

Second WaterWeb Information Summit

Fort Lauderdale, Florida

October 3-6, 1999

Report from Information and Data Quality Working Group

Group Members

Jonathan Barney	Kevin Merrifield
Len Berry	Lenin Montano
Terry Dodge	Jeanne Moody
Dennis Dunn	Xiaosong Wang
Mantha Mehallis	

The Information and Data Quality Working Group was asked to consider ways to improve the quality of water-resources information and the quality of water-resources data. Because most members of this group are computer and/or information specialists rather than experts in water resources, it was generally acknowledged that as a group we were not qualified to make recommendations about the quality of water data per se. Instead, we decided to focus on our areas of expertise. Specifically, the committee's discussions and recommendations fell into three areas:

1. web site design parameters for effective information and data retrieval
2. information about the quality or reliability of data in online databases
3. recommendations for the WaterWeb ring.

WEBSITE DESIGN RECOMMENDATIONS

These guidelines were made with the realization that not everyone has the latest computer hardware, software, and telecommunications. Telephone connections to the internet are often slow or not readily available at all times. This may be particularly true in rural locations of many countries. Access to online water resources databases should not be significantly hindered by not having up-to-the-minute technology.

● **Essential elements for web sites**

- Site map and/or index to contents
- E-mail contact information for comments or questions
- Return to Home Page link on each page

WEBSITE DESIGN RECOMMENDATIONS (continued)

Site maps should have as much information about contents as possible on the first level. Users should not have to drill down through several layers of links or menus to determine whether or not their desired information is available at the given web site. Indexes to site contents should have cross-references to preferred terminology (e.g., Groundwater hydrology, *use* Hydrogeology) instead of making the user guess which synonym or spelling variant is being used. Return to Home Page links are especially helpful for large sites where navigation from one page to another is complex.

- **Web sites should avoid the use of frames**

Frames do not allow the user to readily see the URL of the site being visited. Also, older browsers will not handle the frames format.

- **Page dimensions**

- Width should not exceed 600 pixels
- Home page should not exceed two screens in length

Many computers, particularly laptops, still use 15-inch monitors. Users should not have to scroll sideways as well as up and down to read information on the screen.

- **Optimize page loading speed**

- Loading time should be relative to the importance of the information on the page
- Maximum page size should fall within the range of 30K to 60K or less
- Graphic image sizes should be 10K or less
- Keep the number of graphic images on a page to a minimum
- Use thumbnail images that can be enlarged if desired
- Make a text-only version available for those with low-speed connections.

In designing web sites, designers should try to optimize the loading time of each page. It is understood, however, that some forms of information, especially graphic images, take longer to load than others. Lengthy waits for substantive data, graphs, maps, photographs, or other information to appear on the screen are often unavoidable and are generally perceived as part of the price of access to the needed information. Waiting several minutes for cute animated images, corporate logos, and purely decorative graphics to appear is generally considered a nuisance and may well put the user in a negative frame of mind about the web site and the producing organization. With this in mind, the committee recommends that the maximum page size should be less than 60K and that, if possible, graphic image sizes should be less than 10K. If a site needs to have numerous graphic images, the designer should consider using thumbnail images that can be enlarged if desired and/or spreading the number of images over several pages.

WEBSITE DESIGN RECOMMENDATIONS (continued)

A text-only version for those with low-speed telephone connections should be made readily available. Ideally there will be an option for text-only specified on the Home Page. Text-only can be provided by a script that will strip off tags.

● **Metatags**

- Metatags for Author, Description, and Keywords are essential minimums
- Use Dublin Core metadata element set for metatags

Other metatags to facilitate searching should also be included, but the committee is not recommending specific metatags for such at this time. The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. The building of an interdisciplinary, international consensus around a core element set is the essential feature of the Dublin Core. For further information on the Dublin Core Metadata Initiative, consult the home page at <http://purl.org/dc>

● **Keywords**

Keywords to describe the site as a whole and to link from the WaterWeb ring should consist of broad categories assigned by WaterWeb from a standardized list. Keywords for linking to subpages, database contents, e-text documents, etc. should be assigned by the web site owner from a standardized thesaurus or authority list of keywords. A task force could be appointed to examine various English-language thesauri for water resources and environmental topics and recommend one for preferred use. For foreign-language users and/or web sites, a polyglot dictionary is suggested (either in print copy or online) for determining specific keywords to use in searching for water-related information or data. Some research needs to be done to learn what is already available that could be recommended for general use by the water resources community.

