

## VII.1

# The clean, green net: environmental computer resources under construction

William B. DeVille

### VII.1.1. Introduction

The Internet is a worldwide linkage of computers that use standardized data transfer protocols allowing relatively rapid, relatively inexpensive transfers of information across the network. It is a set of information management technologies that will as surely transform the world as did written languages, printing, industrialization, mass transport and mass communications.

Vannevar Bush, Director of the U.S. Office of Scientific Research and Development during World War II, conceived new technologies for managing and presenting information that would have an impact far beyond merely facilitating easier and faster access to data. His vision was that the result would be new ways of relating data and, ultimately, new ways of publishing information in ways that incorporate these relational links.

The title of Dr. Bush's (1945) article<sup>1</sup> in *The Atlantic Monthly*, "As We May Think", emphasized his belief that future information technologies could result in radical changes in the way humans assimilate, comprehend and use information. He called the encyclopedic array of data made available by advanced information technologies, the "memex". A person viewing the memex might establish new pointers, or links relating data in new ways, that could then be published as a contribution to the memex.

This somewhat philosophical introduction may assist the reader to grasp the underlying construction of the World Wide Web (WWW). The web is a large and growing set of data (web pages, text blocks, graphics, sounds, movies, etc.) among which links may be established using the hypertext markup language (HTML) and uniform resource locators (URLs, the addresses of sites or objects on the web).

*Objective and a caution.* The objective of this chapter is to give an overview – with particular illustrations – of some of the environmental science and engineering resources on the Internet. I also hope to give the reader a feel for potentially useful ancillary

---

<sup>1</sup> If the reader has access to the WWW on the Internet, it would be worthwhile to perform a search for the term "Vannevar Bush" and then select a hypertext link to an electronic version of the article. The web address (known as a URL for "uniform resource locator") for the Lycos web searcher is <http://lycos.com> (copy the address precisely as printed, and enter it into the *open location* or *open dialogue box* in your web browser).

resources, from environmental law and regulation databases to Internet communications lists ranging from professional peer discussion groups, to environmental sites for the general public. Emphasis will be given to the WWW because of the richness and power of the user interface (with graphical interface browser software such as Netscape Navigator or Microsoft Explorer) and the intuitive ease of use of the web's hypertext links.

A cautionary note is due to the continuing rapid evolution of everything on and associated with the Internet, from the software protocols and languages to the implementing and viewing of computer hardware and software. Above all, the scope and content of information resources and the techniques used to present them are constantly growing and changing.

There is a kind of tension in writing about computer information resources within the medium of a book chapter. Books – even those written on topical, time-related subject matter – have a kind of permanency that is lacking in the rapidly evolving world of computer information networks.

Ten years after this book has been published, I would expect to be able to go into a library, locate it through the card catalog (most likely, a computerized one), pick it up, and read it just as it was published. That assumption will not hold for many of the computerized information resources that are available today. The technological media for accessing the information (including the computer hardware, operating systems, network protocols, search engines and software readers) will have changed in 10 years. The databases available today on the Internet's WWW will likely be unusable on the hardware/software systems that will be in use 10 years in the future – just as they would have been impossible to use in today's form, 10 years ago.

A number of examples of information resources on the web are provided in this chapter, including their current web site addresses. Please be cautioned that the Internet addresses of sites listed may change rapidly, even over the next 2 or 3 months. But this is not as much a problem as it might first seem, because the Internet also provides search engines working with continually updated databases of resources that allow the user to find addresses on any topic. It is also very common to see "under construction" labels on web pages, indicating that changes are being made literally on a day-to-day basis.

### **VII.1.2. An overview of the World Wide Web**

The initial conception and development of the web was done at the European CERN laboratories. The web has been called the "killer application" of the Internet, because the use of HTML for formatting and linking files allowed simple and intuitive navigation around the Internet. Web viewers do not have to master the complex commands of the UNIX computer operating system. Furthermore, HTML richened the appearance and formatting of documents by contrast to the plain, unformatted screens to which Internet users had been accustomed. Although HTML-formatted documents are actually plain text with added formatting tags, browsing software (such as Netscape or Explorer) allowed superimposition of a graphical interface and "point and click" navigation among documents and web sites. This ease of navigation is undoubtedly the reason for rapid growth of Internet use, and reminds one of Vannevar Bush's visions of the electronic library.

