Science through the Internet: Researching, Evaluating and Citing Websites

L. E. J. Lee[*1] and E. Misser[2]

[1]Biology Department and [2]Writing Centre, Wilfrid Laurier University, Waterloo, ON, Canada, N2L 3C5.
[*]To whom correspondence should be addressed. E-mail: llee@mach1.wlu.ca

ABSTRACT

This article attempts to convey the joys and frustrations of skimming the Internet trying to find relevant information concerning an academic's work as a scientist, a student or an instructor. A brief overview of the Internet and the "do's and don'ts" for the neophyte as well for the more seasoned "navigator" are given. Some guidelines of "what works and what does not" and "what is out there" are provided for the scientist with specific emphasis for biologists, as well as for all others having an interest in science but with little interest in spending countless hours "surfing the net". An extensive but not exhaustive list of related websites is provided.

INTRODUCTION

In the past few years the Internet has expanded to every aspect of human endeavor, especially since the appearance of user-friendly browsers such as Netscape, Microsoft Internet Explorer and others. Browsers allow easy access from anywhere in the world to the World Wide Web (WWW), which is a collection of electronic files that are the fastest growing segment of the Internet. Correspondingly, we are drowning in a sea of information while starving for knowledge. Can we manage this wealth of information into digestible knowledge? Yes! With help and perseverance. However, given the magnitude and rate at which the Internet changes, this article cannot provide a comprehensive guide to available resources; rather, it serves primarily as a starting-point in the individual quest for knowledge.

WHAT IS THE INTERNET?

The Internet is a worldwide computer network started by the US government primarily to support education and research. Many books and reviews exist that detail the Internet in almost every aspect. Among these, "The World Wide Web–Beneath the Surf" by Handley and Crowcroft (1) gives basic information and history. A succinct overview in a tutorial format has been set up by the University of California at Berkeley Library (2). It provides a quick start to finding information through the Internet. Information about teaching and learning through the "Web" can also be found in study modules set up by Widener University's Wofgram Memorial Library (3). For the science aficionado, concise information containing a primer to the

Internet for the biotechnologist can be found in a recent review by Lee at al., 1998 (4). For more in-depth knowledge, two books of interest to the biologist are Swindell et al., 1996 (5) and by Peruski and Peruski, 1997 (6). However, given the scope and the rate of growth of the Internet, estimated at 40 million servers and predicted to reach over 100 million servers by the year 2000 (7), any review can become obsolete within months of publication. (Table 1 illustrates growth estimates of the Internet).

IMPORTANT INTERNET TERMINOLOGY

What are URLs?

URL stands for Universal (or Uniform) Resource Locator and is analogous to the address protocol used in sending and receiving regular mail. The first portion usually refers to the protocol type, for example:
- HTTP (hypertext transfer protocol) allows users to access the information in hypertext format, namely clickable sites and multimedia (sound, graphics, video).
- FTP (file transfer protocol) permits transfer of files, whether these are text files, image files or software programs.
- GOPHER is an obsolete text transfer protocol without multimedia access that preceded HTTP.

The next portion of the URL is a set of letters or numbers that indicate website address and files. For a more detailed explanation see "Understanding and decoding URLs" by Kirk, 1997 (8).

STARTING-POINTS FOR INTERNET RESEARCH

Due to the size of the Internet, one needs to rely on various software, called search engines, to find appropriate information. A common start-up site that can provide quick subject catalogs by topic area is Yahoo (11). Many single or multiple database search engines perform broad searches on a topic by keyword. Links to these can be found through the Internet Public Library (IPL) (12). The most popular engines include: Lycos (13), Excite (14), Infoseek (15), Dogpile (16), and Metacrawler (17). A recent addition that allows for one-step searching of web-pages and full-text journals is Northern Light (18). This engine is recommended for scientists, but access to its full text articles requires payment.

A comparison of various search engines' performance with overall tips for Internet searching can be

Table 1: Growth of the Internet in terms of Host computers

Date	Hosts	Date	Hosts
05/69	4	07/88	33,000
10/69	5	07/89	130,000
04/71	23	10/90	313,000
06/74	62	07/91	535,000
03/77	111	07/92	992,000
08/81	213	01/93	1,313,000
05/82	235	01/94	2,217,000
08/83	562	01/95	5,846,000
10/84	1,024	01/96	14,352,000
10/85	1,961	01/97	21,819,000
11/86	5,089	01/98	29,670,000
12/87	28,174	07/98	36,739,000

Data from Promo Net (9) & Network Wizards (10)

found at the Okanagan University College Library (19).

