

Embedding metadata in the data: An integrated approach

Marc Vayssières

California Department of Water Resources, 2351 S Street, Sacramento, CA 95816

Biographical Sketch of Author

Marc Vayssières is an Environmental Scientist for the California Department of Water Resources. He works in the Environmental Monitoring Program (EMP) for the Sacramento-San Joaquin Delta, Suisun Bay, and San Pablo Bay conducted under the auspices of the Interagency Ecological Program (IEP). The EMP was initiated in 1971 to monitor water quality, phytoplankton, zooplankton and benthos. Marc's specializes in the management, analysis and reporting of these long term monitoring data.

Abstract

Addressing ecological questions such as the cumulative impacts of human activities on water resources and aquatic communities often requires integrating data from several different monitoring programs. Such integration is not possible without good metadata. We propose that that embedding much of the metadata with the data, in a self-documenting database, is a flexible and advantageous approach to maintaining metadata.

In essence, metadata need to answer these basic questions: who, what, when, where, why, and how. But often the user of the data will also need to know the answer to combinations of these questions: e.g. what-when, where-when, and what-how-when.

Traditionally, metadata are stored in separate text documents describing the databases. In our experience such an approach leads to: a lack of synchronization between data and metadata, extra work to conform to "standards" and to maintain cross-tabulations such as what-when or what-where, and a tendency to treat the metadata as an afterthought. Embedding metadata in the database may begin with adding a few extra columns of information to tables. It is more thorough when it involves structuring of a database to store monitoring data in distinct time series that explicitly carry data attributes such as units, method of analysis, equipment used to acquire the data, etc. Queries, including cross-tabulations of these attributes, can then be used to produce and update metadata documents. Examples from long-term (25-30 years) monitoring databases of water quality and benthic invertebrates in the San Francisco Upper Estuary are used to discuss the advantages of metadata embedding.

* * * * *