifically, the claim that citation indexing was part of a late 19th century classification trend as noted in GEOFFREY C. BOWKER & SUSAN LEIGH STAR, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES 16 (1999).

⁸ And timing did have a lot to do with it. KeyCite was developed in the 1990s, and many of these comparative studies attempted to benchmark it to Shepard's. See Elizabeth M. McKenzie, *New Kid on the Block: KeyCite Compared to Shepard's*, 3 AALL SPECTR. 8 (1998); Elizabeth M. McKenzie, *Comparing KeyCite with Shepard's Online*, 17 LEGAL REF. SERV. Q. 85 (1999). For a more recent comparison, see CAROL A. LEVITT & MARK ROSCH, ARE ALL CITATOR SERVICES CREATED EQUAL? A COMPARISON OF GOOGLE SCHOLAR, FASTCASE, CASEMAKER, LEXISNEXIS, WESTLAWNEXT, AND BLOOMBERG (2012).

⁹ William L. Taylor, *Comparing KeyCite and Shepard's for Completeness, Currency, and Accuracy*, 92 LAW LIBR. J. 127 (2000). For an earlier and less comprehensive effort, see Fred R. Shapiro, *KeyCite and Shepard's—Coverage and Currency of Citations to Recent Cases: A Comparative Study*, 17 Legal Information Alert (1998).

identifies all citing opinions that have a negative effect on the validity or persuasiveness of the cited opinion.” Because he was assessing the citator holistically, this definition suited his purposes, but tends to seem more like the concept of “recall,” perhaps with some consideration of precision involved as well, rather than “accuracy” as operationalized here. Taylor’s study, however, supplies a framework for talking about the types of information found in citators, and his vocabulary is discussed in the context of this paper’s methodology.

That said, in other studies assessing the accuracy of a controlled vocabulary, the term is meant as a study in the accuracy of application—e.g., were all articles about cholesterol-lowering drugs marked with one of the correct subject terms relating to cholesterol-lowering drugs?¹⁰ Again, this is sensible when using “accuracy” to judge an information system as whole, and roughly aligns with Taylor’s recall-heavy definition. As will become clear below, the operational definition for this paper will differ slightly due to the special nature and needs of a legal citation index and what exactly this paper is purporting to test. In general, Taylor and similar older comparisons of citators conclude that each commercial system has its strengths and weaknesses, and that truly effective legal researchers ought to consult both Shepard’s and KeyCite to ensure

¹⁰ Natalie Kupferberg & Bridget McCrate Protus, *Accuracy and completeness of drug information in Wikipedia: an assessment*, 99 J. MED. LIBR. ASSOC. 310 (2011).

correctness. The legal information vendors also had their say in response to Taylor.¹¹

More recent comparisons of citators have looked at the citator's place within the online legal research product context—that is, how the citator behaves in relation to other parts of a unified tool containing hundreds of legal databases. Mart has studied both Shepard's and KeyCite for their relative strengths (or, as it turns out, weaknesses) when accessed through case-law subject headings, called "headnotes."¹² The difference that Mart fixed on was between human-generated headnotes and algorithm-generated headnotes, and she studied how inclusive, in terms of cases cited, each headnote was. This paper does assume that the Shepard's explanatory phrases are both generated or selected and applied by human editors, while the KeyCite explanatory phrases are likely generated or selected and applied algorithmically.¹³ But this is merely an assumption based on a general impression of the data gathered, not on any otherwise verifiable information. And this will not be a focus of the paper, as it presents what is likely a distinction without a difference—the object of this study is the explanatory phrases themselves and whether or not they are accurately applied to the judicial disposition in a case, not how they were generated. Mart's conclusions ultimately

¹¹ Jane W. Morris, *A Response to Taylor's Comparison of Shepard's and KeyCite*, 92 LAW LIBR. J. 143 (2000); Daniel P. Dabney, *Another Response to Taylor's Comparison of KeyCite and Shepard's*, 92 LAW LIBR. J. 381 (2000).

¹² Susan Nevelow Mart, *The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis*, 32 LEGAL REF. SERV. Q. 13 (2013).

¹³ McKenzie, *supra* note 8.

echo those in other, earlier citator comparisons in the narrow sense. Both products are flawed, but flawed in different ways.¹⁴

Some in the profession have questioned whether the companies marketing citators ought to claim that their descriptions of subsequent negative treatment “validate” case law.¹⁵ Updating case law or checking citations is no simple task. Regardless of what the citator claims about the treatment of a case, there will be misapplications or gaps in coverage, and yet there is substantial anecdotal evidence that lawyers accept the companies’ claims about validity more or less unquestioningly.¹⁶ Again, this is not exactly the purpose of the present study, but it is an interesting and challenging perspective from actual lawyers. Validity and accuracy are related concepts, of course, and the role and uses of citators will be discussed following the presentation of this study’s results.

Literature about standardized controlled vocabularies, judicial precedent, and citation indexing in general are in fact most useful to this paper’s task. Studies of law-specific controlled vocabularies have focused on the difficult nature of drawing terms using literary warrant in case-law collections.¹⁷ (And, interestingly, none treat the distinction between literary warrant and end-user

¹⁴ Mart, *supra* note 12.

¹⁵ Alan Wolf & Lynn Wishart, *Shepard’s and Keycite Are Flawed (or Maybe It’s You)*, 75 NEW YORK STATE BAR ASSOC. J. 24 (2003).

¹⁶ *Id.*

¹⁷ Robert C. Berring, *Legal Research and the World of Thinkable Thoughts*, 2 J. APP. PR. PROCESS 305 (2000); Daniel P. Dabney, *The Universe of Thinkable Thoughts: Literary Warrant and West’s Key Number System*, 99 LAW LIBR. J. 229 (2007).

warrant in the law.) On the whole, both the legal information literature and the academic literature concern the generation of controlled vocabularies or the appropriateness of terms chosen, but neither touches on something as specific as what is being studied here. In all, researchers seem to be concerned with studying the systems themselves or their effect on the legal profession or jurisprudence conceptually, rather than the effect that the system might have on particular users in particular instances, or on particular information objects. And of course, these are all worthy and interesting pursuits. Some will even be taken up later in this paper. But there are very few user studies of legal researchers and none, to my knowledge, focused solely on the use of citators.¹⁸ And while this paper is not a user study, it does aim to address this gap in the literature.¹⁹

The literature on the application of standards, specifically ANSI/NISO Z39.19 2005, to collections has been helpful, although ultimately limited. As with some of the studies mentioned earlier, the measure of “accuracy” is more about consistency in application rather than correctness or perceived usefulness. The studies of standards tend to consider issues of whether application of terms is “correct” rather than “useful.” Here, of course, we are concerned with the latter. That said, the standard is also relatively new, meaning that there have been few

¹⁸ But, for an interesting recent survey on lawyer use of information, see ALL-SIS, STUDY OF ATTORNEYS’ LEGAL RESEARCH PRACTICES (2013).

