# The National Agricultural Library's Database of Online Documents Covering Water and Agriculture

Joseph R. Makuch[1] and Stuart R. Gagnon[2]

## Abstract

Agricultural operations can have a negative effect on water quality. The *National Water Quality Inventory, 1998 Report to Congress* (the most recent biennial report in this series) indicates that agriculture is a major source of water quality impairment of assessed rivers and lakes in the United States.

Information is a key ingredient of problem solving. With the emergence of the World Wide Web, many documents produced by state and federal agencies, extension services and other organizations involved with water issues are being published and freely distributed online.

While water quality information may be available electronically, the information is scattered across many Web sites. In addition, the organization and access mechanisms among sites is not consistent. These characteristics make it difficult to rapidly locate and access a comprehensive collection of documents related to a particular water issue associated with agriculture, for example, the relationship between nutrient management practices and nitrate movement into groundwater. Internet search engines are not precise enough to easily locate specific documents.

To improve access to electronic documents covering water and agriculture, the Water Quality Information Center at the National Agricultural Library has developed a prototype database (located at http://www.nal.usda.gov/wqic/wqdb/esearch.html) of more than six hundred of these documents.

## The National Agricultural Library

The National Agricultural Library (NAL) was established as the U. S. Department of Agriculture (USDA) Library with the signing of the Organic Act in 1862. In 1962, the secretary of agriculture officially designated the Department Library as the National Agricultural Library (Fusonie, 1988).

NAL, located in Beltsville, Maryland, is one of four United States national libraries, along with the Library of Congress, the National Library of Medicine, and the National Library of Education. NAL is part of USDA's Agricultural Research Service.

NAL's mission is to ensure and enhance access to agricultural information to improve the quality of life. As the world's largest agricultural library, NAL is an important resource for people working on agricultural issues, including issues related to water and agriculture.

**Agriculture and Water Quality**

Due to the importance of water to agricultural productivity and the need to protect water quality from contamination caused by agricultural operations (U. S. Environmental Protection Agency, 2000; U. S. Geological Survey, 1999; Economic Research Service, 1997 and National Research Council, 1993), water-related documents are a key component of NAL's collection. AGRICOLA is the bibliographic database of agricultural literature produced by NAL and its cooperators (National Agricultural Library, 2000). A search (in September 2000) of AGRICOLA on the Web by subject category code P200 (water resources and management) showed 20,445 titles in the online public access catalog and 19,799 titles in the journal article citation index. Other AGRICOLA subject category codes that encompass water-related documents include codes for aquatic sciences (M000), soil conservation (J800) and irrigation and drainage (P210).

**The Water Quality Information Center at NAL**

Problem solving requires information. A variety of people—scientists, policy makers, economists, engineers and many others—are working on ways to curb water quality problems associated with agriculture. The role of NAL's Water Quality Information Center is to meet the information needs of these people.

The Water Quality Information Center was established in 1990 to support USDA's plan to address water quality concerns. As the focal point of NAL's water quality efforts, the center collects, organizes and communicates the scientific findings, educational methodologies and public policy issues related to water quality and agriculture (Makuch, 2000).

The Water Quality Information Center assists individuals looking for information on water quality topics related to agriculture by providing information via the World Wide Web at http://www.nal.usda.gov/wqic. Examples of water quality information available electronically include bibliographies; listings of upcoming meetings and Internet discussion groups covering water issues; and annotated links to water-related databases and Web sites providing information on topics such as wetlands, water quality models and funding sources for water quality improvement.

The center also manages Enviro-News—an Internet mailing list that provides timely environmental news to approximately five hundred subscribers. Makuch (1999) provides a more complete discussion of NAL's electronic information initiatives covering water topics.

An additional responsibility of the center is to improve NAL's water-related information resources: the center identifies and recommends additions to NAL's collection, forwards items to the library's gift and exchange program and works to strengthen water quality coverage in AGRICOLA.

