University of Washington School of Law

UW Law Digital Commons

2013

My Lawyer Has a First Name, It's G-O-O-G-L-E: Improving Online Access to Case Law Through Court-Provided Metadata

Andrew Martineau

Recommended Citation

Andrew Martineau, *My Lawyer Has a First Name, It's G-O-O-G-L-E: Improving Online Access to Case Law Through Court-Provided Metadata* (2013), https://digitalcommons.law.uw.edu/law-lib_borgeson/73

My Lawyer Has a First Name, It's G-O-O-G-L-E: Improving Online Access to Case Law Through Court-Provided Metadata

Andrew Martineau

Submitted to
Professor Penny A. Hazelton
to fulfill course requirements for Current Issues in Law Librarianship, LIS 595,
and to fulfill the graduation requirement of the
Culminating Experience Project for MLIS
University of Washington Information School
Seattle, Washington
May 13, 2013

My Lawyer Has a First Name, It's G-O-O-G-L-E: Improving Online Access to Case Law Through Court-Provided Metadata

In the latter part of the twentieth century and the early twenty-first, legal publishing underwent a revolution. We have transitioned from a world of numerous, small publishers of legal information in print to a world of few, giant publishers whose focus is shifting towards the digital. This revolution has technological, economic, and social components. The technological revolution resulted largely from advances in information technology, and the development of the Internet especially. Now, lawyers can complete most research tasks exclusively online, through systems like Westlaw, LexisNexis, and Bloomberg Law. The development of these massive legal information databases coincided with fundamental structural changes in the legal publishing business. Over the past fifty years, the number of publishers of legal materials has shrunk to basically just three.[1] On the other hand, we have seen explosive growth in the availability of free legal information online, through both private sector services (e.g. Google Scholar) and government websites (e.g. regulations.gov). Underlying these phenomena is the general societal trend of citizens depending on the Internet as their primary source for information.

The handling of court records has undergone a revolution as well, similarly fueled by advances in information technology.[2] Over the past thirty years, many courts have moved from all-paper systems to electronic case management systems, which keep track of filings, decisions, court calendars, and just about any administrative task.[3] Many courts have also begun publishing documents in publically accessible locations on the Internet, some even opting to forgo official print publication all together.[4] However, the progress and success of these efforts towards the online publication of authentic court documents vary from court to court.[5]

This shift toward courts publishing their own documents in electronic format raises a number of issues. Much has been written, for example, on the topic of medium-neutral citation.[6] Medium-neutral citation ensures that a litigant can cite to a case without having to consult the version published by a commercial vendor like West.[7] Another issue, which is the focus of this article, is one of access. Does the government have a duty to provide access to public court documents? If so, what might this duty entail?

---

[1] *See generally* Kendall F. Svengalis, *Legal Information Buyer's Guide and Reference Manual*, 3 (2012).
[2] *See generally* Gregory M. Silverman, *Rise of the Machines: Justice Information Systems and the Question of Public Access to Court Records Over the Internet*, 79 WASH. L REV. 175, 178 (2004).
[3] *Id.* at 176, 177.
[4] *See* ASS'N OF LAW LIBRARIES GOV'T RELATIONS COMM., STATE-BY-STATE REPORT ON AUTHENTICATION OF ONLINE LEGAL RESOURCES (2010), *available at* http://aallnet.org/Documents/Government-Relations/authen-rprt-updates/2009aallauthenticationreportupdates.pdf.
[5] *Id.*
[6] *See generally* Michael Umberger, *Checking Up on Court Citation Standards: How Neutral Citation Improves Public Access to Case Law.* 31, no. 3-4 LEGAL REF. SERV. Q. 312-340 (2012).
[7] *Id.*

The idea that the government should provide citizens with online access to public government information has been catching on. Five states have enacted the Uniform Electronic Legal Materials Act, which requires "government publishers of primary legal materials in electronic format to authenticate, preserve, and provide permanent access to those resources."[8] But what does it mean to provide "access" to court opinions in an online context? Would merely placing PDF versions of court decisions on a server in an unorganized mess count as access? Or does access imply that the documents in question be findable and useable, rather than just available?[9]

In the context of legal research, access should mean, at the very least, the availability of primary law materials which a citizen could use to successfully research a basic legal question in a reasonable amount of time and in a reasonable manner. This is a "sliding scale" standard, which means that advances in technology and changes in the legal publishing industry should impact the responsibilities courts have in providing citizens with access to court opinions. This article assumes that courts should take reasonable steps towards providing citizens with access to court opinions. What is "reasonable" depends on factors such as cost, technological limitations, the technical knowledge of court staff, and the expectations of the citizens who will access these documents.

Given that citizens expect to find official government information on the Internet, it is no surprise that state and federal courts are moving in the direction of publishing court opinions in electronic format. However, simply hosting court opinions on a government server does not mean that these documents will be findable and useable by members of the public. As stated by Professor Widdison, "Raw cases, lacking editorial input in the form of e.g. catchwords, headnotes, lists of cases cited etc. are indigestible for lawyers and non-lawyers alike."[10]

In order to make court opinions useful, legal researchers have historically depended on third party indexing services, such as the West Key Number System. Now that the public uses the Internet as their preferred research tool, it would be worth looking at some of the major online services that provide access to court documents for free or at

---

[8] Judy Janes, *Why States Should Adopt UELMA*, Spectrum Online, *available at* http://www.aallnet.org/main-menu/Publications/spectrum/Spectrum-Online/uelma.html; *see also* Uniform Law Commission, *Electronic Legal Material Act, at* http://www.uniformlaws.org/Act.aspx?title=Electronic%20Legal%20Material%20Act (last visited May 5, 2013).

[9] Vinton Cerf, who helped develop the TCP/IP protocol, has argued that if "…the Internet plays a key role in the archiving and propagation of information in our system of justice, then it must be accessible to all who are seeking justice. Access means more than simply the ability to use a computer that is connected to the Internet—for example, at a public library. Access means the information is organized and presented in such a way that someone with a motor, vision, hearing, or other impairment that would inhibit convenient access to the Internet is accommodated." Although this argument focuses on accessibility for people with physical impairments, it can be extended to the population generally. True, meaningful access to information requires that the information be in a useable state. Vinton G. Cerf, *Internet and the Justice System*, 79 WASH. L. REV. 25, 27 (2004).

[10] Robin Widdison, *New Perspectives in Legal Information Retrieval*, 10 INT'L J. LEGAL INFO 41, 43 (2002).

little cost. Do these services succeed at providing meaningful access to legal information?

This article will begin by analyzing a sampling of the major free, online databases of court opinions, arguing that these options lack critical tools necessary to conduct case law research. In Part II, this article outlines different approaches the government might take in providing better online research tools to the public. Part III argues that the inclusion of detailed, standardized metadata with electronically published court opinions would go a long way toward improving the quality of free online legal information. Finally, Part IV of this article will address potential obstacles to the widespread implementation of this recommendation.