¹⁹ That said, David L. Armond & Shawn G. Nevers, *Practitioners’ Council: Connecting Legal Research Instruction and Current Legal Research Practice, The*, 103 LAW LIBR. J. 575 (2011) provides an interesting, if sobering, take on the difficulty of engaging practicing lawyers in discussion of legal resources.

substantive studies on it since its publication and adoption by information professionals. The standard itself does offer methods for assessing controlled vocabularies, which will be discussed below.

While the literature on Medical Subject Headings (MeSH) is clearly well-developed, it lacks relevance when applied to legal subject matter in one key respect: medical article and citation indexing does not, at least to this point, employ judgments of subsequent negative treatment.²⁰ This paper might be able to shed light on an issue not yet taken up in other professional communities—which way is best to express when articles or hypotheses, like cases or points of law, are challenged, distinguished, or even discredited. And while law is not a science—and American writers in the field have to be given credit for acknowledging such, at least since the late 19th century—its peculiar crucible of precedent developed through trial and appeal can be seen as an analogue to work in more practice-oriented sciences like medicine.

Finally, a look into whether “subsequent negative treatment” has itself been satisfactorily defined is warranted. In short, the literature on this topic is in one sense quite large, but for the most part focuses on the concepts of legal change or of “compliance,” that is, whether cases actually follow other cases, and

²⁰ Mark E. Funk, Carolyn Anne Reid & Leon S. McGoogar, *Indexing consistency in MEDLINE*, 71 BULL. MED. LIBR. ASSOC. (1983).

to what degree.²¹ This literature tends to come from the political science community, and is probably not relevant here. However, compliance is a larger inclusive category for study of the common-law doctrine of *stare decisis*, or judicial precedent. The legal literature is low on systematic studies of *stare decisis*, and usually (and necessarily) focuses on one jurisdiction in depth. Alternatively, one could say that the entire focus of fields of jurisprudential study is the nature and meaning of *stare decisis*. I mention this because in each cited case studied below, an analysis of the case's meaning and what has happened to it, dispositively, after its decision and publication, will be relevant to determining the accuracy of evaluation. Considerations of the notion of the rule of law and its relation to having publicized, well-described precedent deserve special mention, as well.²²

The scholarly and professional literature, then, has not touched directly on the issue at study here, but is suggestive of some appropriate methodological approaches.

²¹ James F. Spriggs and Thomas G. Hansford, *Measuring Legal Change: The Reliability and Validity of Shepard's Citations*, 53 POLIT. RES. Q. 327 (2000). For particular studies, see Scott D. McClurg & Scott A. Comparato, *Rebellious or Just Misunderstood?: Assessing Measures of Lower Court Compliance with U.S. Supreme Court Precedent*, in ANNUAL MEETING OF THE AMERICAN POLITICAL SCIENCE ASSOCIATION, PHILADELPHIA (2003) and James F. Spriggs and Thomas G. Hansford, *The U.S. Supreme Court's Incorporation and Interpretation of Precedent*, 36 LAW SOC. REV. 139 (2002) are two additional relevant examples. See also Sara C. Benesh & Malia Reddick, *Overruled: An Event History Analysis of Lower Court Reaction to Supreme Court Alteration of Precedent*, 64 J. POLIT. 534 (2002).

²² See, e.g., Joseph Raz, *The Rule of Law and its Virtue*, in THE AUTHORITY OF LAW: ESSAYS ON LAW AND MORALITY 210, 217 (1979), for one view on the importance of publicity and predictability in the law.

Methodology

The story of the dataset used in this study is somewhat tortuous and deserves a clear explanation. In the summer of 2013, I worked as a summer research associate with Fastcase, Inc., the online legal research company. As part of my work with the company, I performed a quality assurance analysis on the Bad Law Bot software, testing several features of the automated citator product.²³ My work on Bad Law Bot led to an increased interest in citators, but was ultimately limited in scope. However, Fastcase CEO Ed Walters allowed me to keep a copy of the spreadsheet used in the quality assurance analysis and to use the data in this study.

Before I began my project at Fastcase, the following data had been collected: 44,624 cases in the Fastcase database had been identified as negative by the Bad Law Bot algorithm then in existence. While the algorithm is proprietary, its construction at the time was quite simple. In essence, the algorithm searched for instances of negative subsequent case treatment using “explanatory phrases” from Table 8 of the Bluebook.²⁴ Of those cases, 3,800 were selected for hand-tagging by Fastcase staff.

²³ For a bit more information about how the product works, *see* Introducing Bad Law Bot from Fastcase, *available at* <http://perma.cc/9AQT-U9LV> (last visited Apr 1, 2015).

²⁴ HARVARD LAW REVIEW ASSOCIATION, THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION 434 (19 ed. 2010). The phrases, according to the Bluebook, are “commonly used to indicate prior or subsequent history and weight of authority of judicial decisions.”

Staff members searched for these cases in the WestlawNext database and recorded the flag color associated with the case from the integrated KeyCite product. 3,465 of the cases were assigned Red Flags by KeyCite. The remaining cases were either not flagged by KeyCite or were assigned a Yellow Flag. Of these cases, approximately 32% (108) were not flagged (No Flag) and 68% (227) were identified with a Yellow Flag.

WestlawNext user guides explain that the Yellow Flag “warns that the case has some negative history, but hasn’t been reversed.”²⁵ This study begins with the set of 227 Yellow Flag cases, which, of course, had also been identified as “negative” by the Bad Law Bot algorithm. To study the phrases explaining negative subsequent case law treatment, 116 of the 227 Yellow Flag cases were selected for further analysis. These cases were selected because their KeyCite entries consisted only of what Taylor called “unrelated opinions.”²⁶ Unrelated opinions are those in which “the citing opinion is from a different legal matter but has some effect on the persuasiveness of the cited opinion.” KeyCite calls these unrelated opinions “negative citing references” as opposed to negative direct history, which Taylor calls “related opinions..., where the citing opinion is a later

²⁵ KeyCite on Westlaw Next, (2012), *available at* <http://perma.cc/WS2F-C73Q> (last visited Apr 1, 2015).

²⁶ Taylor, *supra* note 9.

phase of the same legal matter, such as appeals, rehearings en banc, substituted opinions, or supplemental opinions.”²⁷

To summarize, then, cases identified in WestlawNext’s KeyCite having direct negative history were excluded from the data examined here. This was to keep the focus on the descriptions of precedent in later-citing cases. Another way to think of this is to imagine three hypothetical cases: Case A, Case B, and Case C. Case A states a proposition of law. Case B in some way acts on the proposition of law stated in Case A. Case B, then, is an example of a related opinion. Finally, in Case C, the action of Case B on Case A is somehow memorialized. Case C is an unrelated opinion, from a different legal matter, but describing the proposition of law stated in Case A.

