Delta Regional Monitoring Program

Data Integration

Strawman Proposal

November 20, 2009

List of Acronyms

AEAL	Aquatic Ecosystems Analysis Laboratory
BDAT	Bay-Delta & Tributaries Database
Cal/EPA	California Environmental Protection Agency
CDEC	California Data Exchange Center
CEDEN	California Environmental Data Exchange Network
DWR	California Department of Water Resources
EPA	Environmental Protection Agency
IEP	Interagency Ecological Program
MLML	Moss Landing Marine Laboratories
MWQI	Municipal Water Quality Investigations
NPDES	National Pollution Discharge Elimination System
NWIS	National Water Information System
POD	pelagic organism decline
RMP	regional monitoring program
SCCWRP	Southern California Coastal Water Research Project
SFEI	San Francisco Estuary Institute
SWAMP	Surface Water Ambient Monitoring Program
TMDL	Total Maximum Daily Load
UC	University of California
U.S.	United States (of America)
USGS	U.S. Geological Survey
WDL	Water Data Library (WDL)

General Overview on Data Integration

Data integration and interoperability is a pivotal issue for the development of the Delta Regional Monitoring Program (RMP). Data integration refers to collecting and handling data using documented and standardized procedures, and the process of combining data residing at different sources and providing the user with a unified view of these data. Although a large volume of water quality and ecosystem health data are being collected in the Delta by the individual, program-driven, monitoring efforts, there are major cross-cutting challenges, such as the Pelagic Organism Decline (POD), that cannot be addressed by any individual program alone. A recent review of water quality data to support POD investigations revealed a lack of documentation and standardization of data collection and handling procedures disqualifying the majority of the data from any analyses (M. Johnson, pers. comm.). Data collection and handling protocols were either lacking or different to the extent that the interoperability of data could either not be established or data from different programs were not compatible for key analyses (M. Johnson, pers. comm.). At the Delta RMP kick-off meeting, stakeholders recognized the current lack of data integration across different programs and identified it as a key challenge to a better understanding of Delta water quality and aquatic life use issues. Data integration and interoperability across different programs is a prerequisite for conducting the types of analyses that allow for a more comprehensive view of some of the cross-cutting major challenges faced in the Delta.

In a way, data integration and interoperability is probably THE main issue to tackle for the Delta RMP for the following reasons:

1) One of the main goals identified for the Delta RMP is to provide information about how multiple stressors are impacting resources and such complex issues require data integration.

2) The Delta is a large area with complex processes and understanding stressors and impacts requires an integrated regional perspective.

The California Water Quality Monitoring Council (Monitoring Council) in concert with the Surface Water Ambient Monitoring Program (SWAMP) is developing solutions to strategically improve water quality data integration across the state. This broad-based, statewide approach focuses on improved coordination and standardized data management to improve data access, integration, analysis, and information dissemination. The approach is a statewide information exchange network that consists of regional data centers that maintain a system of databases--called the California Environmental Data Exchange Network (CEDEN). Access to the information will be through theme-based web portals following the vision of the Monitoring Council. As explained later, regional initiatives are expected to contribute to these broad-based solutions and in turn will have access to training, data management infrastructure, and other tools.

This strawman proposal presents an approach for how the Delta RMP effort can facilitate better data access and integration in the region. The strawman proposal is consistent with recommendations of the Monitoring Council, which includes representatives from the California Environmental Protection Agency (Cal/EPA), the Resources Agency, the Department of Public Health, the public, regulated publicly owned treatment works, regulated stormwater interests, agriculture, water supply interests, citizen monitoring groups, and the scientific community: The strawman proposal is based on the following principles:

1. Coordinated organizational approach
2. Coordinated access to monitoring and assessment information
3. Performance-based monitoring and assessment methods
4. Standardized data management protocols including
    a. Data upload and checking
    b. Data storage
    c. Data exchange
    d. Technology coordination and transfer
    e. Data interoperability

