Persistent Identification of Electronic Documents and the Future of Footnotes*

Susan Lyons**

Both the accuracy of scholarly footnotes and the long-term access to digital publications are threatened by link rot. Ms. Lyons discusses one possible solution: widespread use of a system for persistent identification of electronic documents.

¶1 Can the footnote be saved? Over the past decade, the use of Internet citations in the footnotes of law review articles has grown from a trickle to a flood. But it is well documented that Uniform Resource Locators (URLs) experience "link rot," that is, over time the URL is more and more likely to become a dead link, making the footnote citation worthless or nearly so.

¶2 Beyond the question of the accuracy of scholarly footnotes is the issue of long-term access to digital publications. If an electronic resource cited as a reference has moved to another location and still may be found through a Web search engine, then a broken URL is merely an annoyance. But if the cited document has vanished from the Web entirely, a much larger hole is left. As libraries dedicate more of their budgets to electronic resources and as government agencies push to eliminate the distribution of most documents in print, librarians must consider what changes are necessary to ensure preservation and long-term access to electronic resources. This article examines one small piece of the solution: persistent identification of electronic files, especially publications and documents.

¶3 In 1994, there were just four instances of Web citations in three law review articles.[1] By 2003 there were at least 96,946 citations to the Web in law review footnotes. Table 1 charts the explosive growth in Web citations in law review articles over a ten-year period.

** Documents/Reference Librarian, Rutgers University Law School Library, Newark, New Jersey.

1. The three pioneering articles that included Web citations were a student note, an article authored by four professors in a symposium on intellectual property, and an essay by a district court judge. Judge Jack Weinstein, author of the essay, offered readers the most help as they ventured into the uncharted territory of the Web. He provided a choice of both an FTP site and a Web site and included instructions on what to type or click once the reader reached the sites. Jack B. Weinstein, *Limits on Judges' Learning, Speaking, and Acting: Part II Speaking and Part III Acting*, 20 U. DAYTON L. REV. 1, 3 n.5 (1994). Two Web references appear in the symposium article. Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308, 2363 n.221, 2364 n.223 (1994). The student note included one Web citation. David J. Pascuzzi, Note, *International Trade and Foreign Investment in Colombia: A Sound Economic Policy Amidst Crisis*, 9 FLA. J. INT'L L. 443, 449 n.50 (1994).

Table 1

Number of Law Review Articles with Web URLs in Footnotes by Domain Name, 1994–2003[2]

Domain	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total
.com	1	22	182	558	1,098	1,783	2,483	4,588	21,619	2,4576	56,910
.edu	3	23	169	387	666	943	1,339	2,380	11,337	13,593	30,840
.gov	0	14	141	475	1,031	1,789	2,426	4,106	19,935	22,811	52,728
.mil	0	0	8	18	34	72	119	160	825	1,168	2,404
.net	0	11	77	202	297	431	623	1,114	5,190	5,986	13,931
.org	0	17	158	651	1,325	2,227	4,297	5,433	26,198	28,812	69,118
TOTAL	4	87	735	2,291	4,451	7,245	11,287	17,781	85,104	96,946	225,931

The Problem of Link Rot

¶4 If Web citations could be relied upon to point to the cited information over a long period of time, their exponential increase in usage would be of little concern, but that is not the case. The short life span of URLs—most often labeled link rot—has been documented in several articles. In his 1999 study of Web page mortality, Wallace Koehler describes Web pages as "an intermediate form between recorded and unrecorded communications."[3] Koehler found Web pages ephemeral in two respects: they disappear and they morph into new versions. Tracking 361 Web sites in his study, Koehler found 17.7% of Web sites and 31.8% of Web pages had disappeared after one year.[4] Perhaps more disturbing was the lack of constancy among the surviving sites: after one year more than 99% of all the Web pages recorded some change.[5]

¶5 The ephemeral nature of Web pages is a problem that plagues scholarly articles published in many disciplines, including law. A study done in 2000 examined URLs in the footnotes of thirty-one academic journals over a three-

2. The numbers were compiled by running searches for each domain in the LexisNexis US Law Reviews and Journals Combined database. The searches were run in December 2004. For the years 1994 through 1999, the search was run in one-year segments for each domain. In the year 2000, I first encountered the 3000 hit limit for the .org domain and consequently divided that search into six-month segments. For the year 2001, the search was run in six-month segments; for the years 2002 and 2003, in monthly segments.