Other sites containing links to sites of scientific relevance include SciCentral (20), SciWeb (21), BioMed Net (22) and Science Channel (23), among others. A comprehensive list cataloguing selected sites for biomedical sciences can be found at Biosites (24) and at the IPL Biological Sciences Reference (25).

Timely topics in science are provided by Scientific American (26). Abstracts of scientific articles catalogued by the National Library of Medicine can be searched for free using Medline (27), and those catalogued by the National Agricultural Library, using Agricola (28). Some sites allow for free perusal of full text but few such journals exist. A good site for development, cell science and experimental biology can be found at the Company of Biologists (29). Some free online magazines that may be of interest include: In Scight (30) produced by Academic Press in partnership with Science Magazine, ScienceNow (31) sponsored by the American Association for the Advancement of Science, UniSci (32), HMSBeagle (33) from BioMedNet. As well as Network Science (34).

Despite the abundance of websites, effective and efficient searching can be frustrating when a query results in over 100,000 hits. Successful search strategies are typically through experience and discipline, although following the guidelines indicated by (2, 3) and the comprehensive basic guide for general researching and writing from the IPL (35), can be most helpful. Nonetheless, searching through the Internet has become a common and convenient feature, necessitating one to approach each WWW site with caution. Some guidelines are given below.

GUIDELINES FOR DETERMINING RELIABILITY AND VALIDITY OF WEBSITES

The Internet changes daily as resources are added, changed, moved or deleted. Millions of people, young and old, as individuals or within organizations create resources ranging from basic information about themselves, their interests or their products, to complex lists of funding resources, multimedia textbooks, full-text journals, clinical information systems, epidemiological and statistical databases, and the like. One of the most pressing needs is to evaluate these resources for accuracy and completeness. All information should be received with skepticism, unless an evaluation of a site can be performed.

Relevant question in evaluating a site include the following: Is the site affiliated with a reputable institution or organization such as a University, government or research institution? URL's may reveal this information: "edu" includes most educational institutions, "gov" indicates government affiliated sites, and "com" refers to commercial enterprises, while "org" suffixes are used by many non-profit organizations. The two-letter suffix on non-USA sites indicates the country of origin (8). Is there a tilde (~) in the site address? Usually personal webpages are indicated with a tilde, and although not necessarily bad, one should be particularly careful when evaluating such sites. Other questions to keep in mind: Is there a particular bias? Who is the author? What are their credentials? How current is this site? Many sites have been abandoned and sit as "junkyards" of old information. How stable is the site? Is the general style of the site reliable? Consider grammar and spelling.

Critical evaluation of websites

Many websites provide strategies for the critical evaluation of webpages. The University of Florida with a list of short tips (36), Purdue University provides a step by step checklist (37), and Widener University has page-specific checklists (38). Another list of evaluating resources posted by many librarians can be found through the University of Washington Libraries site (39).

The following are some points to consider when visiting sites:

1. Content: is real information provided? Is content original or does it contain just links? Is the information unique or is it a review? How accurate is it? What is the depth of content?

2. Authority: who or what is the source of the information? What are the qualifications?

3. Organization: how is the site organized? Can you move easily through the site? Is the information presented logically? Is the coverage adequate? Can you explore the links easily? Is there a search engine for the site?

4. Accessibility: can you access the server dependably? Does the site require registration? If so, is it billed? Can it be accessed through a variety of connections and browsers? Is it friendly for text viewers? How current is it? Is it updated regularly?

5. Ratings: is the site rated? By whom? Using what criteria? How objective is it? If the site is a rating service itself, does it state its criteria?

CITING WEBSITES

Information from any source should be properly referenced whenever possible as intellectual property and copyright laws usually apply. Electronically stored information presents new challenges since no method exists to easily monitor this vast "global library". However, scholarly activity should maintain a high standard of conduct by following appropriate citation protocols.

Several citation formats exist for referencing webpages. Two common citing conventions are the MLA style from the Modern Language Association of America (40), and the APA style from the American Psychological Association (41). The latter acknowledges a guide by Li and Crane, 1996 (42) to its style for citing electronic documents. Slight variations exist, depending on whether the citation is from individual works, parts of works, electronic journal articles, magazine articles, or discussion list messages. Detailed information for these can be found in Crane's webpages (43), for APA style and for the MLA style (44). A proposed Web extension to the APA style has recently been reported by Land (45).   Consider however, that there are many citation style guides for electronic sources. Some of these sites are listed at the University of Alberta Libraries (46).