## Growth in Online Publications

Begun in the late 1960s as a distributed, and thus less vulnerable, command system for military operations by the U. S. Department of Defense, the Internet was soon expanded to support scientific research (He, 1997). The World Wide Web, an extension of the capabilities of the Internet, was introduced by CERN (European Organization for Nuclear Physics, Geneva Switzerland) in 1991 to assist in data sharing activities (Abbate, 1999). Online documents had been available via the Internet prior to the implementation of the World Wide Web, but this new technology created an environment in which electronic communication, especially collaboration, could become more intuitive (Berners-Lee, 1996). The ability to hyperlink to specific resources coupled with significant graphical improvements, gives the World Wide Web its versatility and has promoted the Internet's increasingly available resources.

The phenomenal growth of the Internet may indeed be related to the ease of use that the World Wide Web provides. According to a May 2000 Nielsen survey, two-thirds of teenage and adult Americans are online (ACNielsen, 2000). However, with such an extensive audience, it is difficult to state how much truly authoritative information is available (Kibirige & DePalo, 2000). Possibly because of its public popularity, the authenticity and accuracy of any given document must be evaluated. Substantive water-related information, though, is made available by a variety of reputable sources.

With the emergence of the World Wide Web, many documents produced by state and federal agencies, extension services and other organizations involved with water issues are being published and freely distributed online. For example, in the United States, Web sites managed by the Environmental Protection Agency, the U. S. Geological Survey, the Natural Resources Conservation Service, and the Economic Research Service provide many full-text documents covering water quality and agriculture.

While water quality information may be available electronically, the information is scattered across many Web sites; in addition, the organization and access mechanisms among sites is not consistent. These characteristics make it difficult to rapidly locate and access a comprehensive collection of documents related to a particular water issue (such as nitrate contamination of groundwater). Internet search engines are not precise enough to easily locate specific documents.

## A Prototype System for Managing Online Documents Covering Water and Agriculture

Databases are repositories of *metadata*, which, in this case, is data describing and pointing toward some other information source. Metadata has been around for a long time. Libraries have used non-electronic metadata in the form of card catalog cards for almost a hundred years. One of the main benefits of using metadata is to collect representations of related items in one place.

To improve access to online documents covering water and agriculture, NAL's Water Quality Information Center has developed a prototype database of more than six hundred water quality documents (in September 2000). The database is accessible at http://www.nal.usda.gov/wqic/wqdb/esearch.html.

**Background of the Database**

Development of the Water Quality Information Center database began in 1998 as part of an initiative to add records for online publications related to water and agriculture to AGRICOLA. The AGRICOLA project is described in Makuch and Hamilton (1999).

As a way to manage titles for the project, Water Quality Information Center staff created a database using the bibliographic software, ProCite. This was software that staff were using to create electronic bibliographies for the center's Web site and so were familiar with it. As we began adding titles to the ProCite database, we realized that this was a unique resource that people working in the water quality field might find useful. We decided that we would continue using the ProCite database to manage the titles that were being selected for AGRICOLA, but we would also make the database publicly available via the Web. In this way, people would be able to easily locate and instantly access information on water and agriculture by searching a database that was specifically focused on the topic.

**Database Records**

Records in the database are simple. Each record contains eleven fields. They are

1. author
2. author affiliation
3. title
4. date sent to Cataloging
5. date added to AGRICOLA
6. date of access
7. location/URL
8. notes
9. abstract
10. call number
11. keywords

Fields four and five are specific to the management of the AGRICOLA project. The other field names are self explanatory. Because of time limitations, not all fields are completed for every record. Issues involving the "keywords" field are discussed below.