These terms are useful ways of talking about the same phenomena. However, since there is no other vocabulary for discussing subsequent negative treatment, a few more terms of art may help in the study that follows. Given that the users of citators are primarily concerned with updating the case they are researching, I have taken to calling this case the Target Case. That is, the Target Case is the case being updated, or “Shepardized.” It is analogous to Case A, but feels somewhat less clinical, and more geared to a researcher’s actual needs. Locating the Citing Case, namely a relevant one, is the result of the use of a citator. We are primarily concerned with a Negative Citing Case, which is

²⁷ *Id.*

analogous to the idea of Case C. Finally, though somewhat less important to this study, is the Acting Case – the case commenting on the Target Case. This is the same as Case B. The relationship between Case B and Case A, the Acting Case and the Target Case, is subsequent negative treatment, and is expressed in a Negative Citing Case, or a Case C. Descriptions of subsequent negative treatment, and their accuracy, are the narrow focus of the study.

For the 116 cases analyzed, 17 individual pieces of data were collected. Some data were collected to insure that the proper case was being analyzed – identification data such as the case’s name, date of decision, and one or more citations associated with the case. Other data fields, as would be expected, record the subsequent negative treatment in the three databases studied (WestlawNext KeyCite, LexisNexis Shepard’s Citations, and Fastcase Bad Law Bot). The 17 data fields are as follows, with brief descriptions of their content:

	FIELD	DESCRIPTION
1	Target Case Name	Name of the Target Case in short form
2	Decision Date	Year, month, and day of decision
3	Target Case Citations	Citations identifying the case from the West National Reporter System or Federal Reports (F. Supp, F., and S.Ct.), from official reports, or, for more recent unreported opinions, the case name, court of decision, and year (eschewing WL or LN citations)
4	Jurisdiction	Federal or State

5	Flag Color on WLN	For all Target Cases analyzed in this study, the Flag Color on WestlawNext was Yellow
6	# of Neg. Treatment on WLN	The number of cases identified as “Negative Citing References” on WestlawNext’s KeyCite
7	WLN Neg. Treatment	The Citing Case identified as “Most Negative” among the Negative Citing References
8	Symbol on Lex	The graphical Shepard’s “signal” for the Target Case
9	# of Neg. Treatment on Lex	After selecting “All Neg” in the Shepard’s report, the number of “Citing Decisions” listed on the report
10	Lex Neg. Treatment	Identified by selecting the most negative Citing Case within the Shepard’s report
11	# of Neg. Treatment on FC	In the Fastcase Authority Check report, the number of cases flagged by the Bad Law Bot algorithm
12	FC Neg. Treatment	The top result in the list of negative Citing Cases identified by Bad Law Bot
13	FC Acting Case	The case identified by Bad Law Bot as the case acting upon the Target Case
14	Classification on WLN	The phrase used in WestlawNext to describe the subsequent negative treatment in the “Most Negative” Citing Case
15	Classification on Lex	The phrase used in the Shepard’s report to describe the subsequent negative treatment in the most negative Citing Case
16	Classification on FC	The phrase describing the relationship between the Target Case and the Acting Case, as identified by Bad Law Bot and generally drawn from Table 8 of the Bluebook
17	Quote from FC BLB	A full quote of the sentence that contains the description of the relationship between the Target Case and Acting Case on Fastcase

TABLE 1

All data have been both collected or checked by me personally and, to the best of my knowledge, is current though at least January 2015 (though some data collection continued into February and March). To gather and check the data, I used my law student access to WestlawNext and LexisNexis (or “Lexis.com”)

from the Lexis Advance platform). And while the data set will be hosted in its entirety in .xlsx format in full on my personal website and will be deposited with an electronic copy of this master's paper,²⁸ I will offer a sample record for a particular case to aid in understanding of why these data might be useful:

	FIELD	DATA
1	Target Case Name	Mulhall v. Advance Sec., Inc.
2	Decision Date	4/22/1994
3	Target Case Citations	19 F3d 586 (note the absence of punctuation and spacing; entering this citation format into the databases tested works well)
4	Jurisdiction	Federal
5	Flag Color on WLN	Yellow
6	# of Neg. Treatment on WLN	14
7	WLN Neg. Treatment	787 F.Supp.2d 961
8	Symbol on Lex	Yellow Triangle
9	# of Neg. Treatment on Lex	8
10	Lex Neg. Treatment	787 F. Supp. 2d 961
11	# of Neg. Treatment on FC	2
12	FC Neg. Treatment	767 F.Supp.2d 1233
13	FC Acting Case	338 F.3d 1304
14	Classification on WLN	Declined to Follow by

²⁸ Available at <http://perma.cc/4LVP-BC8P> (last visited Apr 1, 2015).

15	Classification on Lex	Criticized by
16	Classification on FC	overruled on other grounds
17	Quote from FC BLB	“(citing <i>Mulhall v. Advance Sec., Inc.</i> , 19 F.3d 586, 589 n.8 (11th Cir. 1994), overruled on other grounds, <i>Manders v. Lee</i> , 338 F.3d 1304, 1328 n.52 (11th Cir. 2003)”

TABLE 2

For this Target Case, then, we can reconstruct a picture in textual form from these data. The Target Case was decided in federal court in 1994. KeyCite applied the yellow flag, asserting that the most negative Citing Case “declined to follow” Target Case. Shepard’s applies the Yellow Triangle symbol and asserts that most negative Citing Case “criticized” Target Case. The most negative citing case in both products was the same (787 F. Supp. 2d 961). The Bad Law Bot algorithm, however, identified that Target Case was “overruled on other grounds” by Acting Case, and this fact was recorded in a Citing Case different from the one selected by KeyCite and Shepard’s. Numerous questions possibly flow from these data. The one we are concerned with, however, is: which phrase, if any, is an accurate description of subsequent negative treatment for the Target Case?

To analyze descriptions of subsequent negative treatment, Target Cases having matching Citing Cases were collected, read, and assessed.²⁹ Absent other methods of conducting this type of analysis, this seemed most likely to yield

²⁹ Note that the example from Table 2 meets this criterion and was subject to the analysis described below.

useful, practical results for lawyers and legal information professionals. If the most negative Citing Cases match on at least two of the citators tested, it can be assumed that users should presume these cases present the same subsequent negative treatment. Again, the question here is whether a hierarchical controlled vocabulary like the one employed by Shepard's is a more accurate description of subsequent negative treatment than the uncontrolled systems used by KeyCite and Bad Law Bot.³⁰

³⁰ List of Analysis Definitions Grouped by Shepard's Signals, <https://www.lexisnexis.com/> (last visited Mar 9, 2015). The comparable WestlawNext document for KeyCite does not mention the explanatory phrases, let alone describe or classify them. See Checking Citations in KeyCite, (2010), <http://perma.cc/3C9D-A892> (last visited Apr 1, 2015). Bad Law Bot, of course, does not generate its own explanatory phrases, but rather captures the ones used by courts in Citing Cases.