   A typical search was: **footnote-1 ((http or www) /s com) and date(geq (12/01/02) and leq (12/31/02))**. This search query underestimates the total number of URLs in footnotes as it records only hits that contain at least one URL with that domain, not the total number of instances of URLs with that domain extension in each article. There were at least 96,946 instances of footnotes that cited a Web URL in 2003, but the true number of citations to the Web is likely much higher. It should also be noted that the yearly totals in 2002 and 2003 might reflect a number greater than the number of published articles. Many articles include Web citations in more than one domain category.
3. Wallace Koehler, *An Analysis of Web Page and Web Site Constancy and Permanence*, 50 J. Am. Info. Sci. 161, 163. (1999).
4. *Id*. at 179.
5. *Id.*

year period and found that nearly half the links were dead at the end of three years.[6] A more recent study examined footnotes in three of the leading scientific journals, *Science, The Journal of the American Medical Association (JAMA),* and *The New England Journal of Medicine.* It found that after three years, 13% of all Web links cited in the journals were dead, with the highest morality rates in the **.com** domain (46%) and **.edu** domain (30%).[7] A broader study of the persistence of scientific Web links tracked 515 links used in a graduate biochemistry distance learning course and found that 27.5% of them were broken after twenty-four months.[8]

¶6 Three articles published in 2002 documented the impermanence of Web links in law review footnotes. A study by Mary Rumsey found severe link rot. Examining hundreds of law review footnotes in journals published over a five-year period, she found nearly 70% of all links cited in journals published in 1997 were dead, as were more than 38% of links cited in journals published in 2001, the year before the study.[9] Rumsey also noted that contrary to *Bluebook* rule 18.2 on the use of Internet citations,[10] many authors were using an Internet citation not as a parallel citation, but as the only citation to a source, even when the document had a well-known print analog.[11]

¶7 E. Dana Neacsu also assessed the persistence of Web links in law review articles, finding that twelve of twenty articles examined contained footnotes with broken URLs.[12] Neacsu also noted that while the then new *Bluebook* Rule 18.2 on citing Internet resources discouraged the use of stand-alone Internet citations, "one of its unintended consequences has been to encourage or accept citation to electronic media of suspect reliability." [13] The figures in table 1 show that the greatest increase in the use of Web citations in law reviews occurred shortly after Rule 18.2 was adopted in 2000, with a nearly fivefold increase from 2001 to 2002. Whether this increase is attributable to the new rule, the overall growth of the Web, or some other factor is impossible to say, but it is an interesting observation.

¶8 A study by Simon Canick of citations to a variety of online resources in law review articles found that citations to proprietary databases such as Westlaw and

6. Carol Anne Germain, *URLs: Uniform Resource Locators or Unreliable Resource Locators*, 61 C. & Res. Libr. 359, 362 (2000).
7. Robert Dellavalle et al., *Going, Going, Gone: Lost Internet References*, 302 Sci. 787, 787 (2003).
8. John Markwell & David Brooks, *"Link Rot" Limits the Usefulness of Web-based Educational Materials in Biochemistry and Molecular Biology*, 31 Biochemistry & Molecular Biology Educ. 69, 70 (2003), *available at* http://www.bambed.org/cgi/reprint/31/1/69.
9. Mary Rumsey, *Runaway Train: Problems of Permanence, Accessibility, and Stability in the Use of Web Sources in Law Review Citations*, 94 Law Libr. J. 27, 35, 2002 Law Libr. J. 2, ¶ 23 tbl. 1.
10. The Bluebook: A Uniform System of Citation R. 18.2, at 132 (Columbia Law Review Ass'n et al. eds., 17th ed. 2000).
11. Rumsey, *supra* note 9, at 33–34, ¶ 20.
12. E. Dana Neacsu, *Legal Scholarship and Digital Publishing: Has Anything Changed in the Way We Do Legal Research?* Legal Reference Services Q., 2002 no. 2/3, at 105, 113. [DOI: 10.1300/J113v21n02_06].
13. *Id*. at 112.

LexisNexis and other fee-based digital libraries fare better in terms of persistence over time than do citations to resources freely available on the Web.[14]

¶9 Today the scholarly literature of law, medicine, science, and the humanities rests on a foundation of footnotes riddled with broken URLs. Does it matter that citations to online resources are less reliable? Footnotes do not enjoy universal esteem, especially in literary writing where they may seem an annoying interruption to the flow of the text. Noel Coward once remarked that encountering a footnote "is like going downstairs to answer the doorbell while making love."[15] Yet footnotes serve a certain utility in much of scholarly writing, especially in the sciences and social sciences. Footnotes acknowledge an author's debt to others, serve as portals to readers wishing to explore a tangential point in greater detail, and in scientific literature provide a research trail that others may build upon. In legal writing, hardly a thought can be had or a sentence written without attribution to some greater authority.