Overview on Data Management and Analysis

There are 17 ongoing surface water quality monitoring programs in the Delta collecting data for various purposes and uses (for more information, see the Aquatic Science Center report Summary of Current Water Quality Monitoring Programs in the Delta). Each program varies in its approaches to data management, storage, availability, and access points. Although, in general, most data are stored electronically by now, not all data are readily available online. The main public access points for Delta surface water information include the SWAMP databases/CEDEN, Bay-Delta & Tributaries Database (BDAT), California Data Exchange Center (CDEC), Water Data Library (WDL), and National Water Information System (NWIS).

The SWAMP databases/CEDEN, is a network of regional data centers that includes the San Francisco Estuary Institute (SFEI), Moss Landing Marine Laboratories (MLML), Southern California Coastal Water Research Project (SCCWRP), and UC Davis Aquatic Ecosystems Analysis Laboratory (AEAL). CEDEN is being developed with the goal of making environmental data publicly accessible through a system of databases that are maintained by a network of regional data centers. Although the data centers are in different stages of development, each of the four data centers identified above is providing relevant services for ongoing Delta monitoring. The UC Davis AEAL regional data center manages data from the Central Valley Water Board's Irrigated Lands Regulatory Program and is also designated for managing water quality data collected for Total Maximum Daily Loads (TMDLs) and Proposition-funded projects in the region. The SFEI regional data center provides access to data from the San Francisco Bay RMP and other regional studies through the RMP data access tool and has released beta versions of a map-based Web Query Tool for data access and the Central Valley Monitoring Directory[1]. Services provided by MLML include a Data Help Desk for issues related to database comparability and quality assurance management issues. Both SFEI and MLML provide online data loaders and data checker tools. SCCWRP is maintaining similar data management and access tools for the Southern California region.

The Bay-Delta Tributaries Project (BDAT) contains environmental data (i.e., water quality, biological, and meteorological) concerning the San Francisco Bay-Delta and provides public access to them (http://bdat.ca.gov/). BDAT is maintained by the California Department of Water Resources (DWR). Data currently available for download through the BDAT web interface include both discrete and continuous monitoring data collected by Interagency

---

[1] The Central Valley Monitoring Directory provides access to program and metadata information of current water quality monitoring efforts in the Central Valley watershed. It is being developed as a tool to help improve the coordination and integration of existing monitoring efforts.

[Ecological Program](#) (IEP), [Municipal Water Quality Investigations](#) (MWQI), [State Water Project](#), [U.S. Geological Survey](#) (USGS), and other program participants.

The California Data Exchange Center (CDEC) is DWR's main access point for real-time continuous hydrological data, including water quality data, gathered by various cooperators throughout the State (http://cdec.water.ca.gov/). CDEC includes continuous surface water monitoring data collected by DWR's [Central District](#), State Water Project water quality monitoring, and MWQI programs. It also includes data of the U.S. Geological Survey's Delta Flows Network and continuous [IEP Environmental Monitoring Program](#) data. CDEC data are "real-time" data and become available immediately as they are being recorded.

The Water Data Library (WDL) is maintained by DWR as a database and public access point for water quality and hydrologic data collected by the [Division of Planning and Assistance](#), MWQI, the [State Water Project Water Quality Monitoring Program](#), IEP, and other programs inside and outside the DWR (http://www.wdl.water.ca.gov/). Data are posted in WDL after being checked for accuracy and completeness. WDL is an access tool for statewide data, whereas BDAT serves data from the Bay-Delta region and upstream watersheds.