**Selecting Documents**

As we become aware of new online publications related to water and agriculture, we add them to the database. We learn of new electronic publications primarily by checking water-related Web

sites and through announcements on Internet mailing lists and in print newsletters. A title will be added if it

- covers water and agriculture or closely related field
- is freely available online
- displays in high quality
- is complete (not missing graphics that may be in a print version)
- is published by a known entity, such as a university or government agency

**Putting the Prototype Database on the Web**

We explored the use of Reference Web Poster software to make the database available over the Web. Reference Web Poster, produced by ISI ResearchSoft (which also produces ProCite), allows you to put ProCite (and other bibliographic databases) on the Web. We could not pursue this approach, however, because we did not have access to a Windows-based server required to run the software.

The initial solution we found was to work with Isearch—a freeware search engine that is used for NAL's Web site. Developed by, and available from, the Center for Networked Information Discovery and Retrieval (CNIDR) (located at http://www.cnidr.org/ir/ir.html), Isearch is based upon the Wide Area Information Server (WAIS) model (Nassar, 1997). We worked with staff from the library's Information Systems Division (ISD) to apply Isearch technology to our situation. An ISD staff member wrote Perl scripts to change ASCII text files—saved from the ProCite database—to consecutively numbered HTML documents. One script converts the ASCII bibliography to separate "records" with appropriate spacing. In the same process, HTML tags are inserted at the beginning and end of each record. The second Perl script breaks out (splits) all individual records into separate files, naming each with a new, consecutive record number (Thompson, M., personal communication, May 30, 2000). All updates made to the ProCite database are converted and appended in batches. After an update is made, the files must be re-indexed.

For this database, ISD uses only the Iindex and Isearch modules from CNIDR. Iindex is used to index a subset of the Water Quality Information Center's Web space where the records reside, while Isearch is set to search a specific set of indexes. A backup database is available through a system of maintaining two separate index sets on the site and switching back to the last working set if problems occur during the update (Thompson, M., personal communication, September 25, 2000).

The database search page (http://www.nal.usda.gov/wqic/wqdb/esearch.html) provides instructions and a form to enter search queries. We wrote a brief description of the database and provide links to other databases relevant to water resources and agriculture (http://www.nal.usda.gov/wqic/wqdb/aboutdb.html). Because this is a prototype system, we have not put a great deal of effort into system navigation or screen design.

Figure 1 shows a results screen. Clicking on the document title or filename will take you to the record for that publication. [Note: in the electronic version of this paper, the links are not live in figure 1.] For example, clicking on the first document title brings up the record shown in figure 2. Selecting the URL in the record shown in figure 2 accesses the publication. In this example the

