Click C@refully Before You Quote: Citing Internet-Based Sources

Grzegorz S. Litynski

ABSTRACT

At the end of the 20th century, access to information provided by the World Wide Web (WWW) is changing as never before. The fast availability of current medical literature and the availability of tools for easy access to information, as well as for the easy production of information, have confronted research physicians, scholars, and students with new kinds of problems, many of which concern us personally. Quality control, difficulty establishing basic citation components, lack of standard guidelines for citing, as well as the short lifetime of Internet addresses concern us deeply. Some of these problems could be solved by the concept of an "Online-Library of Medicine" presented in the following paper. Since, however, at the present time there are no good answers to the problems regarding citing Internet-based sources, a Web surfer must keep in his or her mind the motto "caveat lector" (let the reader beware) – or, rather, in the spirit of our time: click c@refully before you cite.

Key Words: Internet, Reference, Internet citation.

INTRODUCTION

The world has never been changing faster than it is today. At the end of the 20th century, we are experiencing one of the greatest technological and social revolutions ever: the beginning of the information age.[1]

Especially revolutionary is easy access to information provided by the World Wide Web. Available resources include commentaries and reviews, bibliographical and statistical databases, discussion groups and newstickers, multimedia textbooks and journals, clinical decision aids and patient simulations, case studies, job boards and information on residency, educational software and information on upcoming meetings, as well as many other sources of information. Indeed, in the past few years nothing has affected medical education and practice more than the Internet. The fast availability of current medical literature and the availability of tools for easy access to information, as well as for the easy production of information, have confronted research physicians, scholars, and students with new kinds of problems, many of which concern us personally. Some of these problems will be discussed in this paper.[2]

Reliability of Internet Sources

We must remember that a significant part (probably more than 95%) of the information present in Cyberspace is published there directly. This means that this information bypassed traditional publishers and established review systems. Literally, anyone with control of a Web site can claim to be an expert. At the same time, Web sites often provide information about diseases that may be treated by products and services sold by businesses that maintain or support the site. Consequently, the information provided may be biased to promote the commercial interests of the site sponsor.[3]

Difficulty Establishing Basic Citation Components

If the author of information appearing on a Web site is known, we can research his or her expertise, using, for example, MEDLINE. Unfortunately, numerous Web sites do not carry an author's name or the authorship is not clear, so that such "tracking down" becomes impossible. Even more complicated is dealing with image and sound files. These electronic documents lack certain compo-

Address reprint request to: Grzegorz S. Litynski, Institut für Geschichte der Medizin, Klinikum der, Johann Wolfgang Goethe University, Paul Ehrlich Str. 20-22, 60590 Frankfurt/Main, Germany. Fax: +49 700 7272 2727, E-mail: glitynski@aol.com, Web site: www.endo-highlights.com

nents that are used to construct Web site citation. Image and sound files do not usually contain information about authorship. Another crucial bibliographic element is publication date, which is difficult to determine. We must distinguish between dates of publication and "last update" and always remember that the edition (version) number is rarely published.

Lack of Standard Guidelines for Citing

There are numerous style guidelines that deal with electronic sources of information; they can be found both in print media and on the Internet. Since there is no standard work widely accepted by Web surfers and librarians, many scholars have summarized existing recommendations, adapted them or extended them for a particular academic discipline, or have presented their own approach.

The most popular style guide for citing electronic sources is that published by Xia Li and Nancy B. Crane, reference librarians at the University of Vermont-Burlington. Other widely accepted styles are those of the APA (American Psychological Association), the MLA (Modern Languages Association of America), the model of Melvin E. Page, a history scholar and experienced Internet user from East Tennessee State University, and various adaptations of Turabian styles. Despite the differences, each style guide attempts to do the same thing: indicate to the reader how to reach the requested original information source.[4]

All of the systems and models of citation of Internet-based sources, however, deal with the style of citation and less with the basic shortcoming of the WWW sources: their extremely short half-life. This leads us to pose the unorthodox questions, Why do we cite? Do we really have to do it?

THORNY PROBLEMS WITH CITING WWW SOURCES

Scholars cite their sources of information for three main reasons: 1) to insure that the information conveyed is accurate, 2) to guarantee their readers access to the full context in which the material was cited, and 3) to credit author(s) ("intellectual honesty"). In short, the essence of citation is verification of information.