The National Water Information System (NWIS) is the USGS Water-Quality Web Service and provides access to water resources data collected nationwide by USGS, including Delta flow and water quality (http://waterdata.usgs.gov/nwis). On NWIS, the unit values are available for 31 days only and then taken offline. Historical long-term monitoring records are maintained as daily values and are updated daily. Data sharing capabilities between NWIS and U.S. EPA's [STORET](#) data warehouse have recently been developed, as a result of a collaboration to provide water-quality station and result information in the same format. California's SWAMP is also currently collaborating with EPA on an exchangeable data format

Discussions regarding data integration for the Delta RMP are more or less at "ground zero." This strawman provides a basis for these discussions by proposing a framework for data integration in the Delta, in the context of the Delta RMP development effort. Rather than contrasting options, the proposal outlines an organic, forward-looking approach to data integration that takes advantage of and builds on the broad-based, statewide approaches by the Monitoring Council and SWAMP.

Resolving data integration issues for the Delta RMP means to accomplish integration with and among the existing distributed water quality monitoring efforts in the region. Proposed functions of the Delta RMP to achieve the overall integration include:

(1) developing monitoring questions that are common among multiple programs
(2) coordinating with existing data management efforts such as SWAMP and CEDEN
(3) guiding analyses and modeling to target the highest-priority questions identified by Delta water quality and ecosystem managers
(4) Use to the extent possible exiting data standards, such as those used by SWAMP, to improve data interoperability, utility and discovery.

Organizational Approach

- Statewide coordination:
  Building on the Monitoring Council vision of a statewide water quality data access solution, organize data integration around themes phrased as easily understood questions, coordinated by issue-specific workgroups:
    1. The Delta RMP will contribute to the statewide data access and integration solution by
        1.1. Providing baseline monitoring data at multiple spatial and time scales that are useful to multiple entities and programs
        1.2. Adhering to standards set forth for statewide data collection and assessments
        1.3. Organizing sub-theme workgroups relevant to Delta-specific issues
    2. In return, the Delta RMP will benefit by
        2.1. Building on existing information management structures
        2.2. Having access to technical monitoring and assessment tools
        2.3. Using state guidelines for data access solutions
        2.4. Implementing statewide monitoring and data management standards

- Regional and local coordination:
  The Delta RMP can and should play a role in fostering and maintaining the integration of local and regional monitoring and assessment efforts by
    1. Obtaining data from participants[2]
    2. Synthesizing information from diverse sources
    3. Providing data in a timely manner to potential users
    4. Promulgating data standards and monitoring protocols such as those used by SWAMP
    5. Coordinating integrated but distributed data management based on existing system of databases
    6. Coordinating regional development of information technology infrastructures within the umbrella of a statewide data exchange network
        5.1. Data management through an existing regional data center, e.g., MLML, SCCWRP, SFEI, or UC Davis AEAL
        5.2. Data access through existing regional data access points, e.g., SFEI, UC Davis AEAL
        5.3. Long-term: data will also be accessible through theme-based web portals under development by the Monitoring Council
        5.4. Coordinating regional participation in web-portal development

- Attachment 1 provides a strawman proposal for a data management and access solution to achieve data integration under the Delta RMP umbrella.
- Attachment 2 depicts proposed phasing of the data integration functions.

---

[2] This may include data that are available through public databases and data that are not currently accessible in a public database (e.g. NPDES receiving water data).

Information Access

A web access/query tool will be needed to retrieve Delta RMP data. SFEI has developed a web access tool for the SF Bay RMP and is currently developing a map-based data query tool. These tools are part of the SFEI regional data center, which is part of CEDEN. The existing tools could be customized and provided to the Delta RMP. They would entail the following components to meet highest-priority needs of data users:

- Map-based interface
- Data and assessment information at a variety of space and time scales
- Ad-hoc query and analysis results
- Ability to download raw data
- Ability to create graphic summaries based on assessment questions (e.g., location of sampling stations with exceedances of fish consumption guidelines, cellular abnormalities in aquatic organisms, etc.)