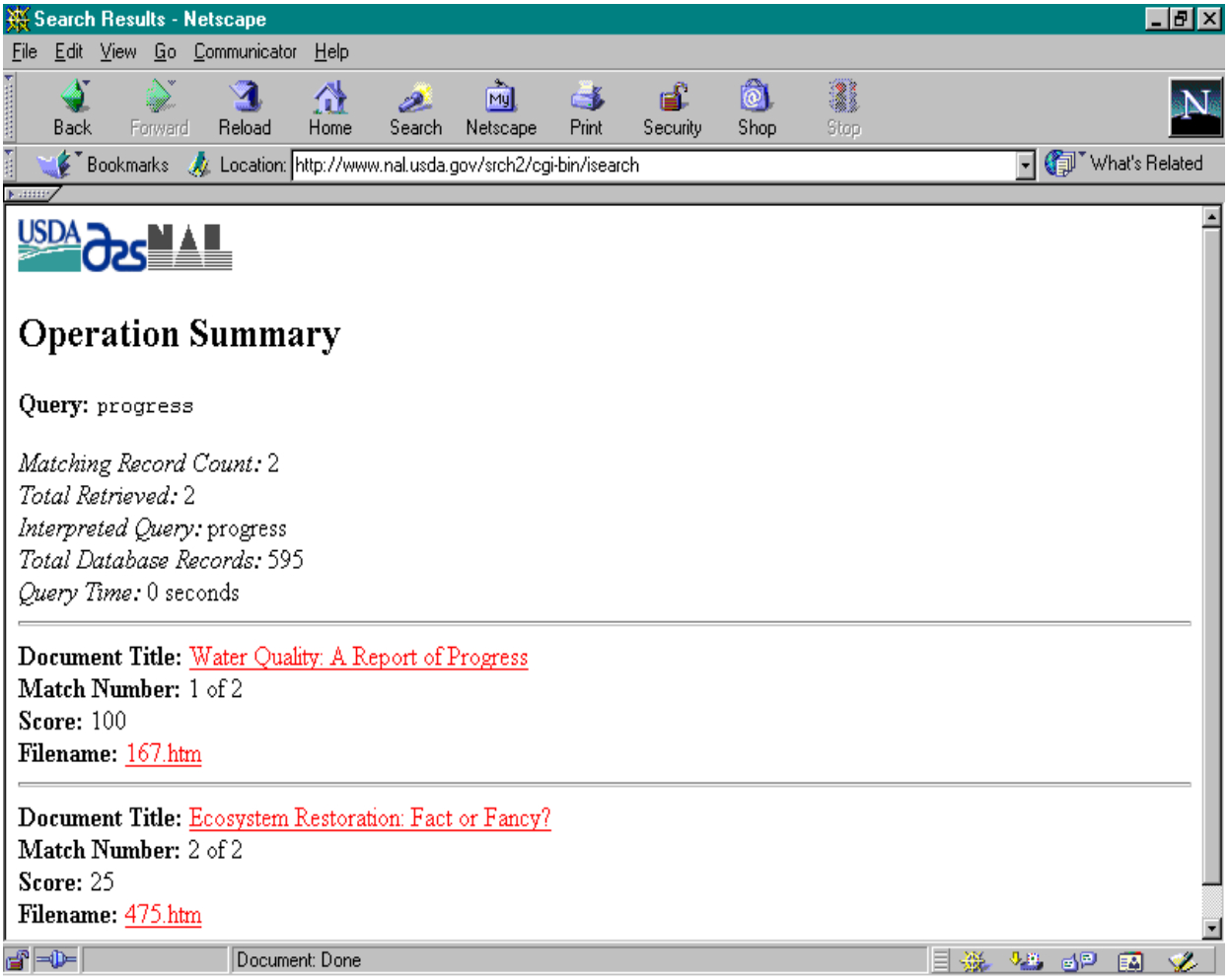


**Figure 1. Example Results Screen for the Database.**

publication resides on the NAL server. However, most of the publications are on other servers, i. e., on the server of the organization that maintains the publication. Also, note that only fields that contain data are displayed. In the example shown in figure 2, the "abstract" and "author" fields are empty, so they aren't displayed.