¶10 Michael Bugeja, a professor of journalism and author of a study on link rot, has noted that "[t]he footnote is basic to research. If we cannot rely on a footnote because the medium is too dynamic, then Internet scholarship will always be a second-class citizen in academe."[16] The problem is not just that footnotes to electronic references may prove short-lived, but that the underlying documents to which they point are disappearing. It is not a question of trusting that the author did indeed find the cited material when an article was written.[17] As readers we may value the sources more highly than the commentary.[18] An article with dead sources is a dead end.

Persistent Identification of Web Documents

¶11 The ephemeral nature of Internet resources threatens to undermine the authority of the scholarly research, yet there is no going back to a print-only world. What solutions are available to preserve the integrity of footnotes citing to online publications? Some law reviews print out all of the Internet documents referenced in an article's footnotes and maintain them in a file for a period of time. This practice is useful for fixing the reference in a stable medium, but as the number of Internet

14. Simon Canick, *Availability of Works Cited in Recent Law Review Articles on LEXIS, Westlaw, the Internet, and Other Databases*, Legal Reference Services Q., 2002 no. 2/3, at 55, 58–59. [DOI: 10.1300/J113v21n02_03].
15. G.W. Bowersock, *The Art of the Footnote*, 53 Am. Scholar 54, 54 (1983–84).
16. *Quoted in* Scott Carlson, *Here Today, Gone Tomorrow: Studying How Online Footnotes Vanish,* Chron. Higher Educ., Apr. 30, 2004, at A33.
17. Describing a critic who had the temerity to question the integrity of his footnotes, historian Edward Gibbon wrote: "If he will take the trouble of calling at my house any afternoon when I am *not* at home, my servant shall show him my library, which he will find tolerably well furnished with the useful authors, ancient as well as modern, ecclesiastical as well as profane, who have directly supplied me with the materials of my History." *Quoted in* Bowerstock, *supra* note 14, at 57. Many a modern author could point only to a library of broken URLs.
18. This is particularly true in legal research for litigation where law review articles are valued chiefly as a handy cumulation of primary sources on a subject under litigation.

references grows, maintaining archives of Web printouts may prove difficult for student-run publications.

¶12 A better solution is to develop an identification scheme that points to the document and not its transient location on a Web server. Persistent identifiers are becoming increasingly available for a significant percentage of Web documents cited in law review footnotes, especially government documents. They apply unique identifiers, similar to bar codes or ISBNs, to digital objects. Persistent identifiers can offer a long-term solution to the problem of link rot *if* they are widely adopted. Today citation to persistent identifiers in footnotes is extremely rare,[19] but this is hardly surprising since they are both new and unfamiliar. As the problem of link rot worsens, however, and authors, editors, and librarians become more familiar with persistent identifiers, usage may increase.

¶13 The concept of persistent identifiers for electronic documents should be familiar to users of Westlaw and LexisNexis. Anyone who has retrieved an unpublished case from Westlaw or LexisNexis has encountered a persistent identifier in the form of a Westlaw or LexisNexis number. Commercial publishers of electronic documents understand the value of persistent identifiers in managing valuable intellectual property. Academic publishers and government agencies are now embracing persistent identifiers to manage documents available on the Web, and within a few years it may be possible to find persistent identifiers for a majority of electronic references cited in law review articles. At present there are several models of persistent identifiers and it is unclear which model will win out.

URIs, URLs, and URNs

¶14 The current pattern of describing identifiers for Web files was worked out in 1992 at a meeting of the Internet Engineering Task Force (IETF), a group that develops standards for the Internet. Tim Berners-Lee, the creator of HTML and the World Wide Web,[20] proposed using the term Universal Document Identifier, a label that would emphasize the importance of universality and persistence of information. When that term was rejected, he proposed Uniform Resource Identifiers. Berners-Lee describes the conflict that arose over the proper term:

> [The IETF] decided that *identifier* wasn't a good label for what the Web used. They wanted to emphasize that people could change the URIs when moving documents and so they should be treated as some sort of transitive address. *Locator* was chosen instead, like a branding, a warning mark on the technology.[21]

---

19. A search in the US Combined Law Reviews and Journals database of LexisNexis on Mar. 30, 2005, uncovered only nineteen instances of citation to PURLs in footnotes, seven citations to DOIs, and no instances of handles or ARKs. Most of the PURL citations were to United States government publications, while all but one of the DOIs were to scientific journals.
20. Tim Berners-Lee, Bio, *at* http://www.w3.org/People/Berners-Lee/#Bio (last visited June 8, 2005).
21. Tim Berners-Lee, Weaving the Web 62 (1999).