I.   Evaluating Free Online Sources of Case Law

Legal research is a complex task: "Undertaking legal research means knowing what to look for, knowing what you're searching for (federal or state law), knowing the terminology, knowing the code (legal abbreviations), knowing where to look, and knowing how one piece of information relates to another."[11] Historically, anyone could complete a basic research task using the print resources available at a law library open to the public. Law libraries meet this goal today by offering a hodgepodge of paper and electronic resources—e.g., a library might provide access to Keycite on certain terminals, but access to American Jurisprudence exclusively in print. However, people increasingly expect to be able to access information with a "public" character remotely and at no cost.[12]

To the novice researcher, it may seem like most of what they need is, in fact, online, in locations like court websites, Google Scholar, Justia, and the Legal Information Institute. These sources of case law will be evaluated using the following criteria identified by Professor Widdison to rate legal information retrieval systems: *accessibility*, *coverage*, *currency*, *reliability*, *searchability*, and *useability*.[13]

Accessibility, according to Widdison, includes issues such as cost, availability, and promulgation.[14] *Coverage* refers to the breadth and depth of the materials accessible in the system.[15] *Currency* captures the system's ability to provide updating tools like Shepard's and KeyCite, as well as the speed at which new legal materials are added to the system after initial publication.[16] *Reliability* concerns issues of stability, accuracy, and authenticity.[17] *Searchability* refers to the ease with which a user can access material in the system using searches by keyword, citation, or other element.[18] Finally, *useability*

---

[11] Steven Whittle, *Finding Law in the 21st Century: An Introduction to the SOSIG Gateway*, 29 INT'L J. LEGAL INFO 360, 372 (2001).

[12] *See generally* Candice Spurlin, *Permanent Public Access to Electronic Government Documents: South Dakota's Response to a National Dilemma*, 50 S.D. L. REV. 113, 119 (2005).

[13] Widdison, *supra* note 10 at 42.

[14] *Id.*

[15] *Id.*

[16] *Id.*

[17] *Id.*

[18] *Id.*

takes into account the general user-friendliness of the system, and how well the system lends itself to customization by users.[19]

Given how quickly the Internet changes, it is worth mentioning that this survey was conducted in the early months of 2013.

A.   Court Websites

Because each court or court system hosts its own unique website, it would be impractical, given the scope and purpose of this paper, to describe them all. Instead, this article will analyze two court websites (one at the federal level, and one at the state level) along with FDsys, the official digital repository for the federal government. The Ninth Circuit website was chosen because it is representative of other federal court sites, whereas the website for the Washington Court system was chosen because it is regarded as having a high-quality electronic case management system.[20] FDsys will be discussed on account of its collection of federal case law, part of a pilot program run by the Government Printing Office (GPO) in conjunction with the Administrative Office of the United States Courts.[21]

i.   Ninth Circuit Court of Appeals

Up until the mid-1990s, the U.S. court system depended on West Publishing to disseminate the official versions of its published decisions.[22] Supreme Court cases were the one exception: the GPO has been printing and selling the *Reports of the US Supreme Court* since 1922.[23] Now, each circuit court hosts its published opinions on its respective website, although the design, layout, and functionality of these databases vary from site to site.[24]

---

[19] *Id.*

[20] In 2007, King County won an award from the Harvard Kennedy School of Government for its electronic court records system. Harvard Kennedy School, Innovations in American Government Awards Program, *available at* http://www.innovations.harvard.edu/awards.html?id=85931 (last visited May 5, 2013).

[21] U.S. Government Printing Office, United States Courts Opinions, *at* http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=USCOURTS ("United States Courts Opinions (USCOURTS) collection is a pilot project between the U.S. Government Printing Office (GPO) and the Administrative Office of the United States Courts (AOUSC) to provide public access to opinions from selected United States appellate, district, and bankruptcy courts. The content of this collection dates back to April 2004, though searchable electronic holdings for some courts may be incomplete for this earlier time period. Once an opinion is located, all associated opinions within the same case can be accessed from the opinion More Information page").

[22] Xiaohua Zhu, *Access to Digital Case Law in the United States: A Historical Perspective,"* IFLA World Library and Information Congress, 78th IFLA General Conference and Assembly, 9 (2012).

[23] *Id.*

[24] See, e.g., the respective websites for the various circuits of the U.S. Court of Appeals: www.ca1.uscourts.gov; www.ca2.uscourts.gov; www.ca3.uscourts.gov; www.ca4.uscourts.gov; www.ca5.uscourts.gov; www.ca6.uscourts.gov; www.ca7.uscourts.gov; www.ca8.uscourts.gov; www.ca9.uscourts.gov; www.ca10.uscourts.gov; www.ca11.uscourts.gov; www.cadc.uscourts.gov.

In terms of accessibility, anyone with an internet connection can access the Ninth Circuit website, search for cases, and, assuming the search was successful, download copies of these unofficial opinions in PDF format. The layout of the website is clear and uncluttered, making it attractive from a useability perspective—but, as we will soon see, the actual usefulness of the website is limited, especially when it comes to its ability to function as a legal research tool.

As to coverage, opinions hosted on the Ninth Circuit website date back to 2005. Whether coverage over this period of time is complete was unable to be verified.

The Ninth Circuit website claims that published opinions appear in its online database by 10:00 am on the day of release, making it a very current source for this material. Unfortunately, individual documents, once accessed, do not include any state-of-the-law information. This means that users would need to consult an outside source to determine whether a case is good law, or whether it has been overturned or superseded by statute.

The search capabilities of the Ninth Circuit website are rather rudimentary. Users can search by case title, case number, case origin, case type, and date filed. Because of the absence of a keyword or subject search, the Ninth Circuit website is only useful for finding cases of which users already possess some familiarity. This limitation alone calls into question the Ninth Circuit website's usefulness as a legal research tool.

Overall, although the Ninth Circuit provides easy online access to its published opinions, its limited search functionality shows that it is not meant to work as a legal research tool. Rather, this is a database designed to provide access to known items.

ii. The FDsys United States Courts Opinions Database

Perhaps federal court websites feel it is unnecessary to provide the public with more advanced case finding tools (e.g. full-text searching, or the ability to search across multiple jurisdictions simultaneously) because of efforts by the GPO to provide just these kind of services. As part of a pilot program created in conjunction with the Administrative Office of the United States Courts, the GPO has begun hosting federal court opinions on its FDsys website.

As of now, FDsys hosts circuit court, district court, and bankruptcy court decisions dating as far back as 2004, but much of this collection is incomplete.[25] Coverage varies greatly across different courts and jurisdictions, and some courts have yet to participate in this pilot program.[26] It is also possible to find opinions older than 2004, but this is limited to opinions directly related to cases decided after the 2004 cutoff

---

[25] *See* U.S. Government Printing Office, *supra* note 21 (according to the GPO, coverage becomes less complete moving back in time, especially as one nears 2004).

[26] *Id.* (for example, FDsys hosts the following Court of Appeals content: Second Circuit opinions dating back to 2010; Fourth Circuit opinions dating back to 2005; Seventh Circuit opinions dating back to 2005; Eighth Circuit opinions dating back to 2001; Ninth Circuit opinions dating back to 2008; Tenth Circuit opinions dating back to 2004; and District of Columbia Circuit cases dating back to 1993. Coverage within these dates is inconsistent: for example, all pre-2004 cases are limited to those directly related to another, post-2004 case. In more recent years, coverage appears to be comprehensive).

(for example, where a case was originally decided in 2002, but appealed or otherwise revisited in 2005). Coverage will hopefully improve as this database moves beyond its pilot phase.