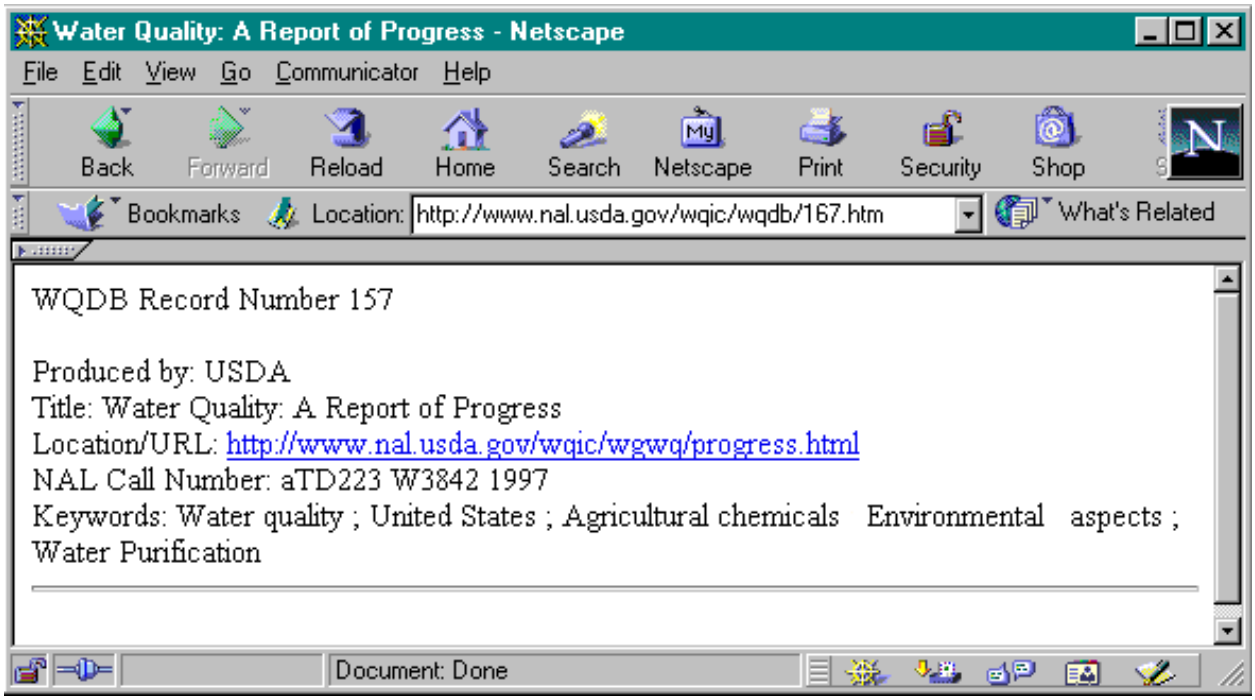


**Figure 2. Example Database Record.**

The keywords for the record shown in figure 2 were taken from the "descriptor" field of the AGRICOLA record for the publication. At one time we attempted to add keywords from AGRICOLA to records in the database to improve the precision of record retrieval by Isearch. However, we abandoned this effort for several reasons: not all of the database titles are in AGRICOLA, descriptors in AGRICOLA were not as specific as we required and a new thesaurus project at NAL offered an opportunity to get the level of specificity we need from a thesaurus that will be used in multiple library programs. The thesaurus will be discussed below.

**Database Usage**

We made the database available on the Water Quality Information Center Web site in May 2000. At this time we created a link from the Water Quality Information Center's home page to the database search page and allowed a shakedown period of public access prior to broadly announcing the new resource. In June 2000, the database's availability was announced to several water-related Internet lists: Enviro-News, Lakes-L and NPSINFO (nonpoint-source information).

Even though the Water Quality Information Center Web site is not file intensive, with few graphics or other server-counted material, it still averages nearly twenty thousand hits per month. In May 2000, prior to the electronic mailing list announcements, the database search page ranked 15$^{th}$ of all the center's pages for number of hits. We attribute this relatively high number to our placement of the new search link at the top of the center's homepage. Following the announcements, in its first three full months (June-August 2000) online, the database search page was one of the top five most requested items on the Water Quality Information Center Web site.

In the first two months, the page ranked second, after the home page. We will continue to monitor usage of the search page and plan to identify strategies for determining which records are accessed most frequently.

## Issues in Online Document Management

Maintenance, as it often is with any system, has been problematic with the database. Once new titles are located and evaluated, adding these to the ProCite database is fairly straightforward. However, any new titles added must also be added to the latest "update group" within ProCite for subsequent uploading to the Web database. Adding these new records to the Web version is complex and time consuming.

To upload new records from the ProCite database, we create a comma delimited file of the update group of records. This flat file is saved to the local network and then uploaded, via FTP, to the HTTP server. Through a Telnet session, we edit the Perl scripts to reflect the new file name and a new beginning sequence number. Then we run the Perl scripts, check the results and move the output to the index directory. We notify our ISD colleagues that our new update is live, requesting a re-index.

We have been uploading these new record groups on a more-or-less monthly basis. In September 2000, we uploaded fifty-eight new records.

