California Water Quality Monitoring Council

Data Management Workgroup (DMWG)

Audit Report

(1/17/2014)

Purpose of the DMWG:

The DMWG provides expertise to establish the overall approach to make use of and integrate existing data management systems into a distributed system of databases, catalogs, and assessment and mapping tools to enable users to access data, metadata, and assessment products from a single entry point, or web portal. In support of the Council's Comprehensive Strategy, key responsibilities of the DMWG include.

- Assist Monitoring Council workgroups identifying methodologies for assessing data management and quality needs.
- Assess and recommend best practices for development of structured data formats and data management strategies complying with appropriate national and state guidelines.
- Identify methods to increase accessibility of water quality and related ecosystem data and opportunities to coordinate and share these data among workgroups, governmental agencies, and non-governmental organizations.
- Assess and recommend IT tools and standards facilitating development of portals meeting Monitoring Council web development guidelines.
- Serve as a resource to assist other workgroups to evaluate technologies in the areas of data management, web applications and geospatial information management.
- Serve as a resource to workgroups for communicating, and where necessary, translating into clear, non-technical language recommendations regarding data management in support of individual workgroup's efforts.

Overall Assessment of Success of the DMWG:

Since its establishment, the DMWG has focused on developing a charter, assessing the state of data, technologies and needs of the existing theme specific workgroups and developing summary documents and recommendations for best practices regarding recommend infrastructure and data standards for Monitoring Council and Portals.

While several work products resulted, the DMWG has found it exceptionally challenging to build and maintain momentum to move these items forward in a timely and effective manner. In large part this is due to a lack of resources and direction from management. As is the case with all of the workgroups, members of the DMWG serve in a volunteer capacity, and as such, it has been extremely difficult to maintain the necessary attention and effort necessary to accomplish key tasks. Furthermore, because the DMWG was not established before several of the theme specific workgroups, significant effort has been required to develop an understanding of the many and varied data and technological solutions that had already been implemented by those groups in developing their portals prior to the DMWG's inception.

The workgroup has been successful in summarizing many of the existing portals and data utilized within them, though the process has not been particularly efficient or effective. In particular, we have found it very difficult to interact with the other workgroups and to obtain the information requested. Additionally, while the DMWG is able to provide advice for implementing particular methods, standards and software, there are typically not the

resources, and in some cases the expertise, available to implement them. Data availability for some workgroups is improving as data become more accessible through systems such as the California Environmental Data Exchange Network (CEDEN). But given data come from a variety of agencies and sources, the degree of availability for integration with the web applications driving the portals is inconsistent. In most cases, the ability to improve upon this situation relies on the availability of resources and management priority to make these data available via web services which can be easily ingested by the portals.

During the summer of 2011 a list of potential workgroup members was developed, representing data management experts from agencies, industry, academic and non-profit sectors. Invitations were initially sent to 29 individuals representing 15 organizations. The initial meeting of the DMWG was held in August 2011. In the last two plus years the membership focused on several key objectives including: Developing a common understanding of current and developing data management systems; establishing workgroup structure and schedule; and establishing subcommittees for: (1) Portals/Tools and (2) Data Standards.

The first three meetings of the workgroup included a series of presentations to provide the membership with an overview of various data management systems and approaches. Meetings throughout focused on developing the workgroup charter and collection of baseline information about the data and technology behind each of the existing and/or developing Monitoring Council Portals. Additionally a joint meeting between the DMWG and the three Ecosystem Health workgroups (Wetlands, Estuaries, and Streams Rivers and Lakes) was held in November 2012 to explore the value of developing a common GIS layer for aquatic resources to be shared by each of the workgroups and to establish effective channels of communication between workgroups. Meetings in 2013 continued to flesh out information about the needs of the existing workgroups as well as to develop recommendations/best practices relating to commonly needed data sources across workgroups and mapping technologies for portals.

Meetings:

Since its inception in the fall of 2011, the DMWG has met 12 times (approximately every other month) with some exceptions when meetings were cancelled due to a lack of progress or agenda. One additional, joint meeting in November 2012 with the three Ecosystem Health workgroups (Wetlands, Estuaries, and Streams Rivers and Lakes) was organized to assess value of a common GIS layer for aquatic resources.

- August 2011
- September 2011
- November 2011
- January 2012
- April 2012
- June 2012
- August 2012
- November 2012: joint meeting with Ecosystem Health workgroups.
- December 2012
- February 2013
- April 2013
- June 2013
- October 2013

Key accomplishments:

- Developed and adopted a workgroup charter;
- Established two subcommittees: (1) Portals/Tools and (2) Data Standards; (subcommittees met as needed via phone/web).

