Citations to the Internet

Karen Patrias, for the CSE Style Manual Subcommittee

In the years since the 1994 publication of the sixth edition of the CSE style manual, *Scientific Style and Format (SSF)*, questions from users on how to cite electronic resources, particularly those found on the Internet, have increased steadily. The seventh edition will include complete details on citing such material. Meanwhile, the National Library of Medicine (NLM) has placed on its Web site a guide to citing the Internet (*www.nlm.nih.gov/pubs/formats/ internet .pdf*) that will prove useful to CSE members. This guide is based on a standard published by the International Organization for Standardization (ISO) in 1997 (ISO 690-21) and a working draft being prepared for the National Information Standards Organization (NISO). Because NLM has specifically applied ISO and NISO principles to scientific material, the NLM guide will form the basis of recommendations to be included in the new *SSF*. This article summarizes the main points found there.

When the Internet changed from a means of communication among a small group of scientists to the ubiquitous tool that it is today, the way in which we regard the information to be found on it also changed. It once consisted primarily of informal electronic mail and datasets to be shared among a few colleagues, but it now includes scholarly publications and other documents that are worthy of citation.

Monographs on the Internet were first just mirror images of large printed texts—such as textbooks, manuals, and technical

KAREN PATRIAS *is senior resource specialist, Public Services Division, (US) National Library of Medicine. She presented some of the material in this article at the 2001 CSE annual meeting.*

reports—but increasingly they are being written specifically for the Internet to take full advantage of the ability to hyperlink, to include complex graphics, and to run multimedia, such as film clips and sound. Some producers of Internet monographs welcome comments or expert opinion from readers and incorporate their comments into the text. Major revisions may be announced as new editions, but minor additions and changes—such as those to add comments, to correct typographic errors, or to update hypertext links—may be transparent to the user.

As with monographs, serials on the Internet may be electronic versions of printed serials or may be created expressly for the Internet. An Internet version may be the equivalent of a print serial or differ in the inclusion of such items as large graphs or data files that are impractical to provide in print or may include additional text that, for whatever reason, was not deemed suitable for the print version. For example, the *British Medical Journal* has adopted a publishing practice called ELPS—electronic long, print short. The same research articles appear in both the print and Internet versions of the journal, but the print version is usually substantially shorter. Another recent practice is to publish the Internet version in advance of the print counterpart to make the contents available more quickly to the intended audience. It should also be noted that a number of Internet serials lack the traditional volume and issue numbers, providing instead an article number or often simply a date. As with monographs, Internet serials may be static—that is, fixed in time and unchanged since publication—or may be updated or otherwise revised over time.

An important concept to recognize in discussing Internet serials is the "version of record". Publishers producing both print and Internet versions usually declare which version is to be considered the

authoritative one. More and more publishers are declaring that the Internet version is authoritative, primarily because of the ease of correcting errors and making other changes. It is important to instruct authors to state which version they used in writing their articles.

Home pages are peculiar to the Internet. A home page is defined as the first or introductory page of a Web site; it usually provides a table of contents or index of the site. Home pages are placed on the Internet by organizations and individuals for a variety of purposes, from providing detailed information about a government agency, a company, an association, or a subject to providing a forum for a personal point of view. Home pages vary widely in size and complexity, reflecting the Web site that they introduce. One explanation for the great variation is that there are no widely accepted standards for the content of a home page. If you wish to cite only a portion or segment of a Web site rather than the entire site, cite the portion or segment according to its format, such as a monograph or journal article.

How do we construct citations for the variety of Internet documents encountered? The title page is the usual place to look for citation information in a print publication, but no standards have been adopted for the Internet for the content of what would equate to a title page. Even established publishers that adhere to strict guidelines for their print publications often seem to abandon them when posting a document on the Internet. A widespread belief is that the Internet is a new publishing medium to which the old rules do not apply. It is the contention of NLM, based on the work of ISO and NISO, that the basic rules of citation are still applicable when referencing the Internet. However, without a traditional title page, the person formatting a citation may have to look further within a document or in non-

traditional locations, such as the source code or page information viewable with a Web browser. One major attempt to provide authoritative information for Web documents is the Dublin Core Metadata Element Set, ANSI/ISO Z39.85-2001.[2] This standard provides for a set of data elements—such as title, creator, publisher, and date—that should be provided in the source code. CSE members involved in the creation of Internet documents should become familiar with the Dublin Core and include it in all publications.