HTML formatting also allows electronic publication of documents that are essentially platform-independent. That is, a document prepared and stored on a UNIX computer may be accessed and viewed by computers using Macintosh or Windows operating systems, and appear essentially the same on all three types of computers, including complex features such as graphics, sounds, animations and so on. The list of new features available is constantly growing as HTML and its associated browsing software evolve.

As is typically the case on the Internet, committees<sup>2</sup> exist that attempt to establish standardized features of the HTML language. As is also typically the case on the Internet, some software developers do not adhere strictly to the recommended standards. This means that some documents may appear differently, or have features that are not consistent between different browsers. Many web pages now provide downloadable documents in the Adobe Acrobat page description format (PDF). This provides greater control of the document formatting and appearance than does HTML, and is especially suitable for long documents.

One of the more exciting developments is the use of an extension of HTML formatting called XML (extended markup language). XML can permit far better control of document formatting and definition of data elements within web pages than does HTML. Imagine, e.g., the difficulties that researchers encounter in compiling and evaluating data that is spread among many host computers on the Internet. XML is already allowing chemists to share information through use of XML data definitions that support agreed-on specifications for the data.

### **VII.1.3. An overview of web resources**

There are already millions of sites on the web, holding billions of published items. It would be wise for web “surfers” to recognize that it is in some ways a new and strange environment, lacking many of the traditional institutional contexts that we use to judge the validity of information.

*Caveat emptor.* Web sites range from home pages created by individuals to those maintained by professional societies, environmental organizations, publishers, universities, governmental agencies and companies. Their quality (from any point of view ranging from aesthetics to trustworthiness of content) extends over a broad spectrum. Most of them offer free access. Some restrict access by requiring passwords, and limit themselves to paying subscribers, organizational membership or some other access criterion. The Internet exercises no control on the type or quality of information on the web. It remains a fact that it is much cheaper and easier to publish electronically on the Net, than to publish through a recognized scientific or engineering publishing firm or a peer-reviewed journal such as *Science* or *Nature*.

This means that the burden of validation of information obtained from the web rests squarely on the user, who must take into account the reliability of the source of the information. There are, however, some fairly conventional indicators of the probable veracity and quality of information based on sources. For example, scientific and engineering professional societies, major university academic departments and governmental agencies

---

<sup>2</sup> <http://www.w3.org/>.

tend to review the content quality of web sites associated with them. Authorship by a person of known credentials can be a good quality indicator (although cases of falsely ascribed authorship have already cropped up on the Internet).

*Familiar academic trappings in a new wrapping.* Many peer-reviewed scientific journals have already appeared on the Net, and this will become more common in the near future. One of the first is the *Journal of Molecular Modeling*,<sup>3</sup> which is available in electronic form on the Internet or on CD-ROM as well as in paper form. This journal's web site advertises itself as "The first fully electronic journal in chemistry – the advanced way of publishing." Abstracts and papers may be downloaded to the viewer's computer in one of the two ways, as HTML files formatted for viewing by a web browser or as PDF (portable document files) viewable by Adobe's Acrobat Reader software. In either case, the papers are well formatted, can include pictures and graphics and generally look much like similar papers printed in contemporary scientific journals. The HTML formatting language allows additional elements such as movie clips with sound, rotating three-dimensional model presentations, and high-quality color images. Ancillary files, such as raw data sets, could easily be attached to the "paper" for downloading by the viewer. It will be interesting to see how such features, which go beyond the capabilities of the printed page, may influence the techniques used by scientists to present their research.

#### VII.1.4. Finding information on the web

*Web search engines.* An obvious question is, "How do I find the information that I'm interested in?" One possibility is to use a web searcher. Several are:

---

Google	<a href="http://www.google.com/">http://www.google.com/</a>
Lycos	<a href="http://lycos.com">http://lycos.com</a>
WebCrawler	<a href="http://webcrawler.com/">http://webcrawler.com/</a>
Altavista	<a href="http://www.altavista.com/">http://www.altavista.com/</a>
Amazon	<a href="http://www.amazon.com/">http://www.amazon.com/</a>
Scirus	<a href="http://www.scirus.com">http://www.scirus.com</a>

---

Of these web searchers, Scirus is a specialty search engine that targets *scientific information only* in earth and planetary sciences, and environmental sciences and covers over 100 million science-related pages.