In terms of accessibility, the GPO allows users to access FDsys at no cost.[27] This, however, may change in the future: the GPO has implemented pay structures in the past, and seems to be flirting with this unpopular idea again.[28]

The long-term reliability of FDsys is questionable, given its current status as a pilot program. Other pilot programs created by the GPO have been abandoned in the past.[29] FDsys does, however, use a system of digital signatures to ensure the authenticity of these opinions.[30] Every night, FDsys pulls cases directly from the US Courts electronic case, processes them, and then makes them available to the public.[31]

In order to search this collection, users must conduct an advanced search from the FDsys homepage. Then, after selecting "United States court opinions," a researcher can use a drop-down menu to specify what kind of search she would like to conduct. Options include full-text, party name, court state, nature of suit, case number, party role, citation, and others. Multiple fields can be searched simultaneously, making it possible to, for example, conduct a full-text search within specific jurisdictions.

The FDsys database adds new cases within a few days of their original publication, making it relatively current, but not as current as, say, the Ninth Circuit website or pay services like Westlaw.[32] Opinions accessed through FDsys lack any state-of-the-law information; this would be difficult, or even impossible, to provide without a more comprehensive collection of federal decisions.

By hosting authenticated versions of federal court opinions on FDsys, the GPO has begun laying the groundwork for what might become a powerful, freely accessible

---

[28] James A. Jacobs, James R. Jacobs & Shinjoung Yeo, *Government Information in the Digital Age: The Once and Future Federal Depository Library Program*." 31, no. 3 J. ACAD. LIBR. 198, 199-200 (2005).

[29] See, e.g., the discontinued program that allowed law libraries free access to documents on PACER, an online system through which users can access electronic court records, including briefs and other docket material, for a fee. John Schwarts, *An Effort to Upgrade a Court Archive System to Free and Easy*, NEW YORK TIMES, Feb. 12, 2009, *available at* http://www.nytimes.com/2009/02/13/us/13records.html?pagewanted=all&_r=0; *see also* Zhu, *supra* note 22 at 14.

[30] U.S. Government Printing Office, *About United States Courts Opinions*, *at* http://www.gpo.gov/help/index.html#about_united_states_courts_opinions.htm ("The secure transfer of files to GPO from the AOUSC maintains the chain of custody, allowing GPO to authenticate the files with digital signatures").

[31] , United States Courts, *Access to Court Opinions Expands,* THE THIRD BRANCH NEWS, Jan. 31, 2013, *at* http://news.uscourts.gov/access-court-opinions-expands.

[32] On May 11, 2013, the most recently published Ninth Circuit case was McCullough v. Graber, 11-16920, 2013 WL 1924434 (9th Cir. May 10, 2013). As of May 11, this opinion could be found on Westlaw as well as the Ninth Circuit website. On FDsys, however, the most recent opinion was Sargsyan v. Holder, 08-72040, 2013 WL 1912465 (9th Cir. May 9, 2013).

legal research tool. As of now, inconsistent coverage means that FDsys works better as a retrieval tool than as a place to conduct research.

iii.   Washington State Court System

The Washington Courts website includes a freely accessible collection of slip opinions. Researchers can search this collection by keyword, and limit by date, court level, and publication status (i.e., whether the opinion is published or unpublished).

The coverage of the Washington Courts database of judicial opinions extends to Supreme Court and Court of Appeals opinions; the opinions of lower courts are not yet included. As of now, this collection dates back to just February of 2013. The website claims that, going forward, all opinions published on the site "will remain on [the] site for a minimum of 15 years."[33] For older opinions, the website directs researchers to the Municipal Research and Services Center of Washington website, which also includes a database of Washington decisions.

The case law hosted on the site takes the form of slip opinions. This means that not only can these cases not be cited to, but that they "are subject to reconsideration, modification orders, editorial corrections, and withdrawal."[34] Consequently, the opinions hosted on this site are of questionable reliability.

The slip opinions seem to be added as soon as they are published, making this source at least as current as Westlaw and Lexis. As was the case with the Ninth Circuit and FDsys websites, citator tools are not provided.

B.   Google Scholar

In terms of accessibility, Google Scholar excels: this is a free service, accessible by anyone with an Internet connection. Likewise, Google Scholar is highly useable. On the Google Scholar main page, users are confronted by a single search box, with options to choose between academic and legal documents.[35] This basic search functionality, unsurprisingly, is also one of Google Scholar's strengths. Users can type in whatever search terms they think will turn up relevant results, and are then able to filter results by date and jurisdiction. Once users settle on a document to view, Google automatically links to cited cases (assuming Google also hosts the cited case). Google has even developed a tool that will add Westlaw-like star pagination to its hosted court cases, allowing users to cite these documents in compliance with court rules.

---

[33] Washington Courts, Supreme Court and Court of Appeals Opinions, *at* http://www.courts.wa.gov/opinions/index.cfm?fa=opinions.displayAll (last visited May 5, 2013).

[34] *Id.* In order to access the official reports, the Washington Courts website directs users to the LexisNexis website. LexisNexis publishes the official versions of Washington case law; at the time this article was written, a bound volume of just the Washington Supreme Court Reports cost $23.50. Although this price seems reasonable at first blush, our common law system makes individual volumes practically useless when separated from a complete set.

[35] Google Scholar, *at* http://scholar.google.com (last visited May 10, 2013).

Google Scholar also provides a range of advanced search functions. Advanced search fields include the ability to conduct Boolean searches, with fields devoted to exact phrase, disjunctive, and conjunctive searching. Users can also tell the system to exclude results containing specified words. Most useful to legal researchers is the ability to limit searches by court, jurisdiction, and date. Google Scholar also lets users search by the name of the judge who authored an opinion. Strangely, the terminology used in the advanced search box is not well tailored for the searching of court documents. For example, the fields refer to all legal documents as "articles," which might be a consequence of the system's roots as a discovery tool for academic scholarship.

Google Scholar's coverage is relatively strong. It hosts all published U.S. Supreme Court cases, federal circuit and district court cases dating back to 1923.[36] State coverage is less comprehensive, limited to published state supreme and appellate court decisions dating back to 1950.[37] The decisions of lower level state and local courts are not yet included.[38] Although it remains unclear from where Google gets its case law, some speculate that its historical collection came from Fastcase, by way of Public.Resource.Org.[39] Google automatically analyzes the citations present in each case, and hyperlinks to cited cases also in its database. Citations to statutes and regulations are not hyperlinked, since Google does not yet include these materials in its database.

In terms of currency, Google Scholar cannot keep pace with pay services like Westlaw. A search on Google Scholar for the most recent circuit court opinions showed that it takes about three days for published opinions to find their way onto Google servers. In comparison, both Westlaw and the respective circuit court websites were completely up-to-date in their collections of published opinions.

Furthermore, Google does not provide users with citator tools comparable to KeyCite or Shepard's. This means that once a user views a case on Google Scholar, she is not given an immediate indication of whether the case is still good law, or whether the opinion is the final decision on the case.[40] This could be particularly dangerous to inexperienced lay users, who might not consider this issue unless the information were provided in a visible, easy-to-understand fashion.

Google Scholar does have a rudimentary citator, which allows researchers to see where a case has been subsequently cited. Although Google indicates the depth in which

---

[36] Georgetown Law Library, *Free and Low Cost Legal Research Guide*, *at* http://www.law.georgetown.edu/library/research/guides/freelowcost.cfm (last visited May 5, 2013).

