San Joaquin Valley Drainage Authority

San Joaquin River
Up-Stream DO TMDL Project
ERP - 02D - P63
Final Draft

Task 11

Final Report

March 25, 2008

Karl Jacobs

Department of Water Resources
Interagency Information System Services
Division of Environmental Services
Department of Water Resources
901 P Street
Sacramento, CA 95814

Abstract/ Introduction

Data associated and relevant to this study were made accessible on the web through the BDAT/CEDEN information systems and/or were placed into a local MS Access database based on the Surface Water Ambient Monitoring Program (SWAMP) where they could be managed locally by researchers.

Three principal deliverables were provided for task 11:

1) Loading and integration time series data relevant to the study into BDAT and its display on the web,
2) Loading and integration of discrete water quality collected by the study on the web and
3) Development of a local database to manage data collected during the study, based on the SWAMP system, and providing a linking mechanism to CEDEN.

<u>Loading and integration of time series data relevant to the study into BDAT and its display on the web</u>

DWR staff loaded data from USGS, Bureau of Reclamation, and DWR automated monitoring systems. Regular time series data provided by Will Stringfellow's group were also incorporated into the BDAT system. These data are now available on the web. A list of monitored parameters are listed in Table 1.

<u>Loading and integration of discrete water quality collected by the study on the web</u>

Discrete data were loaded into the BDAT system and included measured or calculated values for 43 analytes. These data were provided in Excel spreadsheets which were used as a record source and to map to corresponding tables and fields in BDAT's underlying relational database.

DWR staff integrated these data into BDAT by matching various measured parameters and other types of information to those already existing in the system. When an exact match did not exist, a new record was assigned. Using this methodology, data from multiple sources can be standardized; this enables BDAT users to retrieve multiple sets of data using a single set of naming conventions.

<u>Development of a local database to manage data collected during the study, based on the SWAMP system, and providing a linking mechanism to CEDEN</u>

Field monitoring data collected from the monitoring programs specific to this project were loaded into a desktop database utilizing the Version 2.2 relational database template developed by the Department of Water Resources and customized for the State Water Resources Control Board's Surface Water Ambient Monitoring Program (SWAMP). SWAMP standards were followed to the extent possible based on the data provided. The database is equipped with data entry and retrieval forms, report generators, a permanent and temporary side to help facilitate QA/QC processing and a module to transfer data to the BDAT system.

Over 40 analytes were loaded into the database using standard query language and other tools. These data were provided in Excel spreadsheets which were used as a record source and to map to corresponding tables and fields in the underlying relational database. Primary look up information, such as analyte names, laboratory methods, stations, etc were cross walked to those provided by SWAMP. Where matches could not be found, new records were added to theses look up lists.

Methods

Loading and integration of time series data relevant to the study into BDAT and its display on the web

Time series data were provided to DWR staff in Excel spreadsheets which were used to map to corresponding tables and fields in BDAT's underlying relational database.

Parameters mapped into BDAT included matrices (substrates), methods, fractions, and units, the specifics of which were also aggregated within BDAT for purposes of enabling data users to rapidly evaluate data, analyze results, and validate comparisons. The same analytes were not collected at each station, but all available data was included in the data base.

Sampling stations were integrated into BDAT by locating them on a map based on information described in the incoming data sets, and comparing them to stations that have already been identified and in use in BDAT. To matches we assigned existing BDAT standard codes while creating new codes for the remaining stations. Staff were able to glean information about the equipment used at each station from the supporting documentation that was provided to us and likewise integrated that information into BDAT.

When loading the data into the database, staff created a unique record identifier that grouped the station, project, type of data collected, replicate number, and type of result into a single record. The unique identifiers were then propagated to their associated measurement records. Storing data using this methodology conforms to the relational database model and greatly enhances the ability to manage and integrate these data and make possible modifications to them over time.

Will Stringfellow's group provided information on how these data were collected and specifications for which data fields were desired for display on the website.

Loading and integration of discrete water quality collected by the study on the web

Discrete water quality data collected for this study were loaded into BDAT/CEDEN and made accessible on the web. Sampling stations were evaluated for inclusion into BDAT by locating them on a map, based on information described in the incoming data sets, and comparing them to stations that have already been identified and in use in BDAT. If matches were found, staff assigned existing BDAT standard codes while creating new codes for the remaining stations.

The next step was determining the constituent method codes and sample codes. Constituent method codes are generated using the above values after determining whether they are not already present in the database. Samples codes are obtained using the station code, project code, gear (type of data collected), replicate number, and sample purpose code, inserting them into the BDAT samples table, then reading the subsequent serial ID. Again, new codes were created only if the current values in the database could not be matched to the study's data.

Other data items included in the mapping process were matrix (substrate), method of analysis, analytes measured, fractions, units of measurement, time interval between measurements, period aggregation (value is averaged, instant, highest or lowest), and result type (how the value was obtained). Storing data using this methodology conforms to the relational database model and greatly enhances the ability to manage and integrate these data and make possible modifications to them over time.

Will Stringfellow and his staff were consulted to clarify how these data were collected and which data fields were desired for display on the website. In addition, Will and his group were able to clear up minor discrepancies found in the provided spreadsheets

Random checks of the files loaded into BDAT showed the data loaded without errors onto the website and are available for query. Discrete water quality parameters loaded into BDAT can be found in Table 2.