Results

In the 116 case dataset, 31 Target Cases had identical “most negative” Citing Cases in Shepard’s and KeyCite. That is, 26% of the time the two market-leading citator products agreed about which case best provided subsequent negative treatment. While this may seem surprising, I do not think the result is particularly meaningful as a measure of the reliability of either product or both products. First, “related cases” assigned direct or procedural negative history were excluded—nearly half of the original 227 Yellow Flag cases. I would hypothesize that there is greater alignment among related cases, though this question will not be considered further in this paper. In the discussion section that follows, however, the 74% rate of disagreement between citators will be taken up in its proper context.

What is more troubling to the analysis here is the “most negative” designation as a distinctive, meaningful metric. Only KeyCite explicitly identifies a most negative case for a given Target Case. Shepard’s does not make an explicit determination of which case in its report is most negative. That said, the decision of which case was chosen as most negative was far from arbitrary. The Shepard’s system is a hierarchy from most negative to least negative (Red Hexagon → Q in Orange Box → Yellow Triangle), with each analysis definition falling under a

particular visual signal.³¹ On the Shepard's Report page, there is a box summarizing this ordering near the top of the page. And so the most negative case was selected by locating the first result in the most negative set of citing cases. Still, I do not expect a practitioner looks only at the case in a citator listed as, or deemed to be, most negative for the Target Case she is researching. As noted earlier, a user study of attorneys and other expert legal researchers would be needed to learn *how* these people use citators. We do know, of course, that researchers *do* use them.

These challenges aside, the 31 cases with identical most negative Citing Cases can tell us quite a bit about the language used to describe subsequent negative treatment in Shepard's and KeyCite. First of all, 12 of the 31 cases, or 39%, used identical language to describe subsequent negative treatment. Interestingly, these 12 instances all involved the same explanatory phrase: "distinguished by." Two additional cases were assigned near-identical language: "disapproved by" (Shepard's) and "disapproved of by" (KeyCite). That leaves 17 cases in which different language was used to describe subsequent negative treatment for a Target Case. This is the nugget of cases remaining in which a meaningful analysis of a citator's accuracy in application of subsequent negative treatment phrases.

³¹ Analysis Definitions, *supra* note 26.

Four of the Shepard's phrases are present in this final collection of cases:

"criticized by," "distinguished by," "overruled in part by," and "questioned by."

Shepard's defines these phrases as such:

Criticized by – The citing opinion disagrees with the reasoning/result of the case you are *Shepardizing*TM, although the citing court may not have the authority to materially affect its precedential value.

Distinguished by – The citing case differs from the case you are *Shepardizing*TM, either involving dissimilar facts or requiring a different application of the law.

Overruled in part by – One or more parts of the decision you are *Shepardizing*TM have been expressly nullified by the subsequent decision from the same court, thus casting some doubt on the precedential value of the case you are *Shepardizing*TM.

Questioned by – The citing opinion questions the continuing validity or precedential value of the case you are *Shepardizing*TM because of intervening circumstances, including judicial or legislative overruling.³²

The "criticized by" and "distinguished by" phrases should, according to Shepard's, be assigned the Yellow Triangle signal, meaning "caution."³³

"Overruled in part by" should be assigned the Red Hexagon signal, meaning

"warning."³⁴ "Questioned by" should be assigned the Q in Orange Box signal,

meaning "questioned."³⁵ "Criticized by," "distinguished by," and "questioned by"

are considered to be "common analysis phrases."³⁶ So, it can be inferred that, as

³² *Id.*

³³ Shepard's Signal Indicators and Analysis Phrases at lexis.com, (2012), <http://perma.cc/FR7G-FSXA> (last visited Apr 1, 2015). "Caution: Possible negative treatment indicated" is the full definition of the graphical signal.

³⁴ *Id.* "Warning: Negative treatment is indicated" is the full definition of the graphical signal.

³⁵ *Id.* "Questioned: Validity questioned by citing refs." is the full definition of the graphical signal.

³⁶ *Id.*

common phrases, these would likely be encountered by a researcher while he conducts legal research.

Seven KeyCite analytical phrases are used in describing the subsequent negative treatment in the final 17 cases studied: “called into doubt by,” “criticized by,” “declined to extend by,” “declined to follow by,” “disagreed with by,” “not followed as dicta,” and “rejected by.” As mentioned earlier, WestlawNext does not provide definitions of these phrases. The phrases, in the large dataset, appear in combination with either a Red Flag or Yellow Flag, the only two symbols used by KeyCite on WestlawNext. These facts are presented as proof of two propositions. First, that KeyCite’s system of applying analytical phrases to cases is clearly not an example of a hierarchical controlled vocabulary. Second, it will be more difficult to assess whether the phrases have been applied to cases accurately because the definitions of the phrases are necessarily somewhat subjective.³⁷ Words such as “disagreed”³⁸ and “doubt”³⁹ seem to suggest common definitions, while “dicta”⁴⁰ is a legal term of art. “Followed,”⁴¹ “extend,”⁴² and

³⁷ I consulted two dictionaries in crafting the definitions that follow in this section: THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE, (2000) and BLACK’S LAW DICTIONARY, (Bryan A. Garner ed., 9th ed. 2009).

³⁸ As defined by BLACK’S LAW DICTIONARY, *Id.*, “disagreement” means “[a] difference of opinion; a lack of agreement,” which is indeed close to the common use of the term.

³⁹ “Reasonable doubt” is the only technical use of “doubt” in law I am familiar with; of course, it has nothing to do a judge’s determination, but rather a jury’s.

⁴⁰ “Dicta,” is undoubtedly the colloquial plural form of “obiter dictum,” which means “[a] judicial comment made while delivering a judicial opinion, but one that is unnecessary to the decision in the case and therefore not precedential (although it may be considered persuasive).” BLACK’S LAW DICTIONARY, *Id.*

“rejected”⁴³ seem to be hybrids of words used commonly and legal terms of art, and are therefore the most difficult to assess. I have defined the seven KeyCite analytical phrases as follows:

Called into doubt by – The citing case expresses uncertainty about the precedential value of the target case.

Criticized by – The citing case disagrees with the reasoning or result of the target case, although the citing court may not have the authority to materially affect its precedential value.⁴⁴

Declined to extend by – The judge in the citing case has chosen to not increase the influence of the target case.