As we have seen, the Internet does present some unique problems when constructing a reference: Some citation elements are more difficult to locate; some, as will be discussed below, require expansion; and some require the use of elements with which many people are unfamiliar. Regardless, there should be an attempt to find the same elements as for a print publication, namely an author or organization with responsibility for a document, a title, a place of "publication", a publisher or organization issuing the document, and a date of publication.

Author—With the exception of the more traditional monographs and journal articles, most documents found on the Internet do not clearly state the names of their authors. Home pages in particular usually do not, although they often name the person who is Webmaster or contact person for the site. But those persons are not usually considered authors, especially in home pages produced by large organizations. Some sites give a name in association with a copyright statement, such as "copyright 1997 by John A Smith". Such a person should not be assumed to be the author. If the only personal name given in a site is associated with a copyright statement, the person named should be used as the publisher. Most sites display the name of an organization rather than a person. In such cases, if the organization appears to be serving as both author and publisher, place the organization in the publisher position. Regardless, do not use the word *anonymous* in a citation if no author can be determined. [Note: That is a change from

current *SSF* recommendations.]

Title—Monographs on the Internet usually display clearly identifiable titles, and serials have both the title of the article and the title of the journal. Home pages, in contrast, may display only the name of the organization responsible for the site. If so, this name is used as the title. An additional complication is that it is often not easy to ascertain the title in the collage of graphics presented on the site and to differentiate it from an array of advertisements. Here are some basic rules to follow for identifying wording as a title: (1) look for the most prominent (usually the largest) wording on the screen, (2) look for wording followed by a copyright or registered-trademark symbol (© or TM), (3) look at the title bar of the Web browser (generally in the top left corner), (4) look for the title in the source code of the document. If a title cannot be determined, construct a title by using the first series of words on the screen as a title.

Place of Publication—This is defined as the city in which the person or organization issuing or sponsoring the publication resides. In the case of the Internet, the place would be the location of the Web or other site. That information is usually found at the bottom of a home page but may also be at the top of the first screen or at the end of a document. If it is not in one of those locations, it may be obtained from a link within the site, usually a "contact us" or similar link.

Publisher—The advent of the Internet and other online sources has stretched the definition of "publication" and "publisher". In electronic terms, a publisher is defined as the person or organization that produces or sponsors the site. As with the place of publication, this information is usually found at the bottom of a home page, at the top or on a sidebar of the first screen, or at the end of a document. The publisher may also be identified by looking for the organization named after a copyright statement, such as "Copyright 1997 by the American Chemical Society". If such wording as "This site is maintained by XYZ Corporation for ABC Organization" appears, ABC Organization is considered

the publisher and XYZ the distributor. Publisher information is required in a citation; distributor information is optional and may be included as a note.

Dates—Because of the volatile nature of electronic publications, three dates are important in citing them: (1) the date when the publication was placed on the Internet or was copyrighted, (2) the latest date of any update or revision, and (3) the date when the person doing the citing saw the publication. The date of publication must always be included in a citation. The date of copyright should be included only if the date of publication is absent or it differs from the date of publication, for example, "2000, c1998".

The dates of publication and of updates or revisions are often absent from an electronic site, and this makes the third date all the more important. One way to determine a date for an Internet document is to use the Web browser to view the document page information or source. Webmasters who do not put a date in the online display may put one in the source document, either at the top or bottom. Dublin Core elements are placed at the top. Error correction and other changes in electronic publications may occur between scheduled or advertised updates or revisions, and these dates of update or revision may not be known, so a citation must include the date when the electronic publication was seen.

