The Practice of Informatics

JAMIA

Technology Brief ■

The Shadow Uniform Resource Locator:

Standardizing Citations of Electronically Published Materials

JOSEPH V. DICARLO, MD, XAVIER PASTOR, MD, BARRY P. MARKOVITZ, MD

Abstract Citation of scientific materials published on the Internet is often cumbersome because of unwieldy uniform resource locators (URLs). The authors describe a format for URLs that simplifies citation of scholarly materials. Its use depends on a simple HTML device, the "refresh page." Uniform citation would follow this format: [Author I. Title of article. http://domain/year/month-day(e#).html]. The HTML code for such a page is: ⟨HTML⟩ ⟨head⟩ ⟨meta HTTP-EQUIV="Refresh" CONTENT="0; URL= http://Actual-URL/ for-article/ referred-to/ in-citation.html"⟩ ⟨/head⟩ ⟨/HTML⟩. The code instructs the browser to suppress the content of the refresh page and bring up the title page of the cited article instead. Citations would be succinct and predictable. An electronic journal would not need to alter its existing file hierarchy but would need to establish a distinct domain name and maintain a file of refresh pages. Utilization of the "shadow" URL would bring us one step closer to truly universal resource locators.

■ **JAMIA.** 2000;7:149–151.

Citation of scientific materials published electronically (i.e., on the World Wide Web) is often cumbersome, in large part because of the inclusion of the often unwieldy uniform resource locator (URL) or address. None of the published citation guidelines[1–4] have suggested the simplification and standardization of the URLs of scientific and scholarly materials on the Web. Moreover, the most recent recommendations put forth by the International Standard Organization still support a citation approach that includes complete URLs, no matter the length.[4] We describe a straightforward and useful URL format for scientific and scholarly materials that will not require the revamping of a Web site's file hierarchy or a publication's pagination format. This URL format would greatly simplify the task of locating electronically published material. It can be easily applied retroactively to existing materials. Its use depends on a simple HTML (hypertext markup language) device, the "refresh page."

Affiliations of the authors: Stanford University Medical School, Stanford, California (JVD); Hospital Clinic, University of Barcelona, Barcelona, Spain (XP); Washington University School of Medicine, St. Louis, Missouri (BPM).

Correspondence and reprints: Joseph DiCarlo, MD, Division of Pediatric Critical Care Medicine, 750 Welch Road #315, Stanford, CA 94304; e-mail: ⟨jdicarlo@stanford.edu⟩.

Received for publication: 10/4/99; accepted for publication: 11/15/99.

Background

The URL indicates the exact location on a host computer of materials posted or published on the Internet. An Internet browser uses a given URL to bring the desired material to the screen. Citation of an online scientific resource most often includes the full URL for

the given abstract, full-text article, or chapter. Although a user can type a URL directly into a browser's location box, some URLs are so long (consisting of more than 100 characters) that this task is impractical. Alternatively, a short URL can be used to locate the home page of a Web site (e.g., for an online journal), from which the reader searches for the desired article. Citation by either of these methods is of limited usefulness to the reader. For example, the following citation was included in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (1997)[5] to illustrate a recommended citation format:

> Morse SS. Factors in the emergence of infectious diseases. Emerg Infect Dis [serial online] 1995 Jan–Mar [cited 1996 Jun 5]; 1(1): [24 screens]. Available from: URL: http://www.cdc.gov/ncidod/EID/eid.htm.

The listed URL actually brings one to the home page of the journal. From there the site's search function must be used (e.g., entering ''Morse SS,'' which yields a list of 17 documents). A link listed on the search results page provides access to the desired article, at http:/www.cdc.gov/ncidod/eid/vol1no1/morse.htm.

By applying the same recommendations to a recently published article in an electronic journal, the following citation is produced:

> Logigian EL, Kaplan RF, Steere AC. Successful treatment of Lyme encephalopathy with intravenous ceftriaxone. Journal of Infectious Diseases [serial online]. 1999 Aug [cited 1999 Aug 9];180(2). Available from: URL: http://www.journals.uchicago.edu/JID/journal/issues/v180n2/980614/980614.html

The URL is long and complicated and follows a logic unique to the journal.

To date, there has been no attempt to simplify and standardize the URLs of scientific and scholarly materials on the Web. It is unlikely that there ever will be standardization of URLs (or, therefore, of Web site file hierarchies), since extensive revision of existing sites would be required. However, standardized ''shadow'' URLs could be assigned to the title pages of existing and future materials without much difficulty or expense.

Proposed Format

We propose a standardized format for the URLs included in the citations of scholarly and scientific material published on the Web, as follows:

> Author I. Title of article. http://domain/year/month-day(e#).html

The concept is illustrated in this working example:

> Goldstein B. Inflicted head trauma. http://picuBOOK.net/1999/05-26(e1).html

in which ''e1'' signifies that this was the first electronic item or article posted by that electronic publication on May 26, 1999. The URL actually calls up a very brief HTML document, a ''refresh page,'' that seamlessly brings the reader to the cited scientific material. This format requires the following:

- *A distinct domain name* for the online journal or resource. Registration of a domain name, currently $70 for two years, [6] is not expensive. Ideally, each publication would have a distinct domain name. Alternatively, a journal could be located in a subfolder high in a site's hierarchy. Examples of current online journals and their domain names are *British Medical Journal*, http://www.bmj.com; *Pediatrics*, http://www.pediatrics.org; and *Chest*, http://journals.chestnet.org/chest/.
- *A simple refresh page* for each article or chapter in the electronic journal.