[37] George Mason University School of Law, *Free Legal Research Sites, at* http://www.law.gmu.edu/library/research (last visited May 5, 2013).

[38] Google Scholar, *at* http://scholar.google.com/scholar_courts?hl=en&as_sdt=4,48 (last visited May 5, 2013).

[39] Andrew Plump-Larrick, *Case Law in Google Scholar, at* http://lawlibrary.case.edu/2009/11/17/case-law-in-google-scholar (last visited May 5, 2013); Public.Resource.Org, *Announcement: 1.8 Million Pages of Federal Case Law to Become Freely Available*, Nov. 14, 2007, *at* https://public.resource.org/case_law_announcement.html (last visited May 5, 2013).

[40] For example, a search for the Washington case "State v. Eriksen" in Google Scholar turns up six identically named documents, all related to the same case, State v. Eriksen, 172 Wash. 2d 506, 259 P.3d 1079 (2011). But, Google Scholar does not tell users how these various opinions are related to one another, and which is the final decision.

the cited case has been discussed in subsequent cases, it gives no indication as to *how* the subsequent case cited the preceding case.[41] Nevertheless, a determined and cash-strapped legal researcher could use Google Scholar to check whether a case is good law—it would just be a matter of reading through each decision citing the case is question. This would be a time consuming task for a trained lawyer; for a layperson, reading and interpreting dozens of citing cases might be completely impractical.

The reliability of the Google Scholar system is difficult to judge. On one hand, Google has an excellent record when it comes to uptime/downtime.[42] On the other hand, further research is needed to discern the accuracy of the court documents Google does host.[43] Furthermore, even if the court documents hosted by Google tend to be highly accurate, this doesn't change the fact that these are not official, authenticated documents. Finally, it may be too early to judge the stability of the Google Scholar system, given the recentness of its creation. Google is a for-profit organization; if hosting legal documents becomes unprofitable, it would make sense for Google to abandon the project.

Google Scholar, overall, does an excellent job with the services and tools it chooses to provide. Among the free legal information resources analyzed in the article, it is by far the most user friendly. Because Google Scholar does not let users know whether a case is still good law, however, it does not work as a viable alternative to pay services like Westlaw and Lexis.

B.   *Justia*

Being a free source of legal information on the Internet, Justia receives high marks for accessibility.[44] The site includes a range of search options, allowing searches by party name, keyword, and presiding judge, among others. The search bar is surprisingly difficult to use, however, because the functionality of the search bar changes as users click around the site; e.g., in order to search in a particular jurisdiction, it is necessary to first browse to that jurisdiction. Consequently, it is impossible to conduct searches across a specified selection of jurisdictions on the Justia system.

One strength of Justia is that it is somewhat browseable. Users can browse by area of law and by legal topic. For some users, this may be a more efficient way of finding materials than through full text keyword searching.

In addition to hosting collections of federal and state case law, Justia provides access to federal and state statutes and regulations. Justia is more explicit than Google about its coverage; when browsing to federal circuit court decisions, for example, it

---

[41] Interestingly, Google Scholar does have something called a "how cited" feature, but this simply pulls quotes from the citing case, rather than giving an editorial characterization of the citation (e.g. "distinguished by," "discussed in," etc.). This may have to do Google's reliance on automation in providing these tools, whereas the LexisNexis and Westlaw citators rely on a combination of automation and human editors.

[42] 99.99981% uptime from March 2012 to March 2013. MonTools, Uptime/Downtime Stats, *at* http://montools.com/stats/www.google.com (last visited March 30, 2013).

[43] This is an especially important concern given the amount of automation Google relies upon.

[44] Justia, *at* www.justia.com (last visited May 5, 2013).

clearly shows the user that it hosts these decisions dating back to 1901. It even indexes cases not present in the Justia databases; e.g., many federal district court cases are not available in full text, but rather link to the PACER system. Justia does not hyperlink across most of its hosted primary law materials.

Justia does not fare well when it comes to providing currency information. Like Google Scholar, the system does not tell users whether a case is still good law. Unlike Google Scholar, Justia does not display or link to citing cases. Thus, although Justia hosts a large breadth of legal content, statutes and regulations included, and features some subject indexing, it cannot compete with Westlaw or Lexis when it comes to researching case law.

C. *Legal Information Institute*

The Legal Information Institute at Cornell University Law School (LII) has a similar layout to Justia, supporting access through searching or browsing. In this sense, LII is highly accessible.

Unfortunately, LII's coverage is spotty, and seems to be getting spottier.[45] Regarding case law, LII only hosts Supreme Court decisions. When browsing to the decisions of other courts, LII redirects the user to the court's website.[46] LII does host the U.S. Code and Code of Federal Regulations, and generally links to these sources from its Supreme Court decisions. LII also provides access to some secondary materials, which includes a legal encyclopedia, popular name tables, and LII's own editorial material, among others. Thus, although LII contains a great range of legal material, it cannot compete with the likes of Google Scholar when it comes to searching for and accessing case law.

The reliability of the LII system is questionable. LII, unlike Google Scholar and Justia, is a not-for-profit group that depends on user donations to operate, at least in part. Unfortunately, this means that LII often appears underfunded, and, in at least one instance, LII has had to eliminate services for budgetary reasons.[47]

Much like Google Scholar and Justia, LII does not let users know whether a case is still good law. This, combined with its spotty coverage, means that LII is nowhere near an alternative to pay services like Westlaw or LexisNexis.

---

[45] *See* Legal Information Institute, *US Circuit Courts—End of Life Announcement*, *at* http://www.law.cornell.edu/lii/eol/usca ["As of today (March 1, 2013), the LII is discontinuing its search engine for the United States Courts of Appeal. It has proven too difficult for us to maintain given our small staff and the absence of standard approaches across the 13 Circuit Courts of Appeal. We simply can't offer the service at the level of quality that our audience deserves, and are therefore discontinuing it rather than offer subpar results"].

[46] Once at a court website, users are generally confronted with very rudimentary case finding tools. For example, the Ninth Circuit's advanced search requires the user to know the case title, number, or date to conduct a search, making it useful only for accessing "found" items. *See* United States Court for the Ninth Circuit, Published Opinions, *at* http://www.ca9.uscourts.gov/opinions/ (last visited May 5, 2013).

[47] *See* Legal Information Institute, *supra* note 45.

D. Final Remarks on Free Online Sources for Case Law

Although online services like Google Scholar, Justia, and LII host great volumes of legal documents, they do not provide services like Shepard's or KeyCite. Instead, these services rely on full-text searching combined with some basic forms of automated indexing (e.g., by identifying the year of the decision and the name of the court that issues the opinion). To better serve the needs of legal researchers, these free services need the capability to tell users whether a case is still good law. Subject indexing, hyperlinked citations, and the ability to conduct faceted searching would make these free services even stronger, and many of these free services already provide some of these functions. Given some time, and with a little help from the courts, it's easy to envision these services evolving into viable alternatives to Westlaw, Lexis, or Bloomberg Law for the casual legal researcher.

II.   The Government's Role in Providing Access to Judicial Opinions Online

As we have seen, the needs of layperson legal researchers are not being met by free sources for online case law research. Authoritative sources like court websites often host collections of court documents without providing robust finding tools. Private sector sites like Google, Justia, and LII have spotty coverage and lack critical features present in pay services like Westlaw, Lexis, and Bloomberg Law. Given tight government budgets and changing public expectations, solving this problem will be difficult. By committing to making court opinions freely available to the public, courts can utilize the free market to help mitigate these access issues at little expense to themselves.