Declined to follow by – The judge in the citing case has chosen not to comply, conform with, or accept the target case as authoritative.

Disagreed with by – The citing case expresses a difference of opinion or lack of agreement with the target case.

Not followed as dicta – The court in the citing case will not accept the target case as authoritative because of statements in the target case are considered to be unnecessary to the decision in the case and therefore not precedential.

Rejected by – The citing case declines to make use of reasoning from the target case.

With these definitions in place, it is finally possible to make a meaningful assessment of the accuracy of the application of descriptions of subsequent negative treatment. As described in the methodology, determining which system is more accurate is also necessarily subjective—a matter of reading the relevant portions of each case and determining if one explanatory phrase (or both, or

⁴¹ “Follow”, as defined by BLACK’S LAW DICTIONARY, *Id.*, means “[t]o conform to or comply with; to accept as authority.”

⁴² The closest definition of “extend” provided by THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE, *supra* note 34 is “to expand the influence of.”

⁴³ Rejection has technical legal meaning in the law of contracts, but I am not aware of a clear technical decision in terms of precedent. The common meaning from THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE, *Id.*, is “refuse[d] ... to make use of.”

⁴⁴ This definition is in effect identical to the Shepard’s definition.

none), more clearly described the nature of the judicial determination in the citing case as it relates to the target case. While this seems difficult in theory, there were relatively few challenging decisions to make when reading and classifying a statement of subsequent negative treatment.

Target Case Shortname	Decision Date	Target Case Citations	Jurisdiction	Flag Color	Symbol on Lex	Classification on WLN	Classification on Lex
Gregory v. United States	1966-07-28	369 F2d 185, 125 U	Federal	Yellow	Q in Orange Box	Criticized by	Questioned by
McCurdy v. Steele	1973-01-02	353 FSupp 629	Federal	Yellow	Red Hexagon	Declined to Follow by	Questioned by
United States v. DeCoster	1973-10-04	159 USAppDC 326,	Federal	Yellow	Red Hexagon	Called into Doubt by	Questioned by
Brandenburger v. Thompson	1974-03-25	494 F2d 885	Federal	Yellow	Q in Orange Box	Disagreed With by	Questioned by
Finley v. United States	1975-12-02	404 FSupp 200	Federal	Yellow	Q in Orange Box	Disagreed With by	Questioned by
United States v. Dorfman	1982-06-01	542 FSupp 345	Federal	Yellow	Red Hexagon	Rejected by	Criticized by
Melson v. Kroger Co.	1983-11-04	578 FSupp 691	Federal	Yellow	Q in Orange Box	Disagreed With by	Questioned by
People v. Johnson	1984-05-15	204 CalRptr 563, 15	State	Yellow	Yellow Triangle	Not Followed as Dicta	Criticized by
A & C Properties, In re	1986-05-13	784 F2d 1377	Federal	Yellow	Red Hexagon	Declined to Follow by	Questioned by
Barber v. National Bank of Alaska	1991-07-26	815 P2d 857	State	Yellow	Q in Orange Box	Called into Doubt by	Questioned by
Stafford v. Purofied Down Products Corp.	1992-09-02	801 FSupp 130	Federal	Yellow	Red Hexagon	Declined to Follow by	Questioned by
Mulhall v. Advance Sec., Inc.	1994-04-22	19 F3d 586	Federal	Yellow	Yellow Triangle	Declined to Follow by	Criticized by
Doctor's Associates, Inc. v. Distajo	1995-09-28	66 F3d 438	Federal	Yellow	Red Hexagon	Disagreed With by	Criticized by
Atchison, Topeka and Santa Fe Ry. Co. v. Brown & Bryant, Inc.	1998-10-14	159 F3d 358	Federal	Yellow	Red Hexagon	Not Followed as Dicta	Criticized by
Reed v. Reed	2001-06-19	48 SW3d 634	State	No Flag	Red Hexagon	Declined to Follow by	Overruled in part by
Security Ins. Co. v. Old Dominion Freight Line	2003-10-22	314 FSupp2d 201	Federal	Yellow	Yellow Triangle	Declined to Extend by	Distinguished by
Pequignot v. Solo Cup Co.	2009-03-27	640 FSupp2d 714	Federal	Yellow	Red Hexagon	Declined to Follow by	Criticized by

TABLE 3

Of the 17 target cases (shown in Table 3) with matching most negative citing cases, four had overlapping descriptions of subsequent negative treatment. This occurred when the explanatory phrase was equally accurate in describing the subsequent negative treatment. For example, three target cases labeled as “criticized by” in Shepard’s, the KeyCite phrases “rejected by,” “declined to follow by,” and “disagreed with by,” as defined above, were for all purposes identical for the given cases. *United States v. Dorfman*, 542 F.Supp. 345, a target case, is described as “criticized by” in Shepard’s and as “rejected by” in KeyCite. The citing case comments:

The Court rejects the suggestion in *United States v. Dorfman*, 542 F.Supp. 345, 373 (N.D.Ill.), *aff'd*, sub nom. *United States v. Williams*, 737 F.2d 594 (7th Cir.1984), relied on by the government, that “[t]he test ... is not the state of the mind of ‘the government,’ but the state of mind of the affiant.” This Court treats the affiant as a representative of the United States government as to the activities of any of its agents, whether or not the affiant has been made aware of those activities.⁴⁵

The citing case, from a federal district court in the First Circuit, chooses not to apply a particular doctrinal rule used in the target case, from a federal district court in the Seventh Circuit. It fits the definition of “criticized by” and the definition of “rejected by,” presented above. Therefore, the description of the subsequent negative treatment has been accurately applied.

In the set of cases studied, 10 of 17, or 58.9%, of descriptions of subsequent negative treatment in Shepard’s were applied accurately. The KeyCite system applied descriptions of subsequent negative treatment accurately in 8 of 17, or 47.1% of cases. Neither citator applied descriptions of subsequent negative treatment accurately in 3 of the 17 cases, or 17.6% of the time for this set. While the sample size studied is small and no conclusions can be reached about the ultimate comparative quality of this aspect of either citator, these data do suggest that each system describes the actions taken in identical citing cases with differing degrees of consistency.

⁴⁵ *Application for Interception of Wire Commc’ns*, 2 F.Supp.2d 177, 179 (D. Mass. 1998)

The clear definitions in the hierarchical Shepard's system allowed for further comparison of the data from the comparative set described above to a Shepard's-only set. A systematic sample from the remaining 85 cases in the set of 116 described above was taken. Alternating 4th or 5th cases were selected from the table sorted chronologically to help mitigate any bias from the date of a case. In total, 19 cases were selected from the 85 and subjected to the same analysis as the set of 17 above. In 14 of the 19 cases in the Shepard's only set, subsequent negative treatment from a citing case was accurately described by the analytical phrase. That is, 73.6% of the time, Shepard's, according to its own standards, described the subsequent negative treatment accurately. A similar test of the KeyCite system for the same set of cases did not seem appropriate, given that no clear definitions actually exist for the explanatory phrases used in the citator.