¶15 While the Uniform Resource Locator (URL) won out, Uniform Resource Identifier (URI) survives as a broader term for identifiers that can describe either URLs or Uniform Resource Names (URNs), a more persistent identifier. URIs are alphanumeric strings that identify resources on the Web: documents, images, downloadable files, and other resources. The URL is the most familiar type of URI and works by pointing to a digital object's physical location on an Internet server. But electronic resources often move without leaving a forwarding address. The inadequacy of URLs was soon recognized, and in 1994 the IETF formed a working group to develop URNs, a protocol that would uniquely identify digital objects (such as a document or image) regardless of their location and would persist over time.[22]

¶16 URNs are clearly preferable for ensuring long-term access to Web documents, and yet URLs dominate the Web. Why is this so? Organizing and maintaining a system of persistent identifiers requires considerable time and money. Such effort seemed unnecessary in the mid-1990s when many institutional Web sites consisted of only a few hundred pages. As institutional Web sites grew in size and complexity, Web site administrators added new directories, often breaking hundreds of links with each reorganization. For large institutions managing hundreds of thousands of electronic documents and images, persistent identifiers are now worth the cost. Academic publishers, government agencies, and national libraries are among the institutions that have embraced persistent identifiers for electronic documents.

¶17 Persistent identifiers require some centralized institution to curate the electronic files, and this means that many files on the Internet will never get tagged with persistent identifiers. For example, the millions of personal blogs[23] and Web pages that exist are unlikely to receive persistent identifiers.

¶18 A significant number of the URLs cited in law review footnotes point to government documents in both the United States and Europe.[24] The organizations that curate these documents are now in the process of adopting standards for persistent identification and permanent public access.[25] Many of the remaining electronic references cite to the **.org** and **.edu** domains and include reports, white papers, articles, and other documents published by foundations and nonprofit organizations. At least some of these organizations are likely to possess the resources to assign persistent identifiers and manage their electronic publications so as to provide for long-term access.

---

22. *See* Internet Engineering Task Force, Uniform Resource Names (urn), *at* http://www.ietf.org/html.charters/OLD/urn-charter.html (last modified July 31, 2001).
23. As blogs become more specialized, citations to blogs are making their way into law review footnotes. A search on LexisNexis in March 2005 found 121 instances of the words "blog" or "weblog" in law review footnotes over the last five years. Two-thirds of the citations were from the preceding year.
24. *See supra* Table 1. Almost 25% of the URLs in law review footnotes are from .gov or .mil domains.
25. Robin Wilson, *DOI: A Bar Code for Digital Information*, INFO. ECON. J., June 2004, at 38, 39.

The Handle System and Digital Object Identifiers

¶19 One of the first implementations of URNs was the Handle System developed by the Corporation for National Research Initiatives (CNRI), a research consortium of United States government agencies and academic institutions.[26] The Handle System is a registered trademark of CNRI. The first implementation of the system was developed in 1994 with a grant from the Defense Advanced Research Projects Agency.[27] Early adopters of the system include the Library of Congress, the Defense Technical Information Center, and the International DOI Foundation (IDF).[28] Every handle assigned is globally unique within the Handle System. The syntax of a handle consists of a prefix that identifies the naming authority and a suffix that identifies an individual document. The two are separated by a backslash. A typical handle looks something like this: **10.1234/document56789**, where **10.1234** represents an organization that has the authority to name documents in the Handle System and **document56789** is the suffix that represents a specific document within that organization. The suffix may consist of some previously developed naming scheme, such as an ISSN or bar code, enabling organizations to easily convert items in a database or catalog into handles or DOIs.