Some have argued that efforts to provide legal information to the public over the Internet has been hindered by the fact that the "private sector has played a disproportionately large role in cyberinfrastructure funding with the result that efforts are often uncoordinated and fragmented."[48] These critics would like to see the government play a larger role in the legal information industry.[49] It is certainly the case that free legal research services on the web have spotty coverage and fail to be well-integrated into the broader universe of legal information.[50] However, the alternative—having the government fund a freely accessible legal information portal—also seems unrealistic, given budgetary restraints and the complexity inherent in such an undertaking.

Instead, courts should utilize the free market system by taking small, cost-effective steps towards helping sites like Google Scholar thrive. The best way to

---

[48] Eugene Clark, *Future Trends in Legal Research and Scholarship: Implications of the Establishment of a Cyberinfrastructure for E-Research*, 16 JL INF. & SCI. 114, 122 (2005).

[49] For background on this debate, *see generally* Zhu, *supra* note 22 at 14 ("Based on the notion that the law is a public good, some scholars argue the government should replace private publishers as the primary distributor of legal information").

[50] For example, Google Scholar does not deal with statutes in any way, though it has a good collection of case law; LII does integrate statutes into its platform, but its case law collection extends just to Supreme Court decisions.

accomplish this now would be for courts to focus their efforts on providing rich metadata in their electronically-published opinions.

According to Robinson et al., building an interactive site that searches, displays, and organizes information is easy, when the source data is freely available in a standardized format: "Web hosting is cheap, software building blocks are often free and open source, and new sites can iterate their designs rapidly."[51] In other words, the creation of a Westlaw Next-style search interface is easy; the difficult part is amassing the data to be searched, storing this data, and indexing the data in such a way that it can be searched efficiently and effectively.

Focusing on the data itself rather the creation of finding tools and web portals also minimizes the risk of expensive government projects becoming quickly obsolete. The Internet exists in a state of flux and change; once popular websites can become ghost towns overnight.[52] Frankly, when one thinks of highly flexible, quickly adaptable organizations, the government doesn't jump to mind as an exemplar of these qualities. An expensive, labor-intensive government information system could be replaced overnight by an innovative Internet startup. This isn't a bad thing at all; in fact, this phenomenon is a prime reason why we use a free market system. Government information is meant to be utilized in just this way by private parties, which is one reason why government documents at the federal level do not receive copyright protection.[53]

Relatedly, government funding can be inconstant and unpredictable.[54] Some have argued that the GPO (and the government in general) should not be in the business of providing permanent access to government information, due to concerns about adequate funding.[55]

Professor Xiaohua Zhu, while not taking sides on this issue, notes the following additional concerns regarding government-sponsored case law databases:

> …the difficulty of collecting historical case law, the expense of maintaining a large system, the lack of legislation, the often-changing information policies of government agencies, and the different ideologies or value systems about governments' roles in providing public access to the law … [m]any still hold the belief that the government should not compete with the private sector in the area of information dissemination but should leave the business opportunity to commercial information providers.[56]

---

[51] David Robinson, Harlan Yu, William Zeller, & Edward Felten, *Government Data and the Invisible Hand,* 11 YALE J. L. & TECH. 159, 161 (2009).

[52] For example, the once-popular social media sites Friendster and MySpace suffered declines in popularity due to the rise of Facebook. *See generally* Liat Clark, *Researchers Conduct 'Autopsy' of Social Network Friendster*, WIRED.CO.UK, at *http://www.wired.co.uk/news/archive/2013-02-27/autopsy-of-friendster* (last visted May 5, 2013) (noting that social media sites can be prone to sudden mass exoduses).

[53] 17 U.S.C. § 105.

[54] *See* Silverman, *supra* note 2 at 185.

[55] *See* Jacobs et al., *supra* note 28 at 203.

[56] *Supra* note 22 at 14, 15.

Still, a strong case can be made for government involvement in publishing and disseminating government information. As the American Association of Law Libraries (AALL) has recognized, it is vitally important that the public sector retain responsibility for the preservation of authentic government records.[57] In the private sector, businesses come and go unpredictably, and when they go they tend to take their information with them. Governments, on the other hand, tend to be much more stable and enduring. Thus, although the private sector is well-suited for developing innovative finding, searching, and organizational tools for electronic government documents, it would be unwise to rely on private companies for storing these documents in a permanent state.[58]

Already, examples exist of the private sector creating portals to government data when the government-provided portal was seen as inadequate. In response to perceived shortcomings of the Regulations.gov website, Cornell helped launch the "Regulation Room," a website dedicated to providing access to federal regulations in a Web 2.0 context, complete with RSS feeds, blogs, and social networking features.[59]

In summary, one way to guarantee meaningful public access to case law online would be to create a government-sponsored, online legal research service, which would function as a sort of "Westlaw light." Given the state of government budgets, however, this proposal seems unrealistic. At the opposite side of the spectrum, we could take the view that it isn't the courts' role to publish court opinions; rather, that this is a job best left to the legal publishing industry. This approach ignores very concrete problems in the way court opinions are published and accessed, such as the skyrocketing costs of Westlaw-like services, the general movement away from print resources, and the shrinking budgets of law libraries. Instead, courts should embrace a balanced approach, and take reasonable steps toward providing meaningful access to court opinions where they can. By utilizing the free market, and thinking of themselves as providers of data rather than providers of access, courts can improve the state of electronic legal research at little cost. In Part III, this Article will detail *how* structured data could improve free, online case law databases.

III.   Digital Case Law, Structured Data, the Semantic Web

Part III of this Article will detail how structured data could improve free online legal research tools. First, this Article will provide a basic overview XML, a software technology that enables the easy creation of standardized metadata. Next, this Article will describe a few basic, practical ways XML could be applied. This Article will then

---

[57] *See generally* Ass'n of Law Libraries Gov't Relations Comm., *supra* note 4; *See also* Jacobs et al., *supra* note 28 at 203-204 (arguing that both the government and Federal Depository Loan Program libraries have important roles to play in publishing, storing, and providing access to government documents).

[58] *See generally* Jacobs et al., *supra* note 28 at 25 (arguing that preservation of and access to digital government documents should be provided by government actors working in conjunction with depository libraries, whereas the role of the private sector should be "re-packaging, re-organizing, and re-distributing the information").

[59] Claire Cardie, Dan Cosley Cynthia Farina & Mary Newhart, *Rulemaking 2.0*, 65 U. Miami L. Rev. 396-447 (2011).

look at a few current implementations of XML in court systems at the federal and state levels. Finally, this Article demonstrates how widespread adoption of uniform metadata standards would help courts keep abreast with advances in Internet technology, and the Semantic Web in particular.

A. *What is XML, and how does it work?*

XML stands for "eXtensible Markup Language," where markup refers to "information embedded in the text of a document that is not intended for printing or display."[60] Rather, this is *metadata*, which is meant to be read by machine. In this limited sense, XML can tell a computer what a document means, and the greater detail included in the metadata, the more a computer can "know" about a document.[61]

The key strength of XML is that it is a standardized format, yet flexible enough to be applied across different information systems. Thus, each court could continue to use its own case management system, but these systems could "speak" to each other by means of a shared language, i.e., XML.[62] Furthermore, adoption of XML would allow for better integration of court records with government documents published by other branches of government, such as statutes, regulations, and administrative rulings.