These results, it can be argued, offer no definitive conclusions of any kind about how accurately a given citator describes subsequent negative treatment in case law. This is an argument I ultimately accept – a truly systematic study of the topic is likely impossible given the number of variables that could affect its results. For instance, factors such as chronology, the presence of direct procedural history (Taylor's "related opinions"), consistency in application of the phrases by LexisNexis or Thomson Reuters, and the case identified as "most negative" could each present different findings. Each variable might be controlled for on its own, but controlling for all at once seems exceedingly difficult. However, I do believe

that by articulating a methodology for studying citators and offering a detailed picture of interpreting and operationalizing accuracy of application of subsequent negative treatment is valuable in its own right.

There are, of course, other indicia of accuracy in application descriptions of subsequent negative treatment: namely, the consistency with which cases are tagged with a citator's iconography. In the set of 116 cases, KeyCite assigns the Yellow Flag to 18 different explanatory phrases. The 18 explanatory phrases are not evenly distributed—two phrases, “declined to follow by” and “distinguished by” account for 51 of the 116 cases. In the larger set of 227 Yellow Flag cases (containing related opinions and unrelated opinions), KeyCite assigns 57 different explanatory phrases.

Shepard's iconography was only studied in the set of 116 unrelated opinions. Seven of the 116 assigned no subsequent negative treatment, and are excluded from this part of the analysis. Of the remaining 109 cases, 37 were assigned the Yellow Triangle, 20 were assigned the Q in Orange Box, and 52 were assigned the Red Hexagon. Of the 37 cases assigned the Yellow Triangle, four explanatory phrases were applied, with two phrases, “criticized by” and “distinguished by” accounting for 35 of the 37 cases. Of the 20 cases assigned the Q in Orange Box, all 20 were given the “questioned by” analytical phrase. Finally, of the 52 Red Hexagon cases, Shepard's assigns 15 analytical phrases. Some of the Red Hexagon cases included related opinions in the Shepard's report

that were not captured by KeyCite, so for a meaningful comparison, it makes sense to look at the Yellow Triangle and Q in Orange Box cases. For the 57 such cases, only five phrases are applied, and three of the five phrases account for 55 of the 57, or 96.5%. Looking at this metric, the value of a controlled visual vocabulary is apparent when considering the accuracy of application of descriptions, albeit visual descriptions, of subsequent negative treatment. It seems fair to conclude that a user encountering the Shepard's Yellow Triangle and Q in Orange Box iconography for a target case would have a clear understanding of the type of subsequent negative treatment contained in the citing case. The user encountering a Yellow Flag on KeyCite, on a superficial level, anyway, would not likely know what meaning to make of the symbol.

Looking Forward: A Discussion of the Citator's Place

Questions about the use of online citators—particularly their presentation of graphical symbols as unite of meaning—are not new among teachers of legal research.⁴⁶ It seems, from a survey of the literature, that qualifications about the adequacy of these symbols form an integral part of lessons and texts on updating cases and statutes.⁴⁷ Even the most basic legal research text cautions that a citator's "signals and editorial signposts are not authoritative statements of the law..." and that there is "no substitute for reading a citing document and determining for yourself its scope and effect."⁴⁸ These same texts, however, note that citators are "helpful as a starting place"⁴⁹ in determining a case's continuing validity, or should be considered "invaluable resources ... [ensuring] that cases are still good law."⁵⁰

Marketing materials from Shepard's and KeyCite paint a picture of the infallibility of each service's respective citator. One such document produced by LexisNexis asks: "Why do so many judges and lawyers rely on Shepard's Citations Service, exclusively from LexisNexis®, when it comes to validating

⁴⁶ Kent C. Olson, *Waiving a Red Flag: Teaching Counterintuitiveness in Citator Use*, 9 PERSPECT. TEACH. LEG. RES. WRIT. 58 (2001).

⁴⁷ See, e.g., ARMSTRONG AND KNOTT, *supra* note 3 at 135.

⁴⁸ COHEN AND OLSON, *supra* note 3 at 115.

⁴⁹ ARMSTRONG AND KNOTT, *supra* note 3 at 135.

⁵⁰ COHEN AND OLSON, *supra* note 3 at 109.

research?” The answer? “It’s a matter of trust.”⁵¹ Likewise, KeyCite materials tout “West’s 125-year tradition of editorial excellence” and the company’s “leading-edge technological expertise” in suggesting that users can “instantly verify whether a case ... is good law.”⁵² Of course, as Wolf and Wishart pointed out, these statements should be taken for what they are—largely meaningless, if not misleading, marketing attempts to differentiate products.⁵³ It can be tempting to believe these kinds of statements based on the credibility of the source,⁵⁴ not to mention the cost of the product.⁵⁵ But savvy legal researchers know, and legal research instructors profess—there is no sidestepping the need to read a case and use professional judgment to determine the precedential effect of later-citing cases. Still, citators are seen as essential components of legal research databases.⁵⁶

In light of what the results of this study suggest, there is even greater reason to doubt the adequacy of a citator’s determination. More troubling, perhaps, is the implicit critique of the citator as a “good starting place” or an

⁵¹ Citation Services: Deep Analysis and Unique Product Details Lead Researchers to Shepard’s Citations Service at lexis.com, (2012), <http://perma.cc/G4JV-5G2T> (last visited Apr 1, 2015).

⁵² KeyCite on Westlaw Next, *supra* note 25.

⁵³ Wolf and Wishart, *supra* note 15 at 27.

⁵⁴ Bob Berring calls this the “Tinkerbell” effect. He writes: “In the Walt Disney animated feature ‘Peter Pan,’ Tinkerbell was a fairy. She only existed if children believed in her existence. This character, viewed by the author at an impressionable age, stands for the classic bootstrapping of authoritativeness.” Robert C. Berring, *Chaos, Cyberspace and Tradition: Legal Information Transmogrified*, 12 BERKELEY TECHNOL. LAW J. 189, 193, fn. 17 (1997). The term is later used by Mary Whisner, *see* Mary Whisner, *Bouvier’s, Black’s, and Tinkerbell*, 92 LAW LIBR. J. 99 (2000). The extension of the idea to citators specifically is my choice alone.

⁵⁵ Hilke Plassmann et al., *Marketing Actions Can Modulate Neural Representations of Experienced Pleasantness*, 105 PROC. NATL. ACAD. SCI. U. S. A. 1050 (2008).

⁵⁶ Greg Lambert, *Casemaker Unique Among Legal Research Providers*, 89 MICHIGAN BAR J. 54, 56 (2010).