All references should generally contain the same information that would be provided from a printed source (or as much of that information as possible). If author of the site is given, their last name and initials are

placed first, followed by the date that the file was created or modified (full date in day/month/year format or year, month/date if feasible) and the title of the site in quotations. If affiliation to organization is known, this should be indicated. The date the resource was accessed is placed next (day/month/year or year, month/date), and finally the complete URL within angle brackets. Care should be taken not to give authorship to webmasters who are responsible for posting or maintaining information on webpages and are not the originators of the contents. However, they can be referenced as editors with the generic Ed. abbreviation. Finally, in some instances, Internet resources are also published on hard copies, in those cases, the appropriate citation format should be followed and the URL address should also be indicated.

Organization of Bibliography

Bibliographic format varies according to the preference of the publisher, institution, or journal. In general, include authors in alphabetical or in numerical order of appearance. Some prefer separate bibliographies for paper-based "hardcopy" references and for "softcopy" electronic sources. Others permit intermixing (as in the present article). If the author is unknown, site names are listed in appropriate order. Should some information be missing, it is acceptable to omit this information and still cite the reference. For example, some sites may not show authors or dates or have any indication of affiliation. However, the URLs should always be indicated.

CONCLUSION

The Internet holds vast and exciting possibilities for the scientific community and for society as a whole. The power of the individual can be multiplied by the "click of a mouse" as new capabilities are provided by linking various computing systems to the global village. Nevertheless, the Internet as seen through the WWW can be addictive. One "click" effortlessly from one site to another in a seemingly endless and aimless loop. Enjoy or despair, at your own risk!

REFERENCES

1. Handley, M. and J. Crowcroft (1994) "The World Wide Web - Beneath the Surf" UCL Press, London, UK. <http://www.cs.ucl.ac.uk/staff/jon/book/book.html>
2. The Library, University of California, Berkeley (20 October 1998) "Finding information on the Internet: A Tutorial" 10 December 1998. <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>
3. Alexander, J., Powell, J. and Tate, M. A. (March 1997) "A modular approach to teaching/learning the world wide web" Widener University/ Wolfgram Memorial Library. 14 December 1998. < http://www.science.widener.edu/~withers/pyramid.htm>
4. Lee, L. E. J., Chin, P and Mosser, D. D. (1998) Biotechnology and the Internet. *Biotechnology Advances* **16**, 949-960.
5. Swindell, S. R., Miller, R. R. and Myers, G. S. A. (1996) "Internet for the molecular biologist" Current innovations in molecular biology series ; v. 3. Horizon Sci. Press, Wymondham, UK.
6. Peruski, L. F. and Peruski, A. H. (1997) "The Internet and the new biology : tools for genomic and molecular research" ASM Press, Washington, D.C.

7. Gray, M (20 June 1996) "Internet Statistics" Student Information Processing Board, Massachusetts Institute of Technology <http://www.mit.edu/people/mkgray/net/>

8. Kirk, E. E. (31 October 1997) "Understanding and decoding URLs" Milton S. Eisenhower Library, Johns Hopkins University. 3 December 1998.
<http://milton.mse.jhu.edu:8001/research/education/net.html >

9. PromoNet (May 1998) "Promo.Net Home" 12 December 1998. <http://promo.net/>

10. Lottor, M. K. (1998) "Network Wizards" 12 December 1998. <http://www.nw.com>

11. Yahoo! Inc. (1994) "Yahoo!" 15 December 1998. <http://www.yahoo.com/>

12. Schwartz, K. L. (1997) "A+ Research and Writing Links" Internet Public Library. 09 December 1998. <http://www.ipl.org/teen/aplus/linksdirect.htm>

13. Lycos, Inc. (1998) "Lycos" Carnegie Mellon University. 15 December 1998.
<http://www.lycos.com/>

14. Excite, Inc. (1995) "Excite" 15 December 1998. <http://www.excite.com>

15. Infoseek Corporation (1994) "Infoseek" 15 December 1998. <http://www.infoseek.com>

16. Dogpile (No date) "Dogpile" 15 December 1998. <http://www.dogpile.com>

17. Go2Net, Inc. (1998) "MetaCrawler" 15 December 1998.   <http://www.metacrawler.com>

18. Northern Light Technology (1997) "Northern Light" 15 December 1998.
<http://www.northernlight.com>

19. Tyner, R. (May, 1996) "Sink or Swim: Internet Search Tools and Techniques" Okanagan U. College. 08 December 1998. <http://www.sci.ouc.bc.ca/libr/connect96/search.htm>