These and several other search engines provide free access for relatively simple types of keyword searches.

*Caution: No web search engine will find all information that is out on the web.* Of the current 5 billion or so web pages, the search engines "know" about less than 10% – and they may not be very selective in choosing the search "hits" presented to

<sup>3</sup> <http://www.ccc.uni-erlangen.de/jmolmod/index.html>.

the user. For this reason, it is advisable to become knowledgeable about potential sources of information on the web, such as university, professional society or governmental web sites and to examine such web sites as supplements to searches made with web search engines.

One example is a search I conducted for the term “pollution prevention” using the Google search engine. Although by no means a comprehensive listing of sites with pertinent information, and although many of the items found were either of no interest or of dubious quality, some of the search hits were very useful. Google found about 515,000 relevant web links for this simple search. Fortunately, I found what I was looking for in the first page of search results.

### **VII.1.5. Google search results for “pollution prevention”**

*Results:* About 515,000 *Search time:* 0.07 s.

Although I have recently sifted through more than a thousand web sites on the topics of solid waste, hazardous waste and pollution prevention, I had neither the time nor the patience to look at all the 515,000 “hits” provided by Google. Fortunately, Google (like many other search engines) tries to provide relevant “hits” on the first few pages of the search results, i.e. it tries to prioritize the results. The term “pollution prevention” keyed a “canned” response based on someone’s understanding of the context of the terms, and resulted in an intelligently selected set of hits. In this case, the search provided URLs to several sites with useful information on the topic. Even better, several of those sites provided still more links to the topic.

In this case, I was really interested in finding references on pollution prevention for leather tanning. Simply by changing the Google search term to “pollution prevention leather tanning”, the number of references found by Google dropped dramatically, and were far more relevant to my immediate interest.

*Results:* About 2260 *Search time:* 0.20 s.

“*Web worms*” and indexing. Several organizations index the web by means of automated programs (“web worms”) that seek out web sites and send back indexes of their contents to their home computer. This allows web services such as Yahoo<sup>4</sup> to compile lists of “significant” web sites. In this sense, Yahoo is somewhat more sophisticated than some of the other web search services, which rely entirely on automated, non-contextual indexing of the contents of sites.

The problem is that most current indexing is non-contextual, which can result in surprising and often very irrelevant responses to search queries. Efforts are being made to give more intelligence to the web worms and automated indexing. Gary Taubes (1995) wrote an interesting article on this subject.<sup>5</sup>

*CERN’s WWW virtual library indexes.*<sup>6</sup> Another approach to identifying and organizing topical content guides to the web was initiated by CERN, the European

---

<sup>4</sup> The acronym stands for “Yet Another Hierarchical Official Oracle”. The URL is <http://www.yahoo.com/>.

<sup>5</sup> Indexing the internet. *Science* 269, September 1995, pp. 1354–1356.

<sup>6</sup> <http://www.vlib.org/>.

Organization for Nuclear Research. This effort was described as “using amateurs to index the web until we can get librarians to do it right.” Volunteers from around the world were enlisted to help categorize web information resources and provide useful links to them. Over the past several years, some of the host organizations, URLs and the focus of several of the index lists have changed. Nevertheless, they remain useful.

---

Bio Sciences	<a href="http://mcb.harvard.edu/BioLinks.html">http://mcb.harvard.edu/BioLinks.html</a>
Biotechnology	<a href="http://www.cato.com/biotech/">http://www.cato.com/biotech/</a>
Chemistry	<a href="http://www.liv.ac.uk/Chemistry/Links/links.html">http://www.liv.ac.uk/Chemistry/Links/links.html</a>
Earth Science	<a href="http://www.geo.ucalgary.ca/VLEarthSciences.html">http://www.geo.ucalgary.ca/VLEarthSciences.html</a>
Energy	<a href="http://www.crest.org/gem.html">http://www.crest.org/gem.html</a>
Engineering	<a href="http://www.vlib.org/Engineering.html">http://www.vlib.org/Engineering.html</a>
Epidemiology	<a href="http://www.epibiostat.ucsf.edu/epidem/epidem.html">http://www.epibiostat.ucsf.edu/epidem/epidem.html</a>
Statistics	<a href="http://www.stat.ufl.edu/vlib/statistics.html">http://www.stat.ufl.edu/vlib/statistics.html</a>
Technology transfer	<a href="http://www.nttc.edu/gov/other/tech.html">http://www.nttc.edu/gov/other/tech.html</a>
US Federal Government	<a href="http://www.lib.lsu.edu/gov/fedgov.html">http://www.lib.lsu.edu/gov/fedgov.html</a>
US Government Info.	<a href="http://iridium.nttc.edu/gov_res.html">http://iridium.nttc.edu/gov_res.html</a>
United Nations	<a href="http://www.undcp.org/unlinks.html">http://www.undcp.org/unlinks.html</a>

---

These indexes typically link to other pages or web sites, so that an ever-expanding set of possible sites of interest is the norm.

