

Database Management

Drinking Water-Constituents of Concern Data Management-Draft Part 1

Background

This section of the work plan discusses the data management system that will , display, and retrieve drinking water constituents data. Management of these data will entail loading drinking water constituents monitoring data into an existing relational database with an already-created "data browser" and mapping interface to query and display the data over the World Wide Web. This database and browser system has been created for the Interagency Ecological Program (IEP) by a cooperative effort between the Department of Water Resources (DWR) and the California Urban Water Agencies (CUWA).

Drinking water constituent data will be combined with a large number of other analyte and biological data using standardized naming conventions, normalization, geo-referencing and common units. These data will be accessible through a stand-alone interface allowing researchers to query for a full range of potential dependent and independent variables. The primary benefit of this project is not only to provide access to drinking water constituent data, but also to integrate these data with the other data needed to conduct research and develop indicators relevant to the ecology of the Bay/Delta and their tributaries, such as the data collected by other CALFED programs including the ERPP. Data integration can help resolve the ambiguities associated with multiple sources and dramatically reduce the time needed to prepare data for analysis. An integrated database system can substantially save staff time and improve productivity.

The advantage of this approach is that the database design has been carefully planned and proven in production, and the extensive software development for the data browser and mapping (GIS functionality) interface has been performed. This represents a considerable cost savings compared to having to create a new system. The addition of the drinking water constituent data will require relatively minor modifications to the existing database design and browser/mapping interface software.

The Bay/Delta and Tributaries project receives funding from the IEP, CVPIA/CAMP program, and SRWP and shares resources with CUWA. The project's principal function is to integrate the data collected by multiple agencies for comprehensive analyses. The project is currently in operation, holding data from various monitoring programs including UCD, several DFG offices, USFWS, stakeholders, private consultants and other agencies. These data can be accessed by all interested parties using a map interface via the World Wide Web.

The existing data management system already has most of the features needed to implement this project and thus this proposal would avoid duplication of effort or "reinventing the wheel". The benefits to CALFED, are reduced cost, collaboration among several of the CALFED agencies and stakeholders, and the experience gained by these agencies while implementing similar projects.

Data management is important to all aspects of CALFED data collection and dissemination processes. Ultimately, CALFED must make data and information readily accessible to CALFED Bay/Delta technical staff, agency staff and stakeholders. Data will also need to be updated regularly to meet different program reporting timelines in a way that allows information from one program to be related to another, and, in time, to modify adaptive management strategies.

Using a relational database management system (RDBMS) to manage data assists with data analysis because of its unique ability to relate diverse types of information. The diverse types of data already stored on the system include physical and chemical data (e.g., water quality, hydrodynamics, meteorological, etc.), biological data used to gauge

the health of the estuary. Beyond determining cause-and-effect relationships between the physical and chemical and biological data, technical staff can use these data to examine relationships to evaluate impacts of various alternatives. Monitoring key system attributes (or indicators), completing focused research to obtain better understanding, and phasing implementation based on information gained are all central to the adaptive management process. An information system that provides technical staff the ability to relate and query data from all these different data types simultaneously would facilitate analysis and reporting efficiency.

An RDBMS structures data into relational tables and provides the platform from which data can be transferred onto the file server, and accessed immediately via the Internet. Users would then be able to do further queries based on specific criteria, such as location, date and time, data type, and a range of values contained within the data set. Maps, from which a user can select a location and receive data at or in the vicinity of that location, have already been implemented. Internet technology, such as image mapping, was linked with existing database technology, creating a query based on location and efficiently transferring data to the user. This technology fits easily into the growing use of GIS technology in federal, state and local agencies.

In addition to efficient data retrieval, users would be able to query diverse data types. The diverse types of drinking water constituent data (e.g. water quality, hydrodynamics, and meteorological data) would be placed into relational tables. Information within those tables would be related according to key fields, such as location, date and time, data type, and other fields. Users would be able to retrieve all related data based on a single query in a matter of seconds, compared to the hours or days it might take trying to obtain data from various sources. Data outputs would be accessible using a PC, Macintosh, or workstation, using an Internet browser as an interface. Through their browsers, users could perform simple and refined queries, obtaining results quickly and efficiently.

Table 1: Tasks and Descriptions

Tasks	Costs
<i>Component 1. Modify browser as needed</i>	
1. Meet with CALFED, Agency and Stakeholder staffs to determine if any browser modifications are needed	\$2,900.00
2. Conduct workshop	\$3,000.00
<i>Component 2. Add data to the database</i>	
1. Add monitoring data generated in the from the other Part 1 tasks	\$5,000.00
2. Add data as required to support baseline analysis	\$5,000.00
Total Budget	\$15,900.00

Component Details

The costs shown in Table 1 are greatly reduced by using an existing data management system. In Component 1 we will meet with CALFED, agency and stakeholder staffs to identify any modification to the database browser that may be needed to meet the needs of data users. This meeting will revolve around reviewing the data and identifying the requirements of the potential users. The deliverable from this effort will be a document identifying requested changes in a memorandum that will provide a cost and schedule for any needed changes. CALFED, agency and stakeholder staffs will also be trained on how to obtain data from the system.

In Component 2, monitoring data will be loaded into the database that are produced from the other tasks in Part 1 of this proposal. The costs estimate assumes that these data have been checked by the individuals who are providing them, and are in digital format. In addition, other historic monitoring data needed for this project will be loaded. The specific costs are only estimates, since the volume and amount of the produced and needed data are not known.