Development of a local database to manage data collected during the study, based on the SWAMP system, and providing a linking mechanism to CEDEN

Discrete water quality data collected for this study were loaded into a local SWAMP comparable database to help provide compatibility with statewide efforts to manage and integrate data and provide a tool for the researchers of this study to better manage their data.

Sampling stations were evaluated for inclusion into SWAMP primary look-up tables by locating them on a map, based on information described on the description and latitude and longitude of the incoming data sets, and comparing them to stations that have already been identified and used by the SWAMP. If matches were found, staff assigned existing SWAMP standard codes while creating new codes for the remaining stations.

Other data items included in the mapping process were matrix (substrate), method of analysis, analytes measured, fractions, units of measurement, time interval between measurements, period aggregation (value is averaged, instant, highest or lowest), and result type (how the value was obtained). Storing data using this methodology conforms to the relational database model and greatly enhances the ability to manage and integrate these data and make possible modifications to them over time.

Constituent codes were generated, as needed, using the analyte, method, unit, matrix and fraction codes after determining they are not already present in the database. Samples codes were obtained using the station code, date, time, project/agency code, event/sample type codes (type of data collected) and replicate number, inserting them into the database's samples table, then reading the subsequent serial ID. Again, new codes were created only if the current SWAMP values in the database could not be matched to the study's data.

Will Stringfellow and his staff were consulted to answer question regarding the source data set. Random checks of the files loaded into database showed the data loaded without errors into the database and onto the website and are available for query.

Results

Loading and integration time series data relevant to the study into BDAT and its display on the web

Time series data associated with the study are available as a distinct data set at: http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category_preselected_Projects.php?category_code=13&project_code=279 and be queried based on location, analyte collection date or survey type. Data can be downloaded in html, Excel or text formats. In addition, data from this study have been fully integrated with the other time series data available on BDAT. These data can be obtained from http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category.php under the "Time Series" link.

Loading and integration of discrete water quality collected by the study on the web

Discrete monitoring data are now available on the Web as a distinct data set at: http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category_preselected_Projects.php?category_code=16&project_code=279 and can be queried based on location, analyte collection date or survey type. Data can be downloaded in html, Excel or text formats. In addition, data from this study have been fully integrated with the other water quality data available on BDAT. These data can be obtained from http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category.php under the "Water Quality" link.

Development of a local database to manage data collected during the study, based on the SWAMP system, and providing a linking mechanism to CEDEN

The loaded database was turned over to Will Stringfellow and his staff for review, checking and comment. Revisions were provided by DWR. This database can be linked to BDAT/CEDEN as a distinct or integrated data set and made available at http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category_preselected_Projects.php?category_code=16&project_code=279 and can be queried based on location, analyte collection date or survey type. Data can be downloaded in html, Excel or text formats. In addition, data from this study can be fully integrated with the other water quality data available on BDAT/CEDEN. These data can be obtained from http://bdat.ca.gov/Php/Data_Retrieval/data_retrieval_by_category.php under the "Water Quality" link.

Discussion

Many Groups within California, including multiple agencies, academic, private, and stakeholder entities collect large amounts of environmental data. These data currently exist in diverse formats and in different databases with inconsistent, and in some cases, difficult means to access. The Bay/Delta and Tributaries (BDAT) Cooperative Data Management system and California Environmental Data Exchange Network (CEDEN) provides a data organization and distribution services for environmental-related data through a collaborative efforts at local and CA State level. Data associated and relevant to this study were integrated with other monitoring data colletected in the State, made available through BDAT/CEDEN on the web. A local database based on the SWAMP standards was develop and provided to Will Stringfellow's group.

References

www.bdat.ca.gov

http://mpsl.mlml.calstate.edu/swamp.htm

www.sjrdotmdl.org

Table 1: SJR Up-Stream Study Time Series Data Available from the BDAT Websites

YSI temperature of water
YSI electrical conductivity
Stage
Flow
Bubbler temperature of air
Conductivity of water measured using other equipment

Table 2: SJR Up-Stream Study Discrete Water Quality Data Available from the BDAT Websites

Temperature (water)
Specific Conductivity
TDS
DO %
DO mg/L
Depth
pH
ORP
Sonde Turbidity

Sonde Fluorescence
Sonde chl-a corrected for Tri-C
Sonde chl-a corrected for SM
PAR (spherical incident light)
PAR (Flat Quantum Detector)
LUX
CaCO3 /L at 8.3 Alk mg
CaCO3 /L at 4.5 Alk mg
Total Organic Carbon
Dissolved Organic Carbon
VSS + DOC
VSS
TSS
Mineral Solids
Total NH4 - N
PO4 - P using 0.7 micron filter (Ortho-Phosphate)
PO4 - P using 0.2 micron filter
Total Fe
Total N
Soluble NH4 - N
NO3 - N
Total - P
BOD 10
CBOD
NBOD
Total Protein
chl-a by SM
Pheophyton
Algal pigments
chl-a by Tri-Chromic
chl-b
chl-c
chl-a using LE 665 method
Precipitation