“invaluable resource.” Indeed, if the graphical signals, let alone the explanatory phrases, cannot be trusted to be applied accurately—at best, approximately 75% of the time, or at worst, less than 50% of the time—it is difficult to see how descriptions of subsequent negative treatment would reliably set the researcher down the correct path. It could be argued that the presence of *any* negative symbol or phrase would be enough to alert the user that “here there be dragons”⁵⁷ and therefore, a concurrent need to read and understand the “flagged” case before citing it.⁵⁸ This argument is easily countered, however, by the fact that not reading a case before citing it is almost certainly a violation of Rule 1.1 of the ABA *Model Rules of Professional Conduct*, requiring a lawyer to provide “competent representation to a client[,]” a duty defined as consisting of, among other things, “thoroughness and preparation.”⁵⁹ So, if a tap on the shoulder from a citator with no further intelligible explanation means “read the case,” that is something a competent researcher should already be planning to do, anyway.

Citators, then, must add some value to the research process, not because they are thought of as “invaluable,” but rather because they are precisely the opposite—important components of rather costly legal research databases. To

⁵⁷ This phrase comes to mind because of an article on a slightly different topic in legal research, see Peggy Roebuck Jarrett & Mary Whisner, “*Here There Be Dragons*”: *How to Do Research in an Area You Know Nothing About*, 6 PERSPECT. TEACH. LEG. RES. WRIT. 74 (1998).

⁵⁸ More appropriately, the relevant portion of the case. See Robert C. Berring, *Unprecedented Precedent*, 5 GREEN BAG 2D 245.

⁵⁹ AMERICAN BAR ASSOCIATION, MODEL RULES OF PROFESSIONAL CONDUCT 11 (2011). This is assumed from the work done in Bast and Harrell, *supra* note 5. Likewise, there are no cases specifically on point in Kristina L. Niedringhaus, *Ethics Considerations Related to Legal Research Practices: A Selective Annotated Bibliography*, 31 LEGAL REF. SERV. Q. 104 (2012).

explore this problem, our discussion will now diverge onto two paths: first, how citators as currently constituted can be used beneficially (or, one could say, profitably) by researchers, and second, how citators might fit in a legal research environment with a multiplicity of tools for determining the meaning of subsequent negative treatment. Finally, the paths will converge when we will consider areas of further research that could be useful for developing a more extensive literature for evaluating the quality of citators.

The roots of legal citation indexes in the United States stretch to the early 19th century, before they were known as citators.⁶⁰ Ogden credits pre-eminent jurist and scholar Joseph Story with the idea of applying an “explanatory letter” such as D for doubted or O for overruled to a table of cases included in a reporter volume.⁶¹ The earliest popular citation index existed for much the same purpose as the contemporary citators—efficient identification of a case’s subsequent negative treatment. The title of the volumes produced over a lifetime by Simon Greenleaf, *Greenleaf’s Overruled Cases*, explains itself.⁶² Other early citation indexes were also compiled by individuals, almost always at first for their creator’s personal use.⁶³ But the coming of the twentieth century brought a growth in the country’s urban population and commercial activity, which led to an increase in the number of courts (and reorganization of those courts), and

⁶⁰ Ogden, *supra* note 6, is the canonical history.

⁶¹ *Id.* at 4.

⁶² *Id.* at 6. Greenleaf was a close friend and colleague of Justice Story.

⁶³ *Id.* at 23.

eventually to a consequent explosion of the number of published precedential legal documents, such as appellate opinions.⁶⁴ In addition to figuring out how to organize these cases, interested parties in the profession, namely lawyers, law librarians, and the growing national publishers needed to find a way to make sense of this ever-expanding stuff of law.⁶⁵

Until the digital era in legal publishing, and through its first generation, Shepard's was the only national citator capable of accomplishing this comprehensive task.⁶⁶ While not originally to be a facsimile of the print, Shepard's online soon took on many of the functions, and idiosyncrasies, of the print volumes.⁶⁷ But its purpose remained the same as always—identifying precedential procedural and citing history in a comprehensive and current fashion. What digital publication of legal materials added, however, was the ability to generate a so-called “table of authority” for a given document, such as a case. That is, it was suddenly possible to extract information about which cases cited other cases and how often a given case was cited, among other things. KeyCite includes graphical representations of a case's direct procedural history while both services include extensive lists of citing cases, regardless of whether the

⁶⁴ For an interesting, summary discussion of this problem, see Richard A. Danner, *The ABA, the AALL, the AALS, and the “Duplication of Legal Publications,”* 104 LAW LIBR. J. 485, 488 (2012).

⁶⁵ This is an analogy to one of the main theses from BOWKER AND STAR, *supra* note 7 at 16. In that work, the context is epidemiological classification.

⁶⁶ Dabney, *supra* note 6, contains an excellent, detailed history of the development of KeyCite in addition to a history of Shepard's in the late 20th century.

⁶⁷ *Id.* at 170.

treatments within are approving, negative, or neutral. The content of each system's list of citing cases for a given target case differs to whatever extent it does because of differences between the larger research systems in their coverage of unpublished opinions,⁶⁸ an interesting issue, but not one I will take up any further here.

Given the change in the nature of citation analysis possible in the digital realm and the shaky performances of both systems in accurately describing substantive negative history, I wonder if the era of the citator *qua* citator has already passed. Instead of conceptualizing these indexes of citations as tools reliably identifying both the existence and nature of subsequent negative history, and therefore essential steps in "updating" legal research, it seems more appropriate to relieve these tools of the moniker they have acquired and the related heavy, if not impossible, burden associated with it. A legal citation index cannot, on its own, say if a target case is bad law. It certainly cannot say if a case is good law. And, as discussed earlier, any legal researcher relying on it to do so is almost certainly incompetent, in the professional sense, anyway. This raises the question, hinted at earlier, of what legal citation indexes are still good for. To put it more circumspectly, what is the tool-known-as-citator's place in legal research now and in the future?

⁶⁸ Mart, *supra* note 12 at 51, fn. 122.

Currently, legal citation indexes can gather a wide range of subsequent citing authority. KeyCite is more explicit about displaying these sources—including statutes citing the target case as well as secondary sources like law reviews, treatises, or looseleaf series. The same variety of sources is also available in the Shepard's report, though the interface makes them a bit more difficult to find. Both products include a table of authorities cited within the target case. KeyCite also includes a unique feature, measuring the depth of treatment in citing court documents. This measure is likely calculated based on how often the target case is cited or mentioned in the citing document, and how many words separate the citation from the next different citation. Thomson Reuters does not disclose how its depth-of-treatment algorithm functions, naturally, so the prior statement represents my best guess. Neither system provides other types of symbols or explanatory phrases to describe other later-citing authorities.