In addition to the traditional elements used in creating references, citing the Internet requires the use of some elements with which many people are not familiar or that must be defined in new ways.

Content and Medium Designators—It is standard practice for a citation to indicate that a publication is not in print format by placing after the title a word that describes the specific nonprint medium—a medium designator. This alerts the user that special equipment is needed to read the publication. The appropriate medium designator is placed in brackets. Thus, a book on microfilm would have "[microfilm]" after its title. *Internet* is also a medium designator. It is optional, but recommended, to combine

a content designator with a medium designator. Content designators indicate the nature of a work. Examples of this combination are "[serial on videocassette]" and "[home page on the Internet]".

Extent—The length of the item being cited is usually expressed as the total number of pages of a print item or the number of minutes of run time of an audiovisual. Although the extent of an item is optional in any citation, it provides useful information. For example, a 10-page journal article would be viewed as substantive, whereas a 10-page book probably would not.

Many electronic monographs lack traditional page numbers, and home pages are nonlinear, often having innumerable hypertext links. For publications other than home pages, extent is therefore usually shown as the number of screens, lines, paragraphs, or bytes. Alternatively, if an electronic document is printed, it may be expressed in the traditional number of pages. Unless the length is supplied by the publisher, which sometimes occurs when a list of items with their sizes is presented to the user for assistance with downloading or when the item being cited is a PDF document, the extent is calculated by the best means possible and placed in brackets, such as "[about 5 screens]", "[10 paragraphs]", "[about 21 p.]", and "[332K bytes]". Of course, screen sizes, fonts, and printers vary widely, but the purpose is to give the user of the citation an indication

of the length of an item. Note that when the number is approximated, the word *about* appears before the length indicator.

Availability Statement—The location at which an electronic document may be found is expressed as an FTP, Telnet, or Web address, for example, "Available from: Telnet to ovid.com" and "Available from: *http://www.nlm.nih.gov/pubs/cbm/dental-_caries.html*". For Internet addresses, the location displayed by the Web browser is usually the one to use. Sometimes, however, a site found by a hyperlink may not be addressable directly. It is therefore good practice to verify that the address works as a stand-alone entity before including it in a citation.

To illustrate the guidelines provided above, here are some sample Internet citations:

For a book:
Graber MA, Toth PT, Herting RL Jr. University of Iowa family practice handbook [Internet]. 3rd ed. Iowa City (IA): University of Iowa College of Medicine; 1997 Jul, c1992-2000 [modified 2000 Nov 28; cited 2001 Mar 7]. Available from: *http://www.vh.org/Providers/ClinRef/FPHandbook/FPContents.html*

For a journal article:
Cruz AA, Coehlo RP, Lucchesi MC. Upper eyelid shape and position in the association of Graves' disease and myasthenia gravis. Digital J Ophthalmol [Internet]. 2000 [modified 2001 May 10;

cited 2001 Mar 5];6(1):[about 6 paragraphs]. Available from: *http://www.djo.harvard.edu/meei/OA/Cruz/OA.html*

For a home page:
NursingWorld [Internet]. Version 3.2. Washington: American Nurses Association; c1995-2001 [cited 2001 Mar 12]. Available from: *http://www.ana.org/*

In summary, it is a popular misconception that simply adding a uniform resource locator (URL) or other electronic address to a title is sufficient for a citation. However, Internet sites change addresses or disappear with great frequency; a citation must therefore provide other identifying information. NLM states in its Internet citation document that it "hopes promulgation of citing guidelines such as this will encourage authors to include all the necessary information within their sites." Members of CSE should assume similar responsibility for creating Web documents that are easily citable.

The author welcomes comments. Please send them to *patrias@nlm.nih.gov*.

References

1. Information and documentation—bibliographic references—Part 2: Electronic documents or parts thereof. Geneva: International Organization for Standardization; 1997. (ISO 690-2.)

2. National Information Standards Organization (US). The Dublin Core metadata element set. Bethesda (MD): NISO Press; 2001. (ANSI/NISO Z39.85-2001.)