● **Web pages should be tested on a variety of browsers**

INFORMATION ABOUT DATA QUALITY

All online databases should include information on the reliability or quality of their data.

● Essential data information elements

- Data source
- Contact information for the originating source of the data
- Data collection methodology, including
 - Whether data elements were measured or derived
 - Analytical methods
 - Detection limits
- Geographic coverage
- Earliest date covered
- Whether historical data are held constant or subject to change
- Active/inactive indicator
- Sampling frequency

From presentations and discussions during the general sessions, it became apparent that the quality of data can range from meeting highly detailed criteria to “any data are better than no data” depending on the use to which it is being put. Therefore, it is extremely important to be able to ask questions of the originator of the data. A prefatory statement should, above all, include the data source and contact information for the actual office or department that collected the data. If the data collection or analysis was outsourced to another agency or organization, then contact information for the contractor should also be given.

Information about geographic coverage and the earliest data collection data are also necessary to determine whether or not a database is likely to be useful. Historical records can be held constant, or they can be changed for political reasons, to correct recently discovered errors, etc. Because changes to historical data can affect the results of comparative analyses made from the database at different times, the database description should also include information on the constancy of historical data. Some metadata elements (such as active/inactive indicators or sampling frequency) can be indicated within the database itself instead of in the overall description.

INFORMATION ABOUT DATA QUALITY (continued)

● Water Data Rating Scale

Databases could be self-rated on a five-point scale according to the degree to which certain criteria (e.g., sampling frequency, collection method) have been met. This “five-star” rating would be analogous to hotel ratings which indicate only that a given property has certain amenities like restaurants available but imply nothing about the quality of those amenities. A water data scale would, for example, indicate that certain collection procedures were used without implying that all instruments functioned perfectly all the time or that all technicians were always conscientious about following specified procedures. A rating scale would give the potential user a quick indication as to whether or not it would be worth pursuing the use of the particular online database. It is conceivable that two types of rating scales would be needed: one for measured data and one for data derived from laboratory analyses. The Working Group on Information and Data Quality does not feel qualified to carry this idea further. We recommend that a task force of water resources professionals be appointed to develop the idea.

WATERWEB RING RECOMMENDATIONS

- Add pages stating the above guidelines
- Provide templates for water- resources web site designers
- Add links to standards for data collection and presentation (e.g., GEMS, Dublin Core)
- Restrict Ring members to noncommercial organizations
- Include links to other water resources web rings
 - Non-English language web sites
 - Web sites of commercial corporations

To help ensure the quality of information on the WaterWeb Ring, designers of web sites should know what is expected or at least desired. A site should be established on the Ring for design criteria and aids. Links to sites for data standards should be included. If feasible, full text information should be provided on the site. These links could also include thesauri for developing site indexes and for searching for information internally on the site.

Multilingual access to the Ring was discussed at length, but there is no politically sound way to determine which languages should be given preference in a web ring dealing with water resources on a global as well as national and local scales. English should be kept as the standard language for access to information on the Ring. It is believed that polyglot glossaries are available for those needing assistance in finding search terms in languages other than their own.

WATERWEB RING RECOMMENDATIONS (continued)

Opening the WaterWeb Ring to commercial enterprises would mean that the sites of some government agencies and possibly some nongovernmental organizations could not be on the Ring because of regulations forbidding too-close association with a commercial endeavor lest the association be construed as official endorsement of a commercial product or service. On the other hand, many commercial sites may have information of value to the water resources community. It was felt that the best solution to the dilemma is to set up a separate web ring for commercial sites with links between the two rings.