Legal research instructors need to emphasize these features of citation analysis currently available in both major citators while explaining potential problems with the descriptions of subsequent negative treatment in the later-citing cases when provided. Also, the notion of a citator as a tool for updating research must be expanded—the citation index, when integrated into the larger research platform, can serve a vital function throughout the research process. Examining the dates of later-citing cases (are there any that are current?) or the jurisdiction or court level of later-citing cases (will the decisions be binding precedent?) are two

basic examples. Rethinking the citator as a tool for analyzing the influence of a case based on later citations in a variety of sources is needed, rather than calling it a final box to tick to ensure the validity of a case's proposition as good law.

Lawyers and legal information professionals also must continue to critique and qualify the marketing or training materials created by Thomson Reuters and LexisNexis about their respective citator products. Much good work, as we have seen, has already been done in this and related areas, and should continue. Law students and practicing attorneys need to be made aware of the implicit overruling problem,⁶⁹ the inclusion or exclusion of unpublished opinions,⁷⁰ the misleading nature of certain graphical representations of precedent,⁷¹ and of course, the inaccuracy in application of descriptions of subsequent negative treatment such as symbols and explanatory phrases. Additionally, it cannot be left to the information vendors to teach how these products work, especially if citators are being presented as tools for instantly confirming or disconfirming the validity of a point of law in a given target case.

Beyond teaching what citators can and should do well presently, it is also interesting to look to a future where new tools will assume some of the traditional citator's functions, or will push Thomson Reuters and LexisNexis to reimagine their current products. Start-up companies are producing new-wave citation

⁶⁹ Wolf and Wishart, *supra* note 15.

⁷⁰ Mart, *supra* note 12.

⁷¹ Olson, *supra* note 46, which argues that the red hexagon should be thought to mean "go" rather than "stop."

analysis tools based on algorithmic extraction of citation data, presenting information about case citation networks visually.⁷² Researchers no longer need to rely solely on the textual representation of citation data when analyzing precedent.⁷³ Also intriguing is the possible application of machine learning algorithms or artificial intelligence to legal systems. And while two recent articles on the topic do not mention the application to citators specifically, the authors do stress that teasing out the relationship between cases is a potential application of this technology.⁷⁴ While a fully automated citator with accurate descriptions of subsequent negative treatment may be far in the offing, more options for conducting citation analysis outside of the bounds of traditional citators have already arrived.

A product of this paper, I hope, is more commentary on the nature of citators generally, but also on descriptions of subsequent negative treatment specifically. The analysis in this paper does not satisfactorily answer several questions it raises. For instance, is the presence of a hierarchical controlled vocabulary causal in any way of greater accuracy in the application of descriptions of subsequent negative treatment? Could the observed results instead

⁷² The best example of this is Ravel. *See* About Ravel Law, <http://perma.cc/BTF5-884H> (last visited Apr 1, 2015). Fastcase also makes use of visual representations of precedent in its Authority Check tool.

⁷³ SUSAN AZYNDAR, KATRINA LEE & INGRID MATTSO, A NEW ERA: INTEGRATING TODAY'S "NEXT GEN" RESEARCH TOOLS RAVEL AND CASETEXT IN THE LAW SCHOOL CLASSROOM 21 (2015), *available at* <http://perma.cc/EEW3-5AWC> (last visited Apr 1, 2015).

⁷⁴ *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 LAW J. 909 (2013) and Harry Surden, *Machine Learning and Law*, 646 WASHINGT. LAW REV. 87 (2014).

be due to differences in the consistency of application of descriptions by the individual services? Or, would the absence of any definitions of explanatory phrases in KeyCite make the idea of consistent application impossible? On the contrary, might the KeyCite terms actually represent narrower terms of broader Shepard's analogues, and form a kind of implicit hierarchical controlled vocabulary? Or do the phrases exist simply as snippets, attempting to form a distinction among cases without actually categorizing the subsequent negative treatment contained within?

Other areas for research that I see as particularly helpful would be further exploration of what is meant by the idea of comprehensiveness in a citator. To that end, an estimate of how many cases present in the case law databases have ever been cited by another case, and, if a case has been cited, the mean number of times it has been cited would be interesting. Of this number, how many of those citing references would be negative? Knowing how to best ask and answer these types of quantitative questions would likely be of aid to researchers looking to benchmark observed qualitative phenomena. Would the necessarily subjective descriptions of negative subsequent precedent make sense to duplicate if one were to start building a citator today or, as I suspect, would the making of meaning take some other, more inferential path? That is, what would the modern-day Story emphasize as important to the modern-day Greenleaf?

More prosaically, it should be noted that Shepard's has a different iconography and user interface on the new Lexis Advance platform, which was not tested here. (KeyCite on WestlawNext was chosen because Thomson Reuters was beginning to phase out Westlaw "Classic" at the time this research began, and has since completed the transition before publication.) Tests on this system, which seems on cursory examination to incorporate a depth-of-treatment element, could yield different results. Likewise, would an examination of, say, the three most negative cases lead to greater correspondence between the citing cases presented by each service?

Finally, allow me to offer a direct invitation to both LexisNexis and Thomson Reuters: Please, critique my work. Justify the decision to continue to include graphic and verbal descriptions of subsequent negative treatment and the value this particular element of the citator adds to the legal research platforms on the whole. Does the cost associated with producing these descriptions make sense given the inaccuracy in application found here? If so, why should lawyers and legal professionals of all stripes—in large law firms, government, or legal education—continue to pay the ever-rising fees attached to your services?

Conclusion

Citators are expected, among other things, to provide accurate descriptions of subsequent negative treatment, and have been expected to do so since coming into being in the United States. This study sought to answer the question of whether a system, such as Shepard's Citations, employing a hierarchical controlled vocabulary, would apply descriptions of subsequent negative treatment more accurately than a system, such as KeyCite, lacking a hierarchical controlled vocabulary. To reach a meaningful conclusion, a large set of cases with later-citing history was narrowed until a fair comparison could be made. The results of this comparison suggest that the presence of a hierarchical controlled vocabulary produces more accurate descriptions of subsequent negative treatment.

To this point, the relevant scholarly literature had been largely silent on descriptions of negative precedent in law. Some comparisons of citators for other factors have been done, and considerable writing by legal information professionals has critiqued the overall trustworthiness of citators. This paper addresses the gap and provides a framework for assessing particular elements of citators. But scholarship in library and information science should be useful when it can be, and this paper provides suggestions on how to best use citators given the results of the study. Citators remain valuable tools as citation indexes, gathering all later-citing history from a variety of primary and secondary sources. But in

providing descriptions of subsequent negative treatment, citators seem tied to an antiquated historical model originating in the early 19th century and may presently be failing at the task deemed to be central to their existence.