Silverman identifies three broad categories of XML metadata. First, there is procedural markup (also called presentational markup). This metadata tells a computer how to display a document. This would include, for example, instructions about font type, font size, margin size, etc. Second, there is structural markup, also known as descriptive markup. This metadata identifies the general type of data being tagged—e.g., this would tell a computer the title of the document. Lastly, and most importantly for the purposes of this article, comes semantic markup. This markup tells a computer what pieces of data in a document mean, in a limited sense. For example, semantic markup could tell a computer which part of a court opinion is the holding, or which part includes party names.[63]

Metadata in XML format can be added to a document through "tagging."[64] For example, the party names in a court document would be identified by a tag like "PartyName." Here, it doesn't matter how the tag is characterized (e.g. "PartyName" versus "Party") as long as the tags are applied consistently within the court and across different courts. A court document "tagged" in this manner would look something like this, in small part:

(PartyName) Doe, John (/PartyName)

---

[60] Silverman, *supra* note 2 at 187.
[61] *See id.*
[62] Silverman, *supra* note 2 at 185.
[63] *Id.* at 187.
[64] *Id.* at 188.

14

A computer programmed to read XML would be cued in to the existence of metadata by the text "(PartyName)" and would know to stop applying this tag once it gets to "(/PartyName)." "(PartyName)" is referred to as a start tag, and "(/PartyName)" would be an end tag. XML also allows for empty tags, which can be used to place markup referring to the document as a whole, rather than specific blocks of text or metadata. All tags can be nested, which allows for multiple tags to be applied to the same chunk of text as well as the tagging of tags themselves.

Displayed in this way, with the metadata tags visible, XML looks cumbersome. Fortunately, a variety of easy-to-use computer applications (and computer application add-ons) have been developed to make this process more user-friendly. Currently, anyone can download a plugin for Microsoft Word enabling users to tag text simply by highlighting it.[65] Thus, once a court settles on a standardized tagging language[66] (e.g. "PartyName" to be applied to the names of the parties), this process would involve someone, perhaps a judge, clerk, or other court personnel, simply highlighting text and applying tags from a preset list.

B. Applications: citators, hyperlinked citations, and beyond

The inclusion of high quality, structured metadata in electronically published court opinions is a low-cost way to improve the quality of online research tools—whether free or otherwise. At the very least, the inclusion of metadata relating to how cases are cited and discussed within an opinion would allow for third parties to easily create programs to mimic Shepard's or KeyCite, as well as hyperlinking across court documents.

Westlaw's KeyCite and Lexis' Shepard's both rely on a combination of human editors and computer algorithms to function. In both systems, human editors provide some case analysis upfront, but most of the work is handled algorithmically.[67]

It is this human component that essentially bars Google Scholar and its ilk from providing comprehensive citator tools. Computer algorithms require upfront costs to develop, but are otherwise inexpensive; human editors require salaries and health insurance. If courts completed some of the work of Westlaw's and Lexis' human editors upfront, Google would just need to develop an algorithm—a task Google has a track record of doing well. This would also have the effect of lowering Westlaw's and Lexis'

---

[65] Benjamin J. Keele & Michelle Pearse, *How Librarians Can Help Improve Law Journal Publishing*, 104, no. 3 L. LIB. J. 400, n.81 (2012) (discussing the Ontology Add-In for Word, Microsoft Research, *archived at* http://www.webcitation.org/67XhdNrHn).

[66] This predefined list of possible elements, attributes, and tags is called an "XML standard." *See* Silverman, *supra* note 2 at 193.

[67] Laura C. Dabney, *Citators: Past, Present and Future,* 27, no. 2-3 LEGAL REF. SERV. Q. 165, 176 ["today both programs do case analysis editorially, and the rest of the process, including finding citations and headnote assignment (and, in KeyCite, depth of treatment and quotations),
is done programmatically"]; *see also* Susan N. Mart, *The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis*, LEGAL REF. SERV. Q., *forthcoming, available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188541, 14 (2012).

cost of doing business—and given how expensive subscriptions to these services can be, we can hope that these savings would be passed on to customers.

Pay services like Westlaw and LexisNexis do much more than provide reverse-citation indices and state-of-the-law information.[68] They also provide hyperlinks across different sources of law, integrated secondary sources, and advanced search functionality, among other features. Users can search across multiple jurisdictions, state or federal, and multiple sources of law simultaneously. These systems also provide original editorial content, such as headnotes, and comprehensive indexing systems, such as the West Key Number System.

Assuming that other branches of government continue to expand their use of XML,[69] we can imagine free services providing links from court documents to, for example, provisions in the United States Code or Code of Federal Regulations. The more metadata courts decide to place in opinions, the better online research tools can be. By issuing their own headnotes, as is already done by some courts, and embedding these notes with structured metadata, free legal research services could easily provide users with the ability to browse legal materials by topic.[70] If standard subject tags were adopted across jurisdictions, this would enable this browsing to occur across jurisdictions. This practice would have the side benefit of giving courts greater control over how their decisions are found, interpreted, and utilized.

According to Silverman, another benefit of XML is that it "permits information in court records to be shared with the public at the courthouse and over the Internet while respecting the legitimate privacy interests of litigants and others who come before our courts."[71] This is important because privacy concerns have been cited as one reason courts should be cautious about accepting filings and publishing decisions on the Internet.[72] Basically, XML would allow a court to tag elements of a document as public, and accessible to anyone, or private, and accessible to just authorized individuals (such as court personnel, or the parties to the dispute). Such a system would be designed so that information deemed private would be automatically culled from the document as it is

---

[68] Jonathan Jenkins, *What Can Information Technology Do for the Law?*, 21 HARV. J. L. & TECH. 589, 595 (2008).

[69] Already, the FEC has begun offering raw bulk data in XML format, accessible to anyone with Internet access at no charge. *See* Federal Election Commission, Downloadable Data Files, *at* http://www.fec.gov/finance/disclosure/ftp_download.shtml (last visited May 5, 2013). The federal register is available in XML format as well. *See* Mary A. Baish and Emily Feldman, *White House, GPO and NARA Collaborate on Release of XML Version of Federal Register,* American Association of Law Libraries, Washington Blawg, *at* http://aallwash.wordpress.com/2009/10/06/white-house-gpo-and-nara-collaborate-on-release-of-xml-version-of-federal-register-2/.

[70] For example, the Washington Courts Reporter of Decisions adds headnotes to opinions before publication, using its own classification of legal topics. Here, taking the extra step of encoding this information in XML format would not require much extra work at all. *See* Washington Courts, Supreme Court and Court of Appeals Opinions, *at* http://www.courts.wa.gov/opinions (last visited May 5, 2013).

[71] Silverman, *supra* note 2 at 175.

[72] *See id.* at 193; *see also* Peter A. Winn, *Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information*, 79 WASH. L. REV. 307, 327 (arguing for an approach that balances privacy against the public's interest in accessing court documents online).

accessed, whether this access is achieved manually by an individual user or automatically by a Google Scholar-like service.