¶20 The broadest implementation of the Handle System has been through Digital Object Identifiers or DOIs, a schema originally developed by the Association of American Publishers but now administered by the IDF.[29] As of 2004, more than seven hundred publishers were using DOIs to identify electronic articles.[30] The DOI system uses the same syntax as the Handle System and is fully compatible with it. DOIs also add metadata that permit publishers to describe the content of electronic files and control access to the files. The publishing industry has established an organization called CrossRef as a centralized agency to register and resolve DOIs. To date more than 16.2 million DOIs have been registered with CrossRef and more are being added every day.[31] Publishers using DOIs to identify their electronic documents include the American Medical Association, the American Psychological Association, Blackwell, EBSCO, Elsevier, Haworth, Kluwer, MIT Press, Nature, Sage, Springer, and Wiley.[32]

---

26. For more information about CNRI, see Corporation for National Research Initiatives, *at* http://www.cnri.reston.va.us (last visited June 8, 2005). CNRI publishes *D-Lib Magazine*, an online journal (www.dlib.org) about digital libraries.
27. Int'l DOI Found., DOI Handbook, App. 2 (2005), *at* http://www.doi.org/handbook_2000/appendix_2.html.
28. *Id*.
29. For more information about the IDF, see International DOI Foundation, *at* http://www.doi.org (last visited June 8, 2005).
30. *CrossRef Adds New Services, Expands to New Content Areas*, Info. Today, Dec. 2004, at 33, 33.
31. *See* CrossRef Registered DOIs counter, *at* http://www.crossref.org (last visited June 16, 2005).
32. *See* Publishers & Societies, *at* http://www.crossref.org/01company/06publishers.html (last visited June 8, 2005).

¶21 Handles, DOIs, and other persistent identifiers require centralized registration and resolution systems, similar to the Domain Name System used to resolve URLs on the Internet. Current Web browsers are not yet able to resolve DOIs without a special plug-in program. The IDF maintains a DOI resolution Web page (http://dx.doi.org) where users can type in a DOI and be linked to the electronic file it represents.

¶22 Another technology that is used to allow DOIs to work within standard browsers is the OpenURL framework, a standard for transporting metadata, identifiers about a digital object, or both. The OpenURL framework was designed to provide a context-sensitive method of resolving requests for an object where multiple authoritative copies exist. For example, a patron working at a networked computer terminal in the library clicks on a link for a journal article. Identical copies of that article are available through Ebsco, ProQuest, and Science Direct, but the library subscribes to only one of those databases. Through the OpenURL framework,[33] the resolver will recognize the user's access privileges and direct the patron to the appropriate copy. The OpenURL framework is a proposed ANSI/NISO standard (Z39.88 200X) and was quickly embraced by scholarly publishers as a way of enhancing the functionality of DOIs. Publishers embed DOIs within an OpenURL, enabling the DOI to be retrieved using standard Hypertext Transfer Protocol (HTTP).

¶23 To take full advantage of OpenURLs and DOIs, another layer of software is necessary: OpenURL link resolvers. These software packages are produced by a number of vendors including EBSCO, Endeavor, Ex Libris, Sirsi, and Ovid Technologies. While link resolver software is not essential to access DOIs, it greatly facilitates patrons' access to online publications in libraries that subscribe to many online full-text databases. The link resolver programs allow authorized users to seamlessly link to the full text of articles from online indexes and databases or even from a general Web site on the Internet if the user is logged in to the library computer network.[34]

¶24 DOIs have provided the publishing industry with a way of managing their intellectual property and limiting access of their content to subscribers. The usage of DOIs, however, need not be limited to commercial publishers nor is there any reason to restrict its usage to subscription-based systems. The Office for Official Publications of the European Communities (OPOCE) is adopting DOIs, as is the Organization for Economic Co-Operation and Development (OECD).[35] In the United Kingdom, the Stationary Office (TSO), a private publisher licensed to

33. For more information on the OpenURL framework, including links to the standard and to the KEV Implementation Guidelines, see The OpenURL Framework Standard, *at* http://library.caltech.edu/openurl/Standard.htm (last visited June 16, 2005).
34. For a discussion of link resolver software and a comparison of the various products, see generally Christine L. Ferguson & Jill E. Grogg, *OpenURL Link Resolvers,* 24 COMPUTERS LIBR., Oct. 2004, at 17.
35. Julien Perkin, *Avoiding Identity Crises: Digital Standards*, FIN. TIMES (London), Dec. 1, 2004, at 5.

print the official publications of Parliament and other government agencies, is using DOIs and has applied them to all items within the *United Kingdom Official Publications* catalog.[36]