- Delta RMP data would be managed and made available by one of the existing regional data centers. Options include SFEI, UC Davis AEAL, MLML, and SCCWRP. If SFEI and/or UCD AEAL were to become the designated data steward for the Delta RMP, the technical infrastructure for data management and access would already be in place, although the availability of specific tools differs between them. MLML also maintains a data portal, and SCCWRP has data mapping and query tools in place. Existing tools could be customized for the Delta RMP. In the long-term, data access would be available through the theme-based web portals under development by the Monitoring Council.

Monitoring and Assessment Methods

- The examination of status and trends at the regional level require standardized protocols for data analysis, interoperability and assessment. The data analysis framework should also be scalable to support examinations at the statewide, regional, watershed, and site-specific level.
- On the regional scale, the Delta RMP data analysis framework will be integrated with the proposed Strategic Plan for Monitoring and Assessment in the Bay-Delta and its Watershed ("Strategic Plan"), which is currently being developed by a team convened by the CALFED Science Program (lead by Dr. Samuel Luoma). The purpose of the Strategic Plan is to provide a means of better integrating data from widely distributed programs (water quality, supply, biological resources, habitat characteristics), identify gaps in existing programs and assure ongoing integrative, interpretive assessments of the data. The framework will define a core of monitoring to explicitly track environmental changes expected to stem from major water management decisions.

Data Management and Documentation - Data Upload, Checking, and Storage

- Data management of the Delta RMP can build on existing resources and tools, including database templates, trained staff at regional data centers and analytical labs, a data checker tool, and a web-based data submission tool.

- Databases will be required for loading and storing data that are generated and/or processed by the Delta RMP. The database will be SWAMP comparable. Templates such as SFEI's San Francisco Bay RMP database are available.
- Trained staff will be needed to upload and process data according to standardized data reporting formats. Depending on the volume of data to be processed, this may require from 0.5 to 3 Personnel Years.
- Existing regional data center data review procedures requiring multiple checks will be implemented to insure data quality meets intended uses, before data are made available via the web access/query tool. The SWAMP data checker can be used by participating monitoring groups and/or contracting labs for ensuring standardized formatting criteria are met. Regional data center staff (both SFEI and UC Davis AEAL) continue to train analytical staff to use standardized data reporting formats.
- A web-based data submission tool should be implemented for labs to use when submitting their data. This will minimize the time required to upload data to the database.

Data Management - Data Exchange

- The Delta RMP database will be integrated with one of the existing regional data centers; therefore, Delta RMP data can be readily exchanged with CEDEN.
- Data exchange with the Delta RMP will build on existing data management infrastructure and procedures (e.g., BDAT, SWAMP databases)[3]. (See appended data integration strawman figure).
- Compatibility will be ensured by using standardized data formats. The SWAMP format will be the designated data format for the Delta RMP. The SWAMP format is based on the national Water Quality Exchange Outbound XML schema.
- Through CEDEN, Delta RMP data will be available for statewide assessments and information access through theme-based web portals as proposed by the Monitoring Council.
- Since CEDEN data are automatically exchanged with the national Water Quality Exchange system, Delta RMP data eventually become available through the planned National Environmental Information Exchange Network.

Data Management – Technology Coordination and Transfer

- Delta RMP data management will be integrated with CEDEN and the Monitoring Council's Water Quality Data Access Solution
- To the extent possible, the Delta RMP data management should build on existing resources and information infrastructure (i.e. SFEI and UC Davis AEAL regional data centers)
- Delta RMP staff and participants need to be actively involved in efforts to improve technology infrastructure and coordination. The Delta RMP could serve as a vehicle to enhance technology transfer and coordination in the region and making state and national web services available to the region

---

[3] For example, if the IEP Environmental Monitoring Program would directly contribute data to the Delta RMP, these data could still be managed and processed like all other IEP Environmental Monitoring Program data, uploaded to BDAT, and then "harvested" by Delta RMP data management staff.