Like the procedure for adding new titles, the procedure for revising records is not very efficient. Errors found in a record must be corrected in both the ProCite database and the Web version. As with adding new records, making corrections to records in the ProCite database is simple, but making the corrections in the Web database is more involved. For example, fixing a broken link in a record on the Web database requires us to identify the HTML file of the record with a search on the Web database and then use a Unix editor to make the change to the live record. We must also make the correction in the ProCite database so it too is up-to-date.

## Future Directions

Link management improvements and several other changes are being considered. Revising records is often required due to dead links that result when documents are moved to another directory on a server. In the prototype system, our strategies for locating dead links have not been as thorough as we would prefer. We currently rely heavily on random sampling of the database and hit-or-miss identification of problem links. In the future, we will be exploring ways to systematically use the library's link checking software to help ensure that all links to documents in the database are functional.

Berners-Lee (1996) stated that common standards permitted the growth of the Web. The increasing use of metadata standards will make searching for and identifying relevant resources much less difficult. We have begun mapping the database fields to Dublin Core, a recognized

metadata format. We plan to implement and be fully compliant with this and other standardized formats in future versions of the database.

One of the fields that maps between both sets of metadata is our "keywords" field. The corresponding Dublin Core field is "subject" (Dublin Core Metadata Initiative, 1999). We expect to populate this field with descriptors (subject words) from a new NAL thesaurus. Toward this effort, we are working to identify terms representing water-related concepts that are appropriate to describe resources contained in the database. WQIC staff, with colleagues from NAL's thesaurus team, have begun developing a water quality component to an NAL thesaurus which will be used with other NAL products. Thesaurus design work involves structuring each new concept into relationships with the existing thesaurus organization, as well as providing notes for usage and definitions. The water-related terms will we integral to the NAL thesaurus and used as separate concepts.

As we move toward implementing changes in field content and structure, we plan to make improvements to the database itself. Our experiences with the prototype have lead us to look at other ways to handle metadata. MySQL, an open source relational database management system, could add more flexibility to our product (MySQL AD, 2000). Importantly, the next version of the system must avoid the high maintenance requirements that are inherent in the prototype. The database must be easy to manage since the center's staff resources are limited. A high quality product must be delivered to end users while efficiently using staff time.

To provide another avenue for accessing the database, we will explore integrating the database into the Agriculture Network Information Center (AgNIC) Web site. AgNIC is an alliance comprised of NAL and land-grant universities and other agricultural organizations (AgNIC, 1998).

## Conclusions

In the form of a simple, cost-effective database, the Water Quality Information Center has created a prototype product offering water resources professionals rapid, desktop access to documents covering water and agriculture from a variety of sources.

The next step is to enhance the general functionality of the system while continuing to add content. A major improvement would be the addition of keywords to new and existing database records. However, this is a large task that requires more resources than the center currently has available.

As the prototype evolves into a more useful digital library, many questions related to access arise and will need to be addressed. Questions include: Who are the primary users? What will they seek and how will they seek it? How savvy are most of the users? What level of collocating functions do they seek? (preferring documents viewed by authoring organization or by subject-

oriented groupings?) (Tennant, 1999). The answers to these questions will help in designing a system that meets the needs of people seeking information on water resources and agriculture.

## References


1. Abbate, J. (1999, November). The electrical century: Inventing the Web. Proceedings of the IEEE, 87, 1999-2000.

2. ACNielsen. (2000, May 8). ACNielsen survey finds nearly two thirds of U.S. population age 12 or older are online [Announcement]. Chicago, IL: Author. Retrieved September 25, 2000, from the World Wide Web: http://www.acnielsen.com/news/american/us/2000/20000508.htm

3. AgNIC. (1998, July 21). AgNIC alliance governance document. Retrieved October 2, 2000, from the World Wide Web: http://www.cornell.agnic.org/agnic/about/index.html

4. Berners-Lee, T. (1996, October). WWW: Past, present and future. Computer, 29, 69-77.

5. Dublin Core Metadata Initiative. (1999, July 2). Dublin Core Metadata Element Set, Version 1.1: Reference Description. Retrieved September 28, 2000, from the World Wide Web: http://purl.oclc.org/dc/documents/rec-dces-19990702.htm

6. Economic Research Service, Natural Resources Division. (1997, July). Agricultural resources and environmental indicators, 1996-97 (U.S.D.A. Agricultural Handbook No. 712). Washington, DC: Author.