"Here Today and Gone Tomorrow," or Dead URLs and Links

Clearly, neither intellectual honesty nor accuracy can be verified if the source consulted is no longer available. The reader cannot analyze the researcher's use and interpretation of the evidence. Frequently, only months or even weeks (sic!) after publishing, an electronic document gets deleted from a university server. Internet addresses (so-called "URL" – Uniform Resource Locator) are often changed by Webmasters or system administrators for technical purposes, and Web sites may disappear without any trace from cyberspace, due, for example to lack of sponsors. Repeatedly, electronic messages are moved after a certain period of time and stored in the administrator's archives, which are not generally accessible to the public. This is true especially of list-servers' messages or newsgroup postings.

Unfortunately, this phenomena is not a sporadic one. When the author of this paper checked ARS MEDICI, a collection of 20,000 medical Internet addresses published on a CD-ROM in 1997, it turned out that over 50% of all addresses were dead links.[5] Can we imagine any scientific paper citing references of which less than 18 months later only 50% are available? A bibliographical disaster!

"Invisible Changes," or New Contents and Old URLs

Another example: What if someone (computer system manager, Web site author or hacker) modifies a document AFTER a surfer has cited a particular electronic document? For example, the author of a Web site on endoscopy presents some facts on the history of that procedure under URL: www.example-endoscopy.com/history.html. There, a surfer may learn that the first surgeon who performed laparoscopic cholecystectomy (LC) was Phillipe Mouret of Lyon, France, in 1987. Some time later, however, the author of the site discovers that it was E. Mühe of Böblingen, Germany, who had carried out the first LC in 1985, and not the aforementioned French surgeon. Consequently, the author alters the contents of the Web site because of this "new" information but not its address (URL) -- a common routine in Cyberspace known as "updating." Now, it is possible to have two groups of researchers pointing to the above-mentioned Web site: one group, who picked up the historical information before the site was updated ("Phillipe Mouret was the first"), and the other group, who reached the Web site after it was updated ("Erich Mühe was the first"). As a result, we are confronted with an unacceptable situa-

tion -- both groups would quote the same reference (Internet address), but both would refer to totally different facts.

Citing and the Topic of (not only Scientific) Responsibility

The situation might conceivably become more problematic (or even dramatic) if this phenomena occurs on Web sites dealing with clinical issues, for example, application of drugs, interpretation of labor findings, or recommendation of certain operation techniques. The law dealing with information obtained from or through the Web is not well defined. Potential legal issues include responsibility for diagnostic and therapeutic procedures based on Internet information.

Equally troublesome is the question of interaction between physicians and patients over the Internet, when, for example, it may occur that a physician practices medicine without being licensed in the state or country in which the patient resides.

Clearly, there is no foolproof mechanism to prevent erroneous information from being placed in print media, but with printed material there are effective methods to store the information. Indeed, in the case of print media, we can reach the source directly at any time through any number of means, such as interlibrary loans. Due to well developed library systems, it is not a problem to reach, for example, the works of Theodor Billroth (1829-1894), William S. Halsted (1852-1922), or Rudolf Virchow (1821-1902), written more than 100 years ago. In the case of electronic sources, we do not have such a system.

IS THERE ANY WAY OUT OF THIS?

Various solutions have been presented both in print media and on the Web to manage these problems. Many expert users propose downloading (or printing) electronic documents and posting the messages upon request. This is only a partial solution: if a reader cannot retrieve the source by himself or herself, scholarship becomes more an act of faith rather than an act of research. A citation, from a scientific point of view, becomes more or less meaningless. Neither downloading nor printing sources at the moment of their quotation makes the information available for other surfers once it is deleted. It is no wonder that archivists and librarians all over the world have launched various projects to collect electronic sources.[6]

The Concept of a Central Server, "Online-Library of Medicine"

It is nearly impossible to imagine a library collecting all Web information along with countless updates even in a single language. Theoretically, a variety of ways of solving the problem of availability to at least portions of documents are feasible. A thinkable solution would be the foundation of a central server, let us call it "Online-Library of Medicine." It would be controlled by an authority -- institutional or not. An author would submit his or her electronic document; "Online-Library of Medicine" would collect, analyze and register the publication. "Online-Library of Medicine" also would approve the electronic document and make it available on its own server. And only such "approved by Online-Library of Medicine" sources -- both on the author's or the Library's server -- would carry weight as quotable sources for the scientific community. Most importantly, it would be accessible to anyone at anytime. Certainly, approved documents would be available also as printed matter to those having no Internet access -- for example, through a local librarian or interlibrary loan system.