Traditionally, we have viewed court opinions almost as literary creations with a narrative quality, rather than as rigid, uniform sets of facts, rules, and holdings. However, the future may require us to rethink this view. Silverman argues that by "tagging all the information contained in a court document, it is possible to dispense with documents altogether—through dissolving them into structured information."[73] In this way, structured metadata would allow a legal researcher to easily identify and access just the part of an opinion that is relevant to her purposes. Perhaps more importantly, structured metadata would allow an automated computer program to identify, isolate, manipulate, and display individual components of a court document.

Beyond these examples, it is difficult to predict what else third parties might be able to do with court metadata. If we have faith in the free market's ability to spur innovation (especially in sectors of the economy with plenty of competition, as is increasingly the case in the market for free and low-cost online legal research), then XML might prove to be a cost-effective way to give the public access to high-quality, free legal research tools.

At a certain point, a legal information system can become so advanced as to encroach on the traditional role of the lawyer. At this point we might ask whether such a system is actually engaged in the practice of law. Some have argued, in fact, that these restrictions on the giving of legal advice are stifling innovation in the legal technology field.[74] For example, Google Scholar now offers a way to view lists of citing cases for any legal document. As discussed above, this service does not include a characterization of these citations—e.g., whether the citing case overturned the prior case, or affirmed it. Assuming that Google were willing to devote the time and energy to creating its own KeyCite-like service, would this constitute the giving of legal advice?

By providing their own characterizations of case treatment in explicit, machine readable terms, courts can help organizations like Google steer clear of these muddy legal issues. If a Shepard's like system could be built off of the metadata included in court documents, then Google would be displaying this metadata in a particular way, rather than adding its own interpretation of the content.

In summary, XML is a cheap, simple way to embed metadata in electronically published documents. By including good metadata in their opinions, courts can help ensure that their documents will be findable, useful, and machine-readable. Even basic tagging could be a boon to free sources of legal information, enabling features like cross-referencing, reverse-citation indexing, and features mimicking those of Shepard's and KeyCite The more structured data included with electronically published court documents, the better free sources of legal information can mimic the advanced features of Westlaw, LexisNexis, and Bloomberg Law.

---

[73] Silverman, *supra* note 2 at 198.
[74] Jenkins, *supra* note 68 at 591.

C.   Current implementations

Given the low cost and potential benefits of structured metadata, it isn't surprising that some courts have already taken steps towards implementation. For example, California requires all motions to contain case numbers, date and time information, and party names in a machine-readable format.[75] Likewise, Minnesota has adopted a standardized metadata language to be applied to the opinions of its courts, and so has Washington State.[76] The National Center for State Courts hosts XML standards that state courts are free to download and deploy in their own electronic case management systems.[77] Often, states implement XML on a small scale to solve a specific problem, then expand XML adoption to increase interoperability across different systems and government bodies.[78]

At the federal level, both PACER and FDsys use XML to provide metadata with their hosted court opinions.[79] FDsys allows users to view an opinion's metadata in XML format; although this information would be most useful to the likes of Google, to the curious researcher it does offer a glimpse into the inner workings of FDsys database. In FDsys, XML serves to both provide descriptive metadata necessary to conduct complex searches and to authenticate its hosted court opinions.

As the above examples illustrate, the groundwork for widespread adoption of XML has already been laid, both at the federal and state levels. Going forward, we can expect to see more state court systems adopting this technology. Librarians, open-access activists, and the Googles of the world should educate government bodies about the importance of adopting standardized XML languages, as well as the importance of making this metadata freely accessible by users and automated indexing programs over the Internet.

D.   Looking towards the future

The "semantic web" has quickly become a buzzword among information professionals. The term "semantic web" refers to an Internet imbued with machine-discernable meaning.[80] It is not "a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people

---

[75] *Id.* at 593.

[76] *See* King County Superior Court Clerk's Office, *Legal XML*, *at* http://www.kingcounty.gov/courts/ECRlibrary/how-does-ECR-work/technology/legal-XML.aspx (last visited May 5, 2013).

[77] National Center for State Courts, *Court-specific Standards,* at *http*://www.ncsc.org/services-and-experts/technology-tools/court-specific-standards.aspx (May 5, 2013).

[78] Robin Gibson, *Information Sharing and Extensible Markup Language (XML),* FUTURE TRENDS IN STATE COURTS 60 (2006).

[79] *See, e.g.,* United States Courts, *XML Tags: Appellate CM/ECF V.3.1, PACER Docket Report and Case/Party Searches,* Feb. 23, 2010, *at* http://www.pacer.gov/documents/pacer_xml_apv3.1.pdf (last visited May 5, 2013).

[80] *See generally* Jenkins, *supra* note 68 at 602-604; *see also* Richard Benjamins, Pompeu Casanovas, Joost Breuker & Aldo Gangemi, *Law and the Semantic Web, an Introduction* 1-17 (2005).

to work in cooperation."[81] This "meaning" grows out of the extensive use of metadata tags, whether these tags are applied manually or through an automated process.[82] Thus, a new focus on structured metadata would put courts in line with general Internet trends.

The "semantic web" requires three main components: 1) a programming language that can capture relationships among different documents and the concepts held in these documents 2) a set of rules that defines how these relationships can be manipulated and 3) a flexible, standardized format for storing data and metadata.[83]

Thanks to XML, courts have access to the third component of the semantic web. The first two components, however, require the development of more complex tagging languages and legal ontologies. Tagging party names and cited cases is easy; developing a broad structure to capture the relationship among, e.g., various forms of trespassing is much trickier.

Law librarians could play a role in this process. For example, Keele et al. argue that law librarians are particularly well-suited for the task of creating legal taxonomies and ontologies.[84] They note that a "shared taxonomy/ontology emanating from the legal academy could also be "mapped" to taxonomies/ontologies developed for more practical or public uses such as projects in the open law movement or internal governmental use."[85] The integration of a comprehensive legal ontology with large volumes of court documents containing metadata in a standardized format could result in powerful legal research tools, with content linked across varied sources of legal information.[86] For example, third party developers could create applications that search court documents, legislative statutes, and agency regulations simultaneously, and display just the relevant information from these sources in a seamless way.

Once courts become comfortable tagging the "easy" part of a case (i.e. the parts that are more or less objectively identifiable, like citations or party names), they could begin experimenting with more complex and detailed tagging practices. For example, a court might implement a system for tagging specific elements of a cause of action. If this practice were widespread, it would be relatively simple to design a program that could pull up the rules underlying a cause of action in specific jurisdictions.

Beyond this, it is difficult to predict how legal research might change once legal information is integrated into the Semantic Web. However, experts "…who develop new information technologies describe a future with intelligent computerized legal assistants that can scour databases and outline arguments in place of low-level associates, as well as sophisticated software agents that can negotiate contracts without the direct involvement of attorneys."[87]

---

[81] Benjamins et al., *supra* note 79 at 4, 5.

[82] *See id.* at 11, 12.

[83] Jenkins, *supra* note 68 at 603 (citing Tim Bemers-Lee et al., *The Semantic Web,* Sci. AM., May 2001, 36-37).

[84] Keele et al., *supra* note 65 at 399.

[85] *Id.* at 400.

[86] *Id.*

[87] Jenkins, *supra* note 68 at 594.