20. SciLink, Inc. (1997) "SciCentral" 15 December 1998. <http://www.scicentral.com>

21. SciWeb, Inc. (1996) "SciWeb" 15 December 1998. <http://sciweb.com/index.html>

22. BioMed Net Limited (1998) "BioMed Net" 15 December 1998.   <http://biomednet.com>

23. Science Channel (1998) "Science Channel" Elsevier Science. 15 December 1998.
<http://www.sciencechannel.nl/>

24. BioSites (1998) "BioSites" University of California at San Francisco. 15 December 1998.
<http://www.library.ucsf.edu/biosites>

25. Internet Public Library (1998) "IPL Biological Sciences Reference" 15 December 1998.
<http://www.ipl.org/ref/RR/static/sci1500.html>

26. Scientific American Inc. (1996) "Scientific American" 15 December 1998.
<http://www.sciam.com>

27. Medline (1998) "PubMed" US National Library of Medicine. 15 December 1998.
<http://www.ncbi.nlm.nih.gov/PubMed/index.html>

28. Agricola (1998) "Agricola Home Page" US National Agricultural Library. 21 December 1998.
<http://www.nal.usda.gov/ag98/ag98.html>

29. Company of Biologists Ltd. (1998) "COB". 15 December 1998. <http://www.biologists.com/>.

30. Academic Press (1998) "InScight" Science Magazine. 15 December 1998.
<http://www.apnet.com/inscight>

31. American Association for the Advancement of Science (1998) "ScienceNow" 15 December 1998.
<http://www.sciencenow.org>

32. UniScience News Net (1995) "UniSci" 16 December 1998.   <http://unisci.com>

33. BioMed Net Limited (1998) "HMSBeagle" 16 December 1998. <http://biomednet.com/hmsbeagle>

34. Network Science Co. (1998) "Network Science" 15 December 1998. <http://www.netsci.org>

35. Schwartz, K. L. (1997) "Research and Writing Step by Step" IPL. 08 December 1998. <http://www.ipl.org/teen/aplus/stepfirst.htm>

36. Edwards, J. (23 November 1998) "Tips for evaluating a world wide web search" University of Florida. 23 November 1998. <http://www.uflib.ufl.edu/hss/ref/tips.html>

37. Scholz, A. (1996) "Evaluating world wide web information" Purdue University Libraries. 08 December 1998. <http://thorplus.lib.purdue.edu/research/classes/gs175/3gs175/evaluation.html>

38. Alexander, J. and Tate, M. "Evaluating Web Resources" (08 August 1996) Widener University Wolfgram Memorial Library. 03 December 1998. <http://www.science.widener.edu/~withers/webeval.htm>

39. Bartelstein, A. and Zard, A. (June 1996) "Resources: Other sites about evaluating information on the Internet" In: "Teaching Students to Think Critically about Internet Resources" University of Washington Libraries. 30 November 1998. <http://weber.u.washington.edu/~libr560/NETEVAL/resources.html>.

40. Modern Language Association of America (1998) "MLA style" 23 November 1998. <http://www.mla.org/main_stl.htm#sources>

41. American Psychological Association (1998) "How to cite information from the Internet and the World Wide Web" 23 November 1998. <http://www.apa.org/journals/webref.html>

42. Li, X., & Crane, N. B. (1996). Electronic styles: A handbook for citing electronic information (Revised edition). Medford, NJ.

43. Crane, N. B. (29 October 1997) "Electronic sources: APA style of citation" University of Vermont. 23 November 1998. <http://www.uvm.edu/~ncrane/estyles/apa.html>

44. Crane, N. B. (29 October 1997) "Electronic sources: MLA style of citation" University of Vermont. 23 November 1998. <http://www.uvm.edu/~ncrane/estyles/mla.html>

45. Land, T. (15 October 1998) "Web extension to American Psychological Association Style (WEAPAS)" 17 December 1998. <http://www.beadsland.com/weapas/#DV>

46. Tong, J. and Philip, B. (January, 1998) "Library guides-Citation style guides for internet and electronic sources" U. of Alberta Libraries. 08 December 1998. <http://www.library.ualberta.ca/library_html/help/pathfinders/style>