*The deadly lure of hypertext links (how to lose focus on your topic).* Hypertext links are so easy to click on and follow that the most likely outcome will be that the web user tracks information threads that are totally unrelated to the original topic of interest. The result can be serendipitous enlightenment or wasted time and effort.

Suppose the desired information is about solid waste landfill design in Denmark. If the search term provided to the Google<sup>7</sup> search engine is “solid waste landfill design”, the results of such a search will be

*Results:* About 77,400 *Search time:* 0.42 s.

Although many of the 77,400 reference links may be interesting, it would be fruitless to try to find information about Danish landfill design in this way.

A straightforward solution to avoid information overload is to refine the search term, thus “solid waste landfill design Denmark”. A recent Google search using this request provided these results:

*Results:* About 2760 *Search time:* 0.42 s.

Within the first 20 Google results was a link to the technical guidelines for siting and designing landfills in Denmark (as of 1995).

---

<sup>7</sup> <http://www.google.com>.

### VII.1.6. Examples of a few web pages that I have found useful

*Chemical Information Sources from Indiana University.*<sup>8</sup> A concise, useful “Bibliography on Chemical Hazards”<sup>9</sup> is also found there, as is a related “chemical safety or toxicology information” page. Educators may be interested in the “chemistry courses on the Internet” and “clearinghouse for chemical information instructional materials” pages. The page titled “chemical information sources (major tools or databases)” is a good reference tool.

*Hazardous Substances Research Centers.*<sup>10</sup> The US EPA has funded regional academic research centers that focus on problem aspects of hazardous substances. The regional center for the south and southwest portion of the US involves three universities, Louisiana State University, Georgia Tech and Rice University. Much of this research at these regional centers focuses on contaminated sediments. Through the “research briefs” link, I was able to find a useful summary of the research objectives on studies of sediments at Bayou D’Inde in Louisiana. The sediment contamination had resulted from hazardous wastes, and I have worked on the problems at this particular site.

*Enviro\$en\$e.*<sup>11</sup> This web site is a joint project of the US EPA, Department of Defense and Department of Energy. Aside from providing useful links to a variety of US governmental web sites, there are some extremely interesting special topic pages. One that provided me with some particularly useful information is the link to the “solvent substitution data systems (SSDS)” page. This allows queries of the solvent alternatives guide (SAGE), the hazardous solvent substitution data system (HSSDS) and the Department of Defense Pollution Prevention Technical Library. These tools can be important to persons seeking alternatives to hazardous solvents.

*Thomas.*<sup>12</sup> This provides searchable and downloadable access to US Congressional legislative bills. As a governmental bureaucrat, I find it very important to keep apprised of the current status of environmental legislation. Similar web resources are becoming available to track legislation in other countries.

*The US EPA Office of Solid Waste Home Page.*<sup>13</sup> This provides links to solid and hazardous waste regulations, technical requirements, case histories and technical resources. This page also provides a link to information on waste minimization issues, resources and technologies.

### VII.1.7. Some evaluations and conclusions

*The state of the art.* The web is now useful for many purposes, and surely it will continue to develop. One of the major limiting factors at this time is the bandwidth of communication lines for data transfers. Ten years ago, a 1200 baud modem seemed fast – and was quite

<sup>8</sup> <http://www.indiana.edu:80/~cheminfo/>.

<sup>9</sup> <http://www.indiana.edu:80/~cheminfo/12-01.html>.

<sup>10</sup> <http://www.hsrg.org/hsrg/html/ssw/>.

<sup>11</sup> <http://es.epa.gov/>.

<sup>12</sup> <http://thomas.loc.gov>.

