

Subtask 9.1

**SAN DIEGO RIVER WATERSHED
DATA AND INFORMATION MANAGEMENT PLAN**

Prepared for

The San Diego River Watershed Workgroup

Prepared by

Anchor Environmental CA, L.P.

Everest International Consultants

KTU+A

Merkel & Associates, Inc.

TRAC

Michael Welch, Consulting Engineer

FEBRUARY 2004



TASK 9.1

DATA AND INFORMATION MANAGEMENT STRATEGY

A Data and Information Management Strategy (Strategy) is an important tool for managing the large amounts of data typically collected during large-scale planning projects. This memorandum discusses the following basic steps to develop a strategy for the long-term management of the San Diego River Watershed data and information:

- Institute a sorting process to first categorize and then assess the applicability of the different data sets collected for use within watershed assessment
- Implement a database application to store pertinent information about the watershed
- Provide recommendations for future data and information management based on discussions with stakeholders at a Data and Information Management Strategy meeting on November 13, 2003 and at the December 12, 2003 meeting of the San Diego River Watershed Workgroup.

The Sorting Process

A list of the existing data and information available within the San Diego River watershed was compiled within Task 7.1 (Data and Information Collected) and augmented during Tasks 7.2 and 8.5 (Watershed Characteristics Inventory Report and Water Quality Report) of the watershed management planning project. These data will be sorted to determine their suitability and precision for use in watershed assessment based on the following factors:

1. Spatial or Non-Spatial

Is the data specifically tied to a geographic point, line or area (spatial)?

Is it more descriptive or policy based such as plans, reports, and guidance documents (non-spatial)?

2. Date of collection

When was the data collected? How old is it and is it still a valid representation of watershed conditions?



3. Method of collection

How was the data collected? Were they remotely sensed, interpretation of aerial photography, visually, field methods, laboratory methods, etc.?

4. Scale of collection

At what scale, if appropriate, was the data collected -- 1:100000, 1:24000, 1:4800, GPS?

5. Purpose of Data

What was the purpose of the data when it was collected?

6. Appropriate Scale

At what scales is the data appropriate to be used for analyses -- watershed-wide, sub-basin, reach, site specific?

7. Data Limitations

Are there any limitations to the use of the data that are important to understand?

8. Potential areas of data application

For which areas of functional analysis is the data appropriate to use -- habitat, hydrology, hydraulics, water quality, land use/management?

Potential Database Application

The Anchor Environmental Project Team assisted the Watershed Workgroup in developing a spreadsheet (Subtask 7.1 Data and Information Collected) to store information about each of the data sources. A database created to access this information would allow future end-users to query the database and retrieve all records that contain information pertinent to specific watershed assessment analyses. For instance, the end-users could query the database to find all data that has been collected since 1995 which is appropriate for watershed-wide analysis in the functional area of hydraulics. The application would search the database for the following:

- Data that have dates newer than 1995
- Data that have been identified as being appropriate for watershed-wide analyses
- Data that pertain to hydraulics.



The result of the query would be a listing of the original data sources (not the data itself), which would allow the user to directly access the available data in an efficient and cost effective manner. The database, therefore, would be a tool for the watershed management planners and other users that would provide consistent and pertinent data for review and inclusion in assessment activities. The Anchor Environmental Project Team will utilize the spreadsheet developed during subtask 7.1 to perform this function during the upcoming Watershed Assessment task. Also, the data and information management plan will be a component of the watershed management plan, and is important in plan implementation.

Stakeholder Input to Data Management

A variety of issues and options pertaining to data and information management were discussed with key stakeholders during the November 13, 2003 Data and Information Strategy meeting held in the San Diego Regional Water Quality Control Board's Hearing Room, and subsequently at the December 12, 2003 meeting of the San Diego River Watershed Workgroup held at the City of Santee, Conference Room, Building 8. Appendix A shows the groups that were represented at these meetings.

Discussions about data focused on the various monitoring programs currently conducted throughout the watershed and their relationship to potential future modeling activities associated with watershed management. The kind of information collected and the methods utilized by these monitoring programs were reviewed and presented in the form of a large matrix showing each monitoring program/location on one axis and the constituent/information type being collected on the other. The Workgroup also identified that it would be beneficial to include information about the monitoring data being collected (metadata¹) for grant programs or as mitigation requirements, and identify whether the published data represented direct information or interpreted data.

Opportunities for and benefits of better program coordination and consistent data collection were discussed with the groups in the context of data requirements for watershed

¹ "Metadata" is data about data. Metadata is utilized to ensure that results are not misinterpreted or misused. The information described above in the sorting process forms the basis for metadata -- why the data was collected (e.g., regulatory requirement or volunteered), the limitations of the data (e.g., detection limits), the source (NGO or governmental), and other information that allows the researcher to better understand how the results can and should be used.



assessment tools (models) that could be used in future watershed analyses. A matrix of watershed assessment models to data needs was developed and is included as Appendix B. This matrix is intended to be used by stakeholders as a reference for the basic data needs of various assessment tools when they are developing/revising future monitoring and data collection activities throughout the watershed. During the discussion with the Workgroup, the importance of determining data adequacy for the various modeling applications was discussed, as well as the value of two sets of time series data, one for model calibration and one for model validation.