The HTML code for such a page is quite simple.[7,8] For example, the HTML code for the page at http://picuBOOK.net/1999/05-26(e1).html consists of the following code:

```
⟨HTML⟩
⟨head⟩
⟨meta HTTP-EQUIV=''Refresh''
CONTENT=''0;
URL=
http://PedsCCM.wustl.edu/All-
Net/english/neurpage/trauma/inflicthead/
  inflicted.html''⟩
⟨/head⟩
⟨/HTML⟩
```

This code instructs the browser to visually suppress this simple blank page (i.e., display the CONTENT=''0'' seconds) and ''refresh,'' or load, the window with the page at the listed URL. The reader never sees the blank refresh page. He or she sees only the subsequent page containing the journal article.

Advantages

The shadow URL offers a number of advantages over the present nonstandardized practice. With the shadow URL system, citations would be succinct and would appear in a predictable format, and the Web site or electronic publication could still use its established hierarchy of folders. In fact, a site could restructure at any time without its original URLs becoming

obsolete, so long as it maintained a folder of refresh files in a standard position.

With a shadow URL an article can be accessed using the citation alone. For example, the following citation was obtained from PubMed,[9] the search engine at the U.S. National Library of Medicine, for an article published only in an electronic format:

> Furnival RA, Street KA, Schunk JE. Too many pediatric trampoline injuries. Pediatrics. 1999 May; 103(5):e57.

To locate the cited article on the Internet, one must first find the URL for the journal, *Pediatrics*, and then search through its archive. But if a refresh page were posted at http://www.pediatrics.org/1999/05-01(e57).html, the citation could be changed to:

> Furnival RA, Street KA, Schunk JE. Too many pediatric trampoline injuries. http://www.pediatrics.org/1999/05-01(e57).html

and one could easily locate the article on the Web directly from the citation.

Problems

The system we suggest requires that a publication obtain a succinct (and, ideally, recognizable) domain name, preferably similar to and as recognizable as the MEDLINE abbreviation for the journal. This should not be difficult even for publishing houses that maintain multiple journals through a singular portal, since related domain names can be assigned, such as canmedproc.acme.com, amrespj.acme.com, and so on.

Restricted access is still possible with the shadow URL. If restricted access of some kind is to be maintained, the shadow URL can open a title page for the article containing a password prompt or another security check. One minor technical issue should be mentioned: The refresh meta function may paralyze the back button on the browser, so that use of the back button brings one into an endless loop of refresh page referring to title page. This problem is easily obviated by referring to the browser's recent history (by means of the Go menu in Netscape 4.x or Internet Explorer 4.x).

The shadow URL should bring the reader to the original posting, where the conventions for updating electronic material would still be followed. At the bottom of the HTML page, one would find the date the material was first posted (coinciding with the date in the standardized URL) as well as the date it was last updated. If an electronically posted article undergoes extensive revision, it should probably be assigned a new standardized URL (and perhaps deserves a new title).

Finally, if an article is published in both print *and* electronic formats, to which form should the citation refer? Perhaps both should be cited:

> Russo PA, Chartrand LJ, Seidman E. Comparative analysis of serologic screening tests for the initial diagnosis of celiac disease. Pediatrics 1999;104:75–78. http://www.pediatrics.org/1999/07-01(e25).html

The simplified shadow URL does not add much to the length of a citation.

Conclusion

Actual URLs, or Web addresses, for scientific materials will remain complicated and unwieldy. But standardization is possible for citation and search purposes, by means of an exceedingly simple coding device. We suggest that this format be adopted by MEDLINE; electronic journals would be required to obtain a distinct domain name and furnish standardized citation URLs that continue to function even if the Web site is restructured. Utilization of the shadow URL would bring us one step closer to truly universal resource locators.

References ■

1. American Psychological Association. Electronic reference formats recommended by the American Psychological Association [online reference]. Available from http://www.apa.org/journals/webref.html. Accessed Jul 12, 1999.
2. Modern Language Association. MLA style: documenting sources from the World Wide Web. Available from: http://www.mla.org/main_stl.htm. Accessed Aug 9, 1999.
3. Walker JR, Taylor T. The Columbia Guide to Online Style. 1st ed. New York: Columbia University Press, 1998:44–50.
4. ISO/TC 46/SC 9 Secretariat. Excerpts from International Standard ISO 690-2: Information and documentation. Bibliographic references. Part 2: Electronic documents or parts thereof. Available from: http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm. Accessed Jul 20, 1999.
5. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. JAMA. 1997;277:927–34. Also available from: http://www.ama-assn.org/public/journals/jama/sc6336.htm.
6. Network Solutions. Register a web address (domain name). Available from: http://www.networksolutions.com/.
7. Richmond A. HTML's META tag: HTTP-EQUIV Web developer's virtual library [online serial]. Available from: http://WDVL.com/Authoring/HTML/Head/Meta/HTTP.html. Accessed Jul 12, 1999.
8. Clark S. META tag tutorial. WebDeveloper.com [online serial]. Available from: http://www.webdeveloper.com/html/html_metatags_part2.html. Accessed Jul 12, 1999.
9. PubMed: National Library of Medicine search engine. Available from: http://www.ncbi.nlm.nih.gov/PubMed/.