<sup>13</sup> <http://www.epa.gov/osw/>.

usable for transmission of simple text files. But a 14.4 modem is about the bare minimum for hooking up to the web and at this speed, transfer of large graphics files is not feasible. Even the fastest phone line modems available are much slower than the Ethernet connection I have to a web server, which itself is hooked up to a high speed data connection to the outside world – and I am often dissatisfied with the speed of communications.

The rapid growth of the number of people using the Net, together with continuing enhancements to the HTML language will make the bandwidth problem even more critical. Several promising technologies already exist and are being further developed to help solve this problem. Some examples are faster data transmissions over existing copper phone lines, optical fibers or use of cable TV connections. It is, therefore, likely that greater communication bandwidths will become widely available in the near future. When that happens, the current level of information technology will begin to seem primitive and Dr Bush's vision of changes in the way people think may come closer to reality. Dr Bush was an optimist about the impact of such things on society. I hope he was right.

*Rapid improvement over a short term.* Candidly, I did not myself find the WWW to be a very useful information-gathering tool before about 1994. One reason was that I had been using a text-based web browser called Lynx, which provides a far more primitive interface than do current graphical browsers. The primary reason, however, was that most of the material that I now find useful was not on the web until quite recently. Even a couple of years ago, a great many web sites that I routinely use were not yet published or contained little information of interest.

Some of the electronic journals and discussion lists that were formerly available only via email are now available at web sites, but with the added advantages of better formatting and far better facilities for searching the archives of back issues. Even the concept of discussion lists is moving over to some web sites, with improved interfaces such as better thread management (i.e. the ability to follow discussions about a selected topic more easily).

As of now, a number of web sites exist with further communications and research in the environmental science and technology disciplines. I recommend them to the reader.

## References

- Bush, V., 1945. As we may think. The Atlantic Monthly (available in <http://lycos.com>).  
Taubes, G., 1995. Indexing the Internet. Science, 269, 1354–1356.

## Uniform Resource Locators (URLs) – the web sites

### *Information sources*

- Hazardous Substances Regional Academic Research Center (Louisiana State University, Georgia Technical University, Rice University: <http://www.hsrc.org/hsrc/html/ssw/>  
Indiana University – chemical information: <http://www.indiana.edu:80/~cheminfo/>  
Indiana University – Bibliography on Chemical Hazards: <http://www.indiana.edu:80/~cheminfo/1201.html>  
US Congressional Legislative Bills – Thomas web site: <http://thomas.loc.gov>  
US EPA – Department of Defense and Department of Energy, Enviro\$en\$e web site: <http://es.epa.gov/>

US EPA Office of Solid Waste Home Page: <http://www.epa.gov/osw/>  
Journal of Molecular Modeling: <http://www.ccc.uni-erlangen.de/jmolmod/index.html>  
W3C World Wide Web Consortium: <http://www.w3.org/>

### ***Web search engines (web searchers)***

Google: <http://www.google.com/>; Lycos: <http://lycos.com>  
WebCrawler: <http://webcrawler.com/>  
Altavista: <http://www.altavista.com/>  
Amazon: <http://www.amazon.com/>  
Scirus: <http://www.scirus.com>

### ***“Web worms” programs and indexing services***

Yahoo – Yet Another Hierarchical Official Oracle: <http://www.yahoo.com/>

### ***WWW virtual library indexes***

CERN, the European Organization for Nuclear Research: <http://www.vlib.org/>  
Bio Sciences: <http://mcb.harvard.edu/BioLinks.html>  
Biotechnology: <http://www.cato.com/biotech/>  
Chemistry: <http://www.liv.ac.uk/Chemistry/Links/links.html>  
Earth Science: <http://www.geo.ucalgary.ca/VL-EarthSciences.html>  
Energy: <http://www.crest.org/gem.html>  
Engineering: <http://www.vlib.org/Engineering.html>  
Epidemiology: <http://www.epibiostat.ucsf.edu/epidem/epidem.html>  
Statistics: <http://www.stat.ufl.edu/vlib/statistics.html>  
Technology transfer: [http://enviro.nfesc.navy.mil/erb/rel\\_sites.htm#a](http://enviro.nfesc.navy.mil/erb/rel_sites.htm#a)  
US Federal Government: <http://www.lib.lsu.edu/gov/fedgov.html>  
US Government Info.: <http://www.ntc.edu/resources/government/govresources.asp>  
United Nations: <http://www.unep.org/Documents.Multilingual/Default.asp?DocumentID=293>