In order to provide reliable and repeatable data, anonymity must be replaced by clearly stated authorship and date of first publication, and the document itself must be stored on a server under intellectual control.

CONCLUSIONS

Presently, there are no good answers to the problems regarding quality control, short lifetime, and lack of accessibility to Internet sources. We can only follow general rules and make suggestions by analogy to ink-on-paper style. The further cooperation of archivists, librarians, expert users, motivated providers, information system developers, computing professionals, and, finally, health specialists is necessary in order to provide the medical community with Internet data capable of being citing.

It does not matter what solution will be developed in the future to solve the key issue of the availability of a previously consulted information source, "Online-Library of Medicine" or any other solution. But one thing is certain, we urgently need a system making Internet resources available for all surfers, even years after the resources were first published online.

Table 1.

Short Checklist for Reliability of Internet Sources for Quotation.

Topic	Possible Answer	Remarks
Authorship:		
Author's Name	a. known b. unknown	
Author's expertise	a. known b. unknown	Check in Medline.
"Publisher" or Information Carrier:		
Type of the information carrier (provider)	a. educational b. commercial c. private d. others	
Provider online since ...	... (years)	Always prefer services that have been online for longer periods of time -the chance that they may disappear "over night" is slimmer.
Date of Publication:		
First date of publication	a. known b. unknown	If unknown, frequently the references or bibliography attached to the electronic document can give an idea of the year of publication / last update.
Last update	a. known b. unknown	
Citing the Source:		
URL (Internet address):		Right click above the electronic document, go to "Properties."
Additional helpful information:		
Paper previously published in print-media	a. yes b. no or **unknown**	If yes, always prefer quotation of a printed version of a document and add "available online under: ..."

Until that time, we must keep in our minds the motto "caveat lector" (let the reader beware), or, rather, in the spirit of our time: click c@refully before you cite.

AUTHOR'S REMARK

Due to limited space, it was not possible to include in this paper detailed models of all of the above-mentioned styles of citing (APA, etc.) They can be found on the author's Web site under www.endo-highlights.com/internet/citing.html. For their actuality, but not contents, the author takes full responsibility.

A *Short Checklist for Reliability of Internet Sources* is provided and designed to make the surfer more aware of the issue of the quality of Internet-based Sources **(Table 1)**.

References:

1. Coiera E. Information epidemics, economics, and immunity on the internet. We still know so little about the effect of information on public health. *Brit Med J.* 1998;317:1469-1470. Curtis KL, Wellter AC, Hurd JM. Information-seeking behavior of health sciences faculty: the impact of new information technologies. *Bull Med Libr Assoc.* 1997;85:402-410. Obst O. Use of internet resources by German medical professionals. *Bull Med Libr Assoc.* 1998;(Oct)86:528-533.

2. Pines WL. The challenge of the Internet. *Drug Inf J.* 1998;32:277-281. Veldenz HC, Dennis JW. The Internet and education in surgery. *Am Surg.* 1998;64:877-880. Hopp DI. Three topics integral to the use of the Internet for clinical trials: connectivity, communication, and security. *Drug Inf J.* 1998;32:933-939.

3. Eysenbach G, et al. Towards quality management of medical information on the Internet: evaluation, labeling, and filtering of information. *Brit Med J.* 1998;317:1496-1502. McLeod SD. The quality of medical information on the Internet. A new public health concern. *Arch Ophtalmol.* 1998;116:1663-1665.

4. Xia L, Crane NC. Electronic Styles: *A Handbook for Citing Electronic Information.* Medford: Information Today Inc.; 1996. *Publication Manual of the American Psychological Association. 4th edition.* Washington D.C.: American Psychological Association; 1994. Gibaldi J. *MLA Handbook for Writers of Research Papers, 4th edition.* New York: Modern Languages Association of America; 1995. Gibaldi J. *MLA Style Manual and Guide to Scholarly Publishing, 2nd edition.* New York: Modern Languages Association of America; 1998. Harnack A, Kleppinger E. *Online! A Reference Guide to Using Internet Sources.* New York: St. Martin's Press; 1997. Turabian KL. *A Manual for Writers of Term Papers, Theses, and Dissertations. 5th edition.* Chicago: University of Chicago Press; 1996.

5. Ars Medici. *Internet-Adressen Medizin.* search offline – find online 1.0. Stiftung NeoCortex, ed. Bern: Hans Huber Verlag; 1997.

6. Norman F. Digital libraries – a quality concept. *Int J Med Inf.* 1997;47:61-64.