¶25 In the United States, several government agencies have adopted handles. The Defense Technical Information Center assigns handles to digital files managed in its Scientific and Technical Network, a digital library.[37] The Library of Congress has assigned more than 400,000 handles since 1995 and uses them to identify documents within its American Memory Collection.[38] The National Agriculture Library is also using the Handle System.[39]

PURLs

¶26 The biggest drawback of the Handle System and DOIs is that they are not yet fully compatible with standard Web browsers, unless they are coupled with an OpenURL. A DOI or handle, entered into the location bar of Microsoft's Internet Explorer or Netscape, will not resolve to the desired digital object unless the browser has been upgraded with a plug-in. As handles and DOIs gain in acceptance, it is likely that future Web browsers will be fully compatible with these protocols. But the difficulty in achieving compatibility with present Web browsers has led some government agencies to choose another method of persistent identification: PURLs.

¶27 Persistent Uniform Resource Locators (PURLs) were developed by OCLC in 1995 to encourage libraries to include hyperlinks in their online catalogs. As OPAC software developed the capability to include working hyperlinks in the MARC 856 field as part of a holding's bibliographic record, there was concern that the hyperlinks would rapidly become dead links. PURLs offer the promise of a hyperlink of sufficient stability so as to be worthy of inclusion in an online catalog.[40]

¶28 OCLC defines a PURL as follows:

> Functionally, a PURL is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. The client can then complete the URL transaction in the normal fashion. In Web parlance, this is a standard HTTP "redirect."[41]

---

36. Wilson, *supra* note 25, at 38–39.
37. CENDI Persistent Identification Task Force, Persistent Identification: A Key Component of an E-Government Infrastructure 4 (Mar. 10, 2004), *available at* http://cendi.dtic.mil/publications/04-2persist_id.pdf.
38. *Id*. at 5.
39. *Id*.
40. For an early discussion of the problem of broken URLs and the use of PURLs in the 856 field as a possible solution to avoiding cataloging unstable URLs, see Libr. of Congress, Defining a Uniform Resource Name Field in the USMARC Bibliographic Format (Discussion Paper 96, May 6, 1996), *at* http://www.loc.gov/marc/marbi/dp/dp96.html.
41. Stuart Weibel et al., PURLs: Persistent Uniform Resource Locators § 3.0, *at* http://purl.oclc.org/docs/new_purl_summary.html (last visited June 8, 2005).

Like handles and DOIs, PURLs are a high-maintenance solution to the problem of link rot. Someone must set up a centralized PURL server and monitor all of the documents to which the PURLs point. If a link breaks, the PURL must be updated to provide a redirect to the document's new location.

¶29 The largest user of PURLs is the United States Government Printing Office (GPO), which began cataloging online documents in 1997 and assigning PURLs to those documents in 1998. GPO catalogers have gone back and systematically replaced URLs with PURLs for those documents cataloged in 1997. At the end of 2004, GPO had assigned approximately 57,000 PURLs to documents listed in the *Catalog of Government Publications*.[42] GPO is adding to that total at the rate of about 14,000 to 15,000 each year.[43] The National Library of Australia and the South Dakota Library Network also use PURLs.

¶30 A great advantage of PURLs over handles and DOIs is that, as functional URLs, they are fully compatible with current Web browsers and require no special plug-ins or secondary resolution service. This is an important consideration for GPO as it is working to move from being a distributor of print documents to a repository of electronic documents and must use a technology that makes documents immediately accessible to its users. While much remains to be done before GPO can guarantee the permanence and authenticity of electronic documents, assigning permanent identifiers to all documents is an important first step.

¶31 One criticism of PURLs is that they are a stopgap measure and not a true URN. At some point down the road newer transfer protocols may supersede the familiar Hypertext Transfer Protocol (HTTP), and PURLs may not be flexible enough to make the migration to the new Internet protocol. But it is likely that if some other schema becomes the predominant model for persistent identifiers (DOIs, for example), the PURLs assigned by GPO could easily be converted and migrated to the new format. An example of a PURL is **http://purl.access.gpo.gov/GPO/LPS2251**. To convert that into a DOI, it is only necessary to convert the first part of the address into a DOI prefix and the second part (GPO/LPS2251) into the DOI suffix. In this way the effort put into cataloging government documents with PURLs would be conserved.

