Transforming <Data> to Information

Setting the Context for Environmental Data

***Richard P. Hooper** – CUAHSI*

The Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) is developing a comprehensive multiscale data access and analysis system, known as the Hydrologic Information System (HIS). HIS consists of multiple components, including a single portal to access data from multiple providers, a web services library to imbed data within user-specified applications, a "digital watershed" to integrate different data types, and modeling tools to permit the development of a community modeling framework. These components are in various stages of development (one of which, the Hydrologic Data Access System, is described more fully on pages 16-17). The overall goal of HIS is to provide scientists greater data access, thereby accelerating research in hydrologic science.

In the process of developing this system, we came to realize that simply delivering data and metadata is not sufficient. A context must also be provided. This ranges from providing the physical setting for data (Is this an arid or humid landscape? What is upstream or downstream of this station?) to providing an interpreted data product that transforms the raw data into a more useful form. The advantages of contextualization go beyond the hydrologic sciences and are applicable to environmental data in general. The effort expended to provide this context is well worth it because, if done well, the data will be transformed to information that will be useful to a broader range of stakeholders, decision-makers, and citizens.

Examples from USGS

For hydrologic data, the U.S. Geological Survey provides some excellent examples of contextualization. Consider its "WaterWatch" feature shown on the next page. Each dot represents the current conditions at a real-time stream gauge expressed as a percentile of historic discharges for the date. Only stations with at least 30 years of record are displayed on the map so that the percentiles will be stable. Coding of the percentiles is designed to highlight extremes: the wettest and driest deciles of discharge.

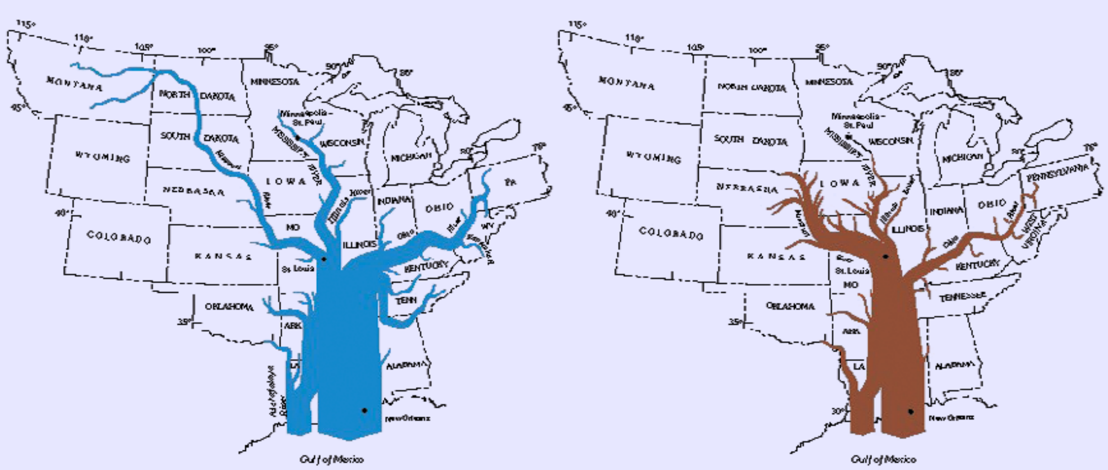
This display very effectively conveys which areas in the country are dry and which are wet in a manner readily interpreted by ordinary citizens. Experts can quibble about the statistics (for example, serial correlation is ignored because each day is effectively treated as independent of the next), but such issues are clearly secondary.

Another example from USGS contrasts the sources of sediment and water in the Mississippi Basin (Meade et al., 1995). The viewer can see at a glance (below) that water comes primarily from the Ohio River while sediment comes from the Missouri River. The data are spatially contextualized and the concept of the connectedness of a river basin is reinforced.

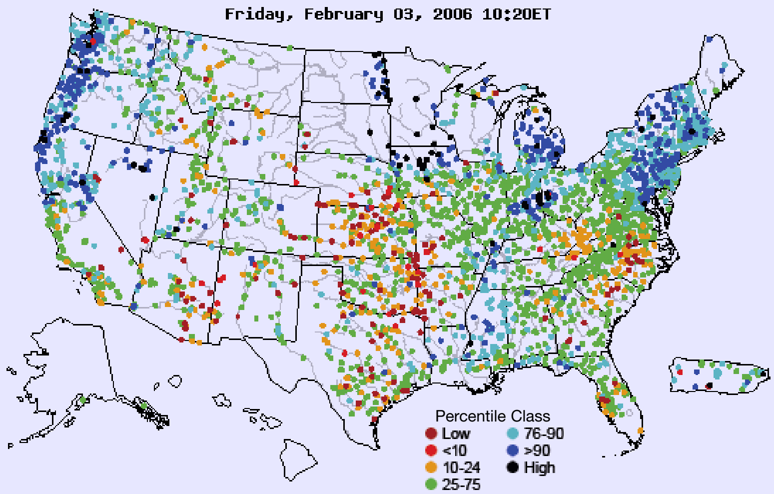
These examples should challenge us to develop effective presentations for our data that place them within a context. And these displays just scratch the surface. Using the same visual strategy as the Mississippi maps (i.e., the width of the line indicating discharge), consider animating the data for daily discharge. At the scale of the Mississippi, rainstorms would appear as bulges making their way down the river, much like a rat going through a snake. Chemical concentrations could also be shown using color so that, for example, spills could be depicted as they disperse and move through a river network. The possibilities are numerous.

Relevance for Different Users

Moving beyond simply environmental or regulatory contexts, we are finding that interpreted data products are necessary even for audiences of research scientists. Consider precipitation: specialists may be interested in the data originating at individual precipitation stations or in the raw radar reflectance values used to estimate precipitation rates. Most environmental scientists,



Sources of water (left) and sediment (right) in the Mississippi River (from Meade et al., 1995).



The USGS WaterWatch display (water.usgs.gov/waterwatch/) depicts real-time streamflow compared to percentiles of historic daily streamflow.

however, are simply interested in how much it rained over a particular area during a particular time. Clearly the raw data must be averaged, combined, or manipulated in some manner to develop this estimate, but the appropriate technique depends on many factors: the temporal resolution, terrain characteristics, and many others. How do we develop the "expert system" required to make this calculation? Alternately, if some estimates are provided, how can they be sufficiently documented so that any user understands their limitations? Similar issues exist for many data series, such as groundwater levels, where the data are measured at a point or over a small "footprint," but a spatially extensive property must be estimated.

There are no simple solutions to this challenge. Knowledge of the local system, considerable technical expertise, and a fair amount of creativity and effort will be required to develop the contexts for effective data delivery. Progress will come incrementally as monitoring programs increasingly develop displays that automatically update, such as USGS WaterWatch, or computer models that convert data to a more useful, visually meaningful product.

CUAHSI's Approach

Within the CUAHSI community, HIS is being designed to support teams of scientists organized around river basins or aquifers. Part of the charge to these groups will be to develop a "Basin User's Manual" to provide the environmental setting of the basin, including geology, soils, cultural development, and land-use history. Another product will be the construction of water, energy, and chemical budgets for the basins that will require construction of "benchmark" data sets of stores and fluxes that will be accessible to the hydrologic community. With this incremental approach, we hope to develop a set of representative basins to be used by researchers as well as to provide examples that can be replicated in other river basins around the world.

Reference

Meade, R.H. (ed.), 1995. Contaminants in the Mississippi River, 1987-92: U.S. Geological Survey Circular 1133, pubs.usgs.gov/circ/circ1133/