The value of metadata (data about data) was also discussed with the group and a series of sample metadata ranging from poor to excellent were distributed for review. The lack of complete metadata makes it very difficult for someone other than the entity collecting the data to determine if a data set is appropriate for a particular use, or not, and why. The Workgroup proposed that the following information be included: what category (parameters if appropriate) of data is being collected, how is it being collected (equipment, protocols, detection limits), where is it being collected, who is collecting it, and why are they collecting it (voluntary, regulatory monitoring, single event, ongoing time series).

Options for Data Dissemination and Storage

There are essentially two end point options for data dissemination and storage: centralized and de-centralized. Within the centralized option, one entity is identified as having the responsibility of warehousing data and making it available to other stakeholders, a one-stop-shop for information pertaining to the watershed. Within the de-centralized option, each stakeholder is responsible for maintaining and making available to other stakeholders links to their data. This option, while not providing the convenient one-stop-shop of the centralized option, is something that can be more easily maintained and expanded upon as needed with proper coordination.

Recommendations

The findings and recommendations that evolved from the November 2003 stakeholder meeting include the following:



1. The general consensus among the stakeholder group was that the de-centralized option is preferable for the San Diego River data.
2. The group agreed that a website, hosted by a workgroup member and containing information about all of the data being collected, should be evolved and contact data for each entity be provided.
3. The major stakeholders within the watershed should continue to discuss this topic and develop an agreed upon metadata content that can be used to document their respective data sets in the future.

During the December Workgroup meeting, the San Diego River Park Foundation (SDRPF) agreed to serve as the host site for the data and information database. At that meeting, the attendees concurred with the recommendations from the November meeting and added the following:

4. Form a Data Management Workgroup or Subcommittee comprised of representatives from the various entities conducting monitoring within the Watershed. The group should initially focus on the following issues:
 - Make decisions about how data should be sorted, but not eliminated
 - Determine appropriate format for data links (e.g. FTP, HTML, phone)
 - Identify data gaps affecting the various aspects of watershed assessment
 - Determine appropriate coordinate system to use for monitoring point data (e.g., latitude/longitude would be more user friendly to the public as it could be incorporated into handheld GPS devices, Stateplane would be more GIS friendly to the research and consulting industry)
 - Develop standards and discuss standardization of data
 - Define the appropriate types of data to be collected and shared
 - Develop methods to reduce redundancy in existing monitoring data
 - Develop methods to increase the applicability of monitoring data and fill data gaps



- Develop methods for sorting data into various categories
- Determine mechanisms for providing data updates from individual entities to the larger group or public (e.g. scheduled updates, links to updates, solicited vs. unsolicited requests, direct or indirect access)

5. SDRPF, as the Manager of the database, will be responsible for the following at a minimum:

- Development and maintenance of website interface to data
- Creation and maintenance of website links to data (e.g. appropriate contact information, website access, FTP access)
- Development of a query mechanism to facilitate data browsing

Funding for this project has been provided in full or in part through a contract with the State Water Resources Control Board (SWRCB) pursuant to the Costa-Machado Water Act of 2000 (Proposition 13) and any amendments thereto for the implementation of California's Nonpoint Source Pollution Control Program. The contents of this document do not necessarily reflect the views and policies of the SWRCB, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.



APPENDIX A

PARTICIPATING ORGANIZATIONS

The following is a list of organizations that attended the public meetings and that assisted in the development of this Data and Information Management Plan.

California Department of Water Resources
California History and Culture Conservancy
City of El Cajon
City of San Diego Storm Water Program
City of San Diego Water Department
City of Santee
County of San Diego, Flood Control
County of San Diego, Parks and Recreation
County of San Diego, Watershed Planning Program
Friends of Famosa Slough
Pacific REMS
Padre Dam Municipal Water District
Riverview Water District
San Diego Regional Water Quality Control Board
San Diego River Park Foundation
San Diego State University, Department of Geography
Southern California Wetlands Recovery Project
U.S. Bureau of Reclamation
U.S. Geological Survey



APPENDIX B

WATERSHED AND STREAM MODELS AND DATA NEEDS

MODEL	Purpose					DATA NEEDS																						
	Watershed Hydrology	Watershed Pollutant Loading	Stream Flow/Stage	Stream Water Quality	Stream Sedimentation	Rainfall	Watershed Topography	Soil Conditions	Land Cover/Vegetation	Land Use	Groundwater Conditions	Evaporation	Climatology (Temperature, Wind, Radiation, Pressure, Cloudiness)	Atmospheric Deposition	Point Source Loadings	Livestock Farms	Septic Systems	Agriculture, Recreation Practices	Headwater Inflow/Pollutant Concentration	Stream Slope, Cross-sections	Stream Bed/Bank Conditions	Stream Crossings (Bridges, Culverts)	Floodplain Conditions	Sediment Characteristics	Stream Flows (Calibration)	Stream Pollutant Concentration or land use loading (Calibration)	Stream Sediment Transport Rate or Geometry Change (Calibration)	
HSPF	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	
HEC-1	•					•	•	•	•	•										•	•	•						
HEC-2			•																	•	•	•	•					
HEC-RAS			•																	•	•	•	•					
HEC-6				•																•	•	•	•					•
QUAL2E				•									•		•					•	•	•				•	•	

NOTES: Priority items are Stream Channel Geometry (particularly at crossings) and simultaneous Continuous Stream Flow, Water Quality parameters to be modeled and Sediment Characteristics.

Wshed Hydro- WARMF, HSPF, HEC-1, TR-20

Pollutant load- WARMF, HSPF

Flow/Stage- HEC-RAS, HEC-2, TR-55

Stream Sed- HEC-6

Stream WQ- QUAL2E