IV.   Obstacles to Implementation

The widespread adoption of detailed, uniform metadata standards would be an inexpensive means to help improve the public's online access to judicial opinions. Nevertheless, the large-scale implementation of uniform metadata standards will be difficult. Limited court budgets are one obvious barrier to implementation. Additionally, the general culture of the judiciary may make it difficult to convince courts to take on the extra work of tagging each opinion. Finally, some may argue that advances in information technology could potentially make the recommendations of this article unnecessary.

One potential barrier to the adoption of uniform metadata standards for court opinions stems from the culture of the legal community.[88] The computer literacy of judges across the United States is likely subject to great variation. Further, judges may be hesitant to implement initiatives that would add to their workload. Thus, although it would be most efficient for judges (or their clerks) to create XML tags as they write an opinion, this task would likely fall to court staff—perhaps the court reporter could take on these duties.[89] Still, it would be time consuming to sift through each opinion in order to tag important elements—including every citation. Courts may be understandably reluctant to either hire additional employees or add to the workload of existing staff.

The basic tagging of a court document would be very simple, however—just a matter of highlighting text and applying tags from a preset list. Basically, this process would be similar to a law student marking up a court opinion with highlighters of different colors. Additionally, in the world of online publishing, the onus is increasingly on an author of a work to make her work findable on the web.[90] Finally, many aspects of the process can be automated, and this will be increasingly the case as technology continues to improve.[91]

Another argument against courts spending time and resources to create detailed metadata in their opinions is that computers may soon be able to handle many of these tasks automatically. As discussed above, one concrete benefit of tagging citations is that this practice could enable the easy creation of citator tools. But what if a program existed that could provide state-of-the-law information automatically, by some sort of computer algorithm?

---

[88] *Id.*

[89] In fact, the National Court Reporters Association took an interest in XML as far back as the year 2000. *See* Eddie O'Brien, *Legal XML: New Standards for Legal Documents,* J. COURT REP., May 2000, *available at* http://www.ncra.org/News/content.cfm?ItemNumber=10898 (last visited May 5, 2013).

[90] *See* Keele et al., *supra* note 65 at 400 [suggesting that the authors of law review articles should be "encouraged to create their own metadata (and apply a supplied ontology/taxonomy) in the authoring process…"].

[91] *See* Widdison, *supra* note 10 at 67 ("Furthermore, we may soon expect our computers to be able to identify key fields within a legal document with a high degree of accuracy. In relation to court decisions, for example, such fields will include findings of fact, statutes, regulations and cases cited, and statements of applicable law—the components of a machine-generated headnote, perhaps?")

In fact, at least three online legal research services purport to do just that (though, two of these are already defunct).[92] The automated citator still in use at the time of writing is the one developed by Fastcase, a low-cost alternative to the likes of Westlaw and Lexis. This citator works by combing opinions for Bluebook signals indicating negative treatment (e.g. *overruled by, abrogated by, overruled on another ground).* So, for example, when the computer algorithm reads Craigmiles v. Giles, it takes note of the following citation (emphasis added): "*See, e.g., Lochner v. New York,* 198 U.S. 45, 57, 25 S.Ct. 539, 49 L.Ed. 937 (1905), ***overruled in part****, Day-Brite Lighting, Inc. v. State of Missouri,* 342 U.S. 421, 72 S.Ct. 405, 96 L.Ed. 469 (1952)."[93] The algorithm then knows that Lochner is no longer good law, and this information would be displayed to a researcher viewing the Lochner opinion.

This process for finding negative case treatment carries with it two critical limitations. First, it relies on judges (or their clerks) to cite to overruled cases according to strict Bluebook guidelines. Although this practice is the norm, a preliminary search turned up exceptions. For example, *Planned Parenthood of Se. Pennsylvania v. Casey* cites to the overruled case *Adkins v. Children's Hospital* without employing standard Bluebook signals. As a result, automated citators would miss the negative treatment of these cases, unless they were cited according to Bluebook rules elsewhere.

Second, and even more alarmingly, it will miss negative case treatment for cases that have not been discussed much in other opinions. This is because, when a court overrules a decision, machine-readable Bluebook signals indicating negative treatment of that decision are nowhere to be found. In other words, if case *A* overrules case *B,* this is usually indicated just by a judge stating "we hereby overrule case *B"*—and the exact wording of such a statement will vary from opinion to opinion. In these situations, the information indicating that the prior case is overruled hides away within the prose of the decision. It is only when case *C* cites to case *B* that machine-readable Bluebook signals crop up. This is why, as discussed above, KeyCite and Shepard's require human editors (to some extent, at least).

This limitation has the danger of misleading novice legal researchers. It is especially problematic that the negative treatment most likely to be missed by such a citator would involve recent cases. For this reason, fully-automated citators should not be used as an authoritative source of state-the-law information.[94] By including specific case treatment information in the form of structured metadata, courts can ensure that

---

[92] Bonnie Shucha, *Review of Free Citators Available in Altlaw & PreCydent,* WisBlawg – From the UW Law Library, May 21, 2008 *at* http://www.law.wisc.edu/blogs/wisblawg/2008/05/review_of_free_citators_availa.html (last visited May 5, 2013); Fastcase, *Meet Our Newest Team Member, Bad Law Bot!, at* http://www.fastcase.com/badlawbot (last visited May 5, 2013).

[93] Craigmiles v. Giles, 110 F. Supp. 2d 658, 662 (E.D. Tenn. 2000) *aff'd,* 312 F.3d 220 (6th Cir. 2002).

[94] In fact, Fastcase's citator, the Bad Law Bot, carries with it the following disclaimer: "Keep in mind that Bad Law Bot determines negative case history by using algorithms, and that it is not intended to be a complete replacement for a full editorial citator or for reading all later-citing cases… If a case has been overturned but no court opinion has cited to it yet, Bad Law Bot won't be able to find any citation signal information." Fastcase, *supra* note 92.

fully automated citators offer better and more complete state-of-the-law information to researchers.

Conclusion and Recommendations

Although free online sources of case law like Google Scholar host great volumes of legal documents, they do not provide services like Shepard's or KeyCite, extensive cross-referencing, or subject indexing. In order to aid the development of free but powerful online legal research tools, courts should consider including detailed, structured metadata in each electronically-published court decision. This practice would have three primary benefits: 1) it could spur improvements in Google Scholar, LII, and other online, free databases of legal resources 2) it would give courts greater control over how opinions are interpreted and used, and 3) it could potentially cut the cost of premium research services, by lowering costs, decreasing barriers to entry for this market, and increasing competition.

The America Association of Law Libraries (AALL) has a proven track record of advocating for open access to government information; going forward, the AALL should continue to lobby for the widespread adoption of uniform metadata standards. In particular, AALL's Digital Access to Legal Information Committee, now focused on UELMA and medium neutral citation, could take the lead on this issue. AALL's Government Relations Committee could also play a role, perhaps by educating judges and court systems on the importance of metadata standards, and how detailed tagging of court opinions could help improve public access to case law.

A number of professional and non-profit organizations have taken on the XML cause. LegalXML.org develops uniform metadata tagging languages, which are currently implemented by some court systems. The National Court Reporters Association has taken an interest as well, unsurprising given that court reporters deal with these sorts of technologies first-hand. AALL should cooperate with these and other like-minded organizations to lobby for the expanded use of XML. By teaming up with LegalXML.org, law librarians could get involved in developing and improving the XML standards used in the legal field.