7. Fusonie, A. E. (1988, Spring). The history of the National Agricultural Library. Agricultural History, 62, 189-207.

8. He, J. (1997, September/October). Introduction of the Internet and World Wide Web. Experimental Techniques, 21, 29-33.

9. Kibirige, H. M. & DePalo, L. (2000, March). The Internet as a source of academic research information: Findings of two pilot studies. Information Technology and Libraries, 19, 11-16.

10. Makuch, J. R. (2000, June 8). About the Water Quality Information Center. Beltsville, MD: Author. Retrieved September 25, 2000, from the World Wide Web: http://www.nal.usda.gov/wqic/aboutwq.html

11. Makuch, J. R. (1999, July). Electronic sources of water resources information from the National Agricultural Library (Paper No. 993181). In 1999 ASAE/CSAE-SCGR annual international meeting: Emerging technologies for the 21st century, Toronto, Ontario, Canada, July 18-21. St. Joseph, MI: American Society of Agricultural Engineers.

12. Makuch, J. R. & Hamilton, E. P. (1999, December). Improving access to water resources and agriculture publications on the World Wide Web. Presented at the American Water Resources Association, Symposium on Water Resources and the World Wide Web, Seattle, WA. Retrieved September 18, 2000, from the World Wide Web: http://www.awra.org/proceedings/www99/w09/

13. MySQL AB. (2000). What is MySQL? Retrieved September 28, 2000, from the World Wide Web: http://www.mysql.com/information/

14. Nassar, N. (1997, May). Searching with Isearch: Moving beyond WAIS. WEB Techniques, 2, 51-55. Retrieved September 22, 2000, from the World Wide Web: http://www.webtechniques.com/archives/1997/05/nassar/

15. National Agricultural Library. (2000). What is AGRICOLA. Beltsville, MD: Author. Retrieved September 18, 2000, from the World Wide Web: http://www.nal.usda.gov/general_info/agricola/agricola.html

16. National Research Council, Board on Agriculture, Committee on Long-Range Soil and Water Conservation. (1993). Soil and water quality: An agenda for agriculture. Washington, D. C.: National Academy Press.

17. Tennant, R. (1999, April 15). Access: Where the rubber meets the road. Library Journal, 124, 32-34.

18. U. S. Environmental Protection Agency, Office of Water. (2000). National Water Quality Inventory 1998: Report to Congress (EPA/841-R-00-001). Washington, DC: Author.

19. U. S. Geological Survey. (1999). The quality of our nation's waters: Nutrients and pesticides (U.S.G.S. Circular 1225). Denver, CO: Author. Retrieved September 27, 2000, from the World Wide Web: http://water.usgs.gov/pubs/circ/circ1225/

# # #

Submitted October 3, 2000.

Disclaimer: The use of trade, firm, or corporation names in this paper is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by

the United States Department of Agriculture, Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

1. Joseph R. Makuch, Ph.D.
Coordinator, Water Quality Information Center
National Agricultural Library
10301 Baltimore Ave.
Beltsville, MD 20705
Phone: 301/504-6077
Fax: 301/504-7098
Email: jmakuch@nal.usda.gov
URL: http://www.nal.usda.gov/wqic

2. Stuart R. Gagnon, M.S.L.S.
Librarian, Water Quality Information Center
National Agricultural Library
10301 Baltimore Ave.
Beltsville, MD 20705
and
McKeldin Library
University of Maryland
Phone: 301/504-6077
Fax: 301/504-7098
Email: sgagnon@nal.usda.gov
URL: http://www.nal.usda.gov/wqic