More Information

California Water Quality Monitoring Council/SWAMP – My Water Quality
http://www.swrcb.ca.gov/mywaterquality/

> Access point to the planned Theme-based Web Portal of the Monitoring Council, hosted by SWAMP. The site is still in development and only the Safe to Swim theme is viewable.

Maximizing the Efficiency and Effectiveness of Water Quality Data Collection and Dissemination, and Ensuring that Collected Data are Maintained and Available for Use by Decision-makers and the Public – Recommendations of the California Water Quality Monitoring Council
http://www.swrcb.ca.gov/water_issues/programs/monitoring_council/docs/sb_1070_full_report_final.pdf

> This report presents the Monitoring Council's vision for a statewide integrated data management and access solution. It was prepared in response to State Senate Bill 1070, which called for improving coordination of water quality monitoring and a single public access point to water quality information on the internet. The recommendations were submitted to the Secretaries for Environmental Protection and Resources.

Services of a Regional Data Center
http://ccamp.net/ceden/index.php/Services_of_a_Regional_Data_Center

> Ultimate vision for the core services that a fully operational Regional Data Center should provide. As described above, certain functions are already in place at some RDCs and can be transferred and/or adapted for use by the RDC that would be charged with managing and/or coordinating Delta RMP data (for example, integrated regional contaminant and flow data management for the Delta).

DRAFT

Attachment 1

Delta RMP Strawman Proposal   8   November 20, 2009
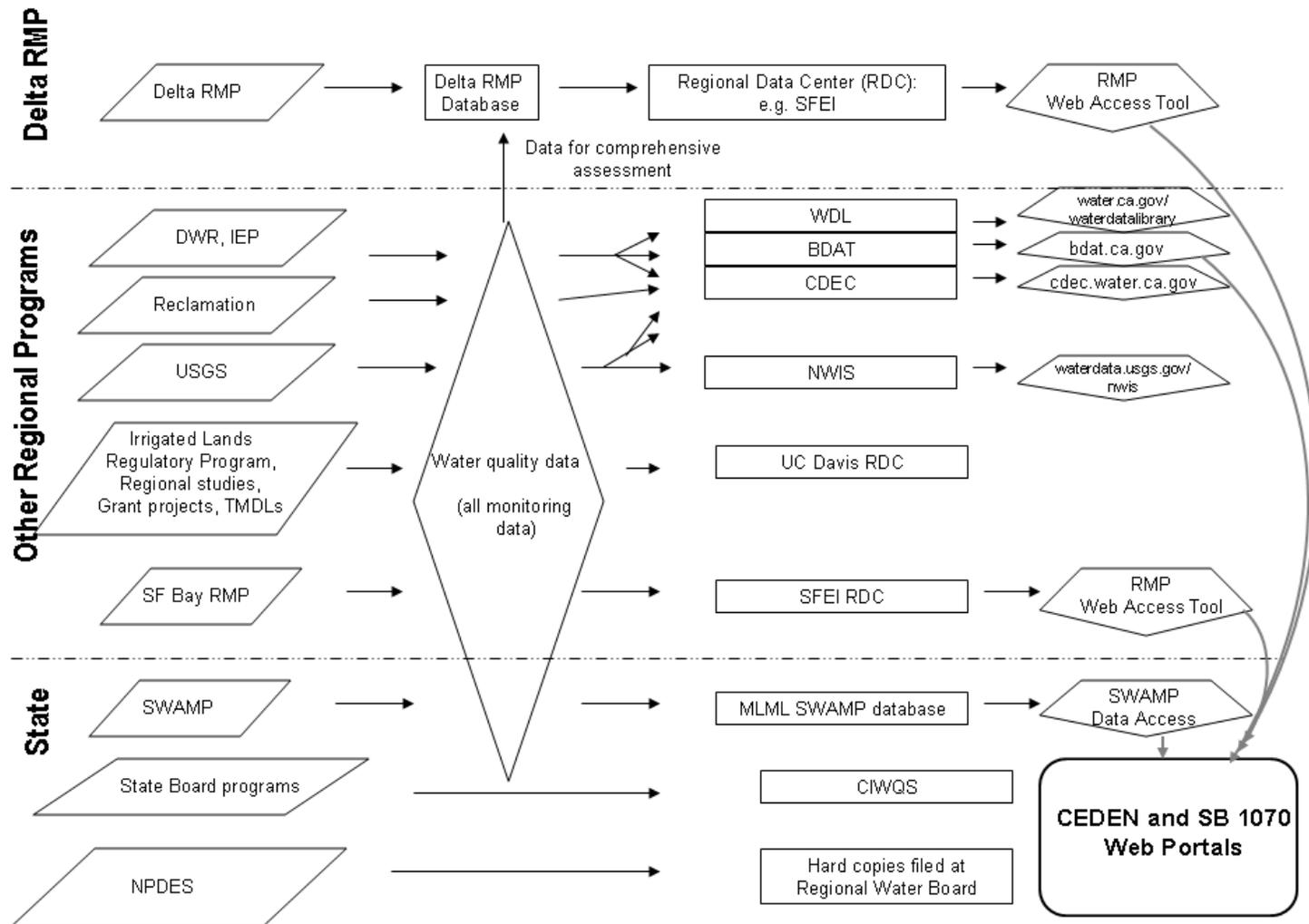
Attachment 2