¶32 Another criticism of PURLs is that they sometimes break, just as URLs do, but PURL breaks are usually brief and temporary. To maintain the integrity of the PURL redirects, GPO runs a PURL validation report every weekend to ferret out broken links. A GPO specialist repairs the broken PURLs the following week.[44]

---

42. E-mail from Cynthia Etkin, Senior Program Analyst, U.S. Government Printing Office, to Susan Lyons, Documents/Reference Librarian, Rutgers University Law School Library (Dec. 29, 2004) (on file with author).
43. *See* Library Programs Service, FY 2002 Annual Report (2002), *reprinted in* Admin. Notes, Oct. 15, 2002, at 6, 11, *available at* http://www.access.gpo.gov/su_docs/fdlp/pubs/annrprt/02lpsar.html [http://purl.access.gpo.gov/GPO/LPS2251].
44. Posting of Cynthia Etkin, Senior Program Analyst, U.S. Government Printing Office, to GOVDOC-L@lists.psu.edu, Broken PURLs (May 27, 2004), *available at* http://lists1.cac.psu.edu/cgi-bin/wa?A2=ind0405d&L=govdoc-l&T=0&O=A&P=5786.

As government agencies are continually modifying their Web sites, it is always possible to hit a PURL on a bad day, that is, between the time the link breaks and it is detected and repaired. In the past a more serious problem occurred when an agency did not simply reshuffle a document to a new directory but removed the document from its Web site. To combat the problem of disappearing documents, GPO has established its own electronic archive for those documents it catalogs. If an agency completely removes a document from its Web site, the PURL redirects to the copy in GPO's electronic archive.[45]

¶33 Another problem is that many electronic government documents are fugitive documents that escape notice and cataloging by the GPO. Librarians from both the American Library Association's Government Documents Roundtable and the American Association of Law Libraries' Government Documents Special Interest Section are working with GPO to identify fugitive documents as part of the Electronic Documents Working Group. Members of this group monitor specific agency Web sites and notify GPO of new electronic publications that are not cataloged.[46] GPO has discussed plans to automate the capture of fugitive documents, especially "born digital" documents, by using Web harvesting software to locate new documents on agency Web sites, but this process has not yet been fully implemented.[47]

¶34 Are PURLs in fact more persistent than URLs when used in scholarly footnotes? Unfortunately they are so rarely used it is difficult to know. A search of the LexisNexis US Law Reviews and Journals Combined database found only nineteen instances of PURLs used in footnotes. Two of the PURLs linked to an OCLC PURL server, and twelve were maintained by GPO and linked to government documents. All of the PURL links were active and pointed to documents that were as old as five years.[48]

¶35 Finally, while PURLs have been adopted by GPO, other United States government agencies have pursued other schemas for persistent identification. Handles and DOIs are used by the Department of Defense and Library of Congress,[49] while the National Library of Medicine has developed another scheme for persistent

45. *See* U.S. Gov't Printing Office, Federal Depository Library Program Electronic Collection (FDLP/EC) Archive, *at* http://www.access.gpo.gov/su_docs/fdlp/ec/ecarc.html (last visited June 8, 2005). The GPO archive, once fully implemented, may address the technical problem of preserving electronic documents but not the issue of documents removed from government Web sites because of security concerns. The GPO reported that in the wake of 9/11 many once public documents were permanently removed, including approximately 6000 documents from the Department of Energy Web site. LIBRARY PROGRAMS SERVICE, *supra* note 43, *reprinted in* ADMIN. NOTES, Oct. 15, 2002, at 6, 13.

46. *See* U.S. Gov't Printing Office, Electronic Documents Working Group, *at* http://www.access.gpo.gov/su_docs/fdlp/tools/ewgroup/index.html (last visited June 8, 2005).

47. GPO issued a solicitation to vendors for Web harvesting software on Dec. 30, 2004. It plans to begin with a pilot project of harvesting items at a Web site of the Environmental Protection Agency, eventually expanding to cover most federal Web sites. Posting of Judith C. Russell, Superintendent of Documents, to fdlp-l@gpo.gov, GPO Solicitation for Web Discovery and Harvesting (Jan. 7, 2005), *available at* http://listserv.access.gpo.gov/scripts/wa.exe?A2=ind0501&L=gpo-fdlp-l&F=&S=&P=70.

48. Two of the hyperlinks in the LexisNexis documents failed to work because of added spaces in the online text but the links worked when retyped in the browser.