Delta RMP Development Steps

Proposed Strategy

Alternatives Report

Program Implementation

Develop Goals, Objectives & Strategy

Develop Framework Options

Present Options to Water Boards

Program Planning

Phase I (Pilot Phase) "Proof of Concept"

Phase II Long-term Program

2010   2012
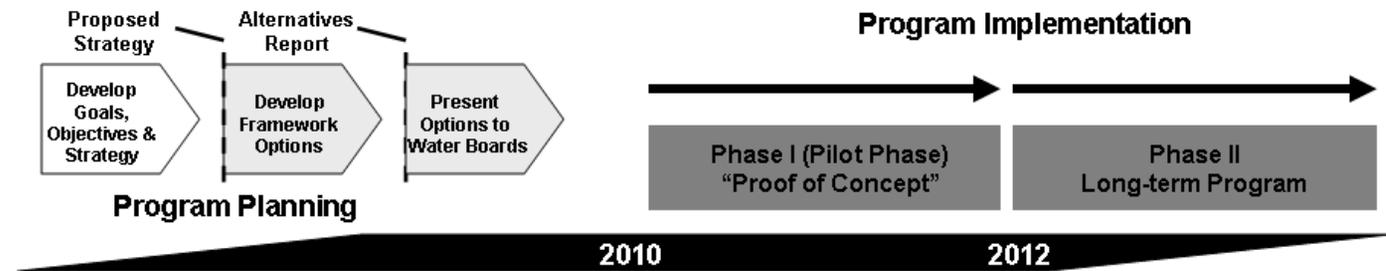
Proposed Phasing of Delta RMP Data Integration Functions

1 Develop Monitoring/Assessment Questions

1.1 Develop Monitoring & Assessment Questions

Initial Monitoring & Assessment Questions complete for   Phase I   Phase II

2 Coordinate With Existing Data Management

2.1 Implement Statewide Monitoring and Data Management Standards
2.2 Coordinate Integrated Data Management
2.3 Upload, process, and manage RMP Data (RDC)

3 Improve Data Interoperability, Utility, and Discovery

3.1 Solicit and Gather Data From Participants
3.2 Synthesize Information from Diverse Sources
3.3 Provide Data to Potential Users
3.4 Promulgate Data Standards & Monitoring Protocols
3.5 Coordinate Regional Participation in Web-Portal Development
3.6 Customize Existing Data Mapping and Query Tools
3.7 Implement Web-based Data Submission Tool

Data Access

Theme-based data access/Web Portals