49. CENDI PERSISTENT IDENTIFICATION TASK FORCE, *supra* note 37, at 5.

identification: Archival Resource Keys (ARKs).[50] This method of persistent identification works within standard browsers. The first part of the ARK is a standard URL that links to the server of the institution hosting the ARK. This is followed by an alphanumeric string that remains unchanged, even though the item may move to a different server. In addition to conveying metadata about the digital object, ARKs also contain a commitment statement from the host organization indicating the object's level of permanence. ARKs are also being used and developed by the California Digital Library, a project of the University of California.[51]

Conclusion

¶36 While persistent identifiers have not been widely used by authors, they offer great potential. There have been at least 55,000 citations to documents in the **.gov** and **.mil** domains in law review articles over the last decade, nearly a quarter of the total citations, and documents in both domains are cataloged by GPO and assigned PURLs. Until such time as persistent identifiers become ubiquitous and widely known, it is unlikely that many authors will cite to them. Yet there is an opportunity to increase the use of persistent identifiers in student-edited law reviews today. Many of the electronic URLs cited by authors are dead by the time an article is being reviewed for publication, and student editors are required to track down active links for electronic citations. Librarians can assist student editors to find persistent identifiers where they are available. Government documents to which PURLs have been assigned can be located by pasting the URL into a PURL search form available at the GPO Web site (www.access.gpo.gov/su_docs/fdlp/tools/purlsear.html). Where footnotes cite articles available through electronic databases, inclusion of a DOI, in addition to a full bibliographic citation, will aid future researchers. Publishers that use DOIs generally include the DOI on the first page of the article or in the bibliographic summary at the beginning of the article.

¶37 The problem of link rot will likely get worse before it gets better. Persistent identification schemas offer the potential of providing stable and persistent links to large bodies of literature cited in scholarly publications. In the next few years one method of persistent identification may win out and become widely adopted by all curators of large repositories of electronic files, including government printing agencies, commercial publishers, universities, and other organizations that maintain digital libraries. At the point when a single standard becomes widely accepted, authors will include persistent identifiers in their footnotes.

¶38 In the interim, law librarians can encourage greater use of persistent identifiers by assisting student editors in finding PURLs and DOIs for the electronic sources cited in law review articles.

---

50. Nat'l Library of Austl., Persistent Identification Systems § 1.4.5, *at* http://www.nla.gov.au/initiatives/persistence/PIpart1.html (last visited June 8, 2005).
51. For a description of the ARK schema, see generally J. Kunze & R.P.C. Rodgers, The ARK Persistent Identifier Scheme (Feb. 19, 2005), *at* http://www.cdlib.org/inside/diglib/ark/arkspec.pdf.

Appendix
Glossary of Terms and Acronyms

ARKs—Archival Resource Keys—a persistent identification scheme developed by the National Library of Medicine and also used by the California Digital Library.

CNRI—Corporation for National Research Initiatives—A government and academic research consortium that established the Handle System.

CrossRef—An organization established to provide centralized registration and resolution services for DOIs. CrossRef primarily services the online publishing industry.

DOIs—Digital Object Identifiers—An application of the Handle System, widely used by commercial and academic publishers.

Handle System—A registered trademark of CNRI that describes an open source method of applying persistent identifiers to electronic files. DOIs build upon the Handle System.

IDF—International DOI Foundation—An open membership consortium that manages DOI registration. It includes both commercial and noncommercial partners.

IETF—Internet Engineering Task Force—A group that develops standards and protocols for the Internet. The IETF is responsible for the URL standard and has established a working group to develop standards for URNs.

OpenURL Framework—A syntax to create Web-transportable packages of metadata, identifiers about an information object, or both. Such packages are necessary to enable context-sensitive or open link technology.

PURL—Persistent Uniform Resource Locator—Developed by OCLC, PURLs provide a persistent way of identifying and locating electronic documents using the redirect feature built into the HTTP protocol. It requires that the curating institution maintain its own PURL server and update it regularly. The United States Government Printing Office is the largest user of PURLs.

URI—Uniform Resource Identifiers—Alphanumeric strings that identify electronic resources on the Internet. URLs and URNs are both types of URIs.

URL—Uniform Resource Locators—A URI that identifies an electronic file by its location on a particular Web server. While easy to establish, URLs fail if the file is moved to a new directory or server, resulting in the familiar "404 Not Found" message. Tim Berners-Lee defines a URL as "[a] term sometimes used for certain URIs to indicate they might change."[52]

---

52. BERNERS-LEE, *supra* note 21, at 218.

URN—Uniform Resource Name—A URI that would identify a particular file regardless of its location. There is no clear standard for a URN. ARKs, handles, and DOIs are types of URNs.