- Through the subcommittees, inventoried and assessed data and technologies in use by existing and forthcoming theme specific workgroups and developed recommendations regarding mapping technologies for portals.
- Held a joint meeting between the DMWG and the three Ecosystem Health workgroups to assess opportunities to develop a common/shared water data layer, such as the California Aquatic Resources Inventory, (CARI) for use by the Monitoring Council Workgroups.
- Developed issue paper for which web mapping framework to utilize as a replacement for the Google Maps API v.2 framework used on a number of My Water Quality portal pages.
- Developed an outline for a data management best practices guide for theme specific workgroup portals which was vetted with several of the workgroups.

Recommendations:

- In our 2012 report, we suggested that all existing and future theme specific workgroups should maintain a designated data liaison who also participates in meetings of the DMWG to ensure a consistent two-way exchange of information between workgroups. While this was attempted in 2013, it was only partially effective. In all cases the designated individuals were members of the DMWG who for various reasons were already involved with these other workgroups – typically in the capacity of serving a technical rather than a domain-specific role. The DMWG continues to struggle with bridging the gap between the scientists who possess a strong understanding of the questions, data and analysis required in their portals and the technical expertise to translate these effectively to the database, mapping and analytical tools and frameworks that the DMWG can evaluate in light of their requirements. Additionally, given the small number of individuals actively involved in the DMWG, there is simply not time available for these individuals to participate in multiple workgroups at the level needed to effectively evaluate their needs with the data management options available. Furthermore, because the existing workgroups have already developed (fully or partially) their own data frameworks and delivery mechanisms, recommending changes (let alone imposing standards) has proven ineffectual, primarily because the resources to implement such changes are not available.

  As an example the DMWG identified a need to address the deprecation of the Google Map 2.x API which the majority of the existing portals were developed on. The deprecation occurred on November 19, 2013. In anticipation of this change the DMWG conducted an extensive evaluation of mapping platforms and put forward recommendations for a transition to an Open Source stack which would avoid both the forced changes experienced with the Google mapping option as well as the need for any portal developer to have access to expensive commercial software. Despite the recommendation of the DMWG, portals remain a hodge-podge of platforms including use of a commercial (Esri) platform which available to the State Board, who currently maintains many of the portals as well as some legacy Google maps and perhaps others. This limits the opportunity to maintain a consistent software stack (and thus a consistent look and feel) across all portals regardless of who takes the lead in developing them.

  While the DMWG has the ability to assess and recommend tools, approaches and software platforms to achieve consistency and interoperability across the workgroup portals, there is not a means to implement these recommendations effectively. It is clear that maintaining multiple platforms leads to a number of problems relating to the long-term maintenance and upkeep of the portals and should be addressed at the level of the Council going forward.

- A second critical focus for the DMWG should be to ensure water data can be effectively shared (machine to machine) between agencies and other data providers. The DMWG recognizes that data will continue to be collected, managed and maintained by individual agencies and/or organizations as appropriate to their respective mandates and that these agencies will remain the authoritative source for these data.

While choices relating to the internal storage and organization of data will necessarily vary by agency or organization, adoption and documentation of Interoperability standards to support a federated approach to data management should be a primary objective for the DMWG in the next year. As with data formats, common and shared data transfer protocols (e.g. web services, REST endpoints, JSON, XML, etc.) should be defined and documented to ensure that data are accessible to the monitoring council portals via services rather than requiring workgroups to contact data managers within multiple agencies and organizations to manually access, acquire, structure and/or provide data for use by the theme specific workgroups. Essential to both of these goals is the need for management within the relevant agencies to understand the need for developing the necessary updates to data and transfer protocols within their organizations to ensure that necessary data are provided in an appropriate and documented manner. At present, many agencies have perfectly functional systems for their own internal analysis and reporting needs, so developing such services for external access are viewed as unnecessary effort which receives little or no priority and lacks the necessary human and/or financial resources necessary for the implementation and maintenance of such services.

The DMWG noted barriers to sharing of data, particularly outside of State agencies remains a source of trepidation in some cases. Specific concerns include potential for: misunderstanding of data quality and appropriate use; legal liability, extra workload associated with preparing data for use by non-experts or in portals; lack of required expertise (e.g. preparation of data for web access, establishment of web services, etc.).

These issues remain significant barriers to the process as more of these data are made discoverable, and potentially comparable through efforts such as CEDEN, CERES the State Geoportal and other systems. Questions around data typically boil down to those of: (1) who is the authoritative source of a given data set; (2) how will data quality be confirmed and maintained; (3) how is versioning of data handled, meaning as data is changed, updated or edited, who does that and how (if at all) is the previous version maintained.

These concerns may be addressed in part through the development model language regarding data availability (e.g. web services), use constraints, metadata and data documentation standards at the level of the Monitoring Council. The DMWG recommends the Council shepherd a process to develop model language, in consultation with the theme specific workgroups, accounting for specific issues or limitations of data sharing and use relevant to their needs. Furthermore, it will be critical for direction and resources to come from the Council to make this a priority. Without such direction, it is likely the status quo will continue and effective data sharing will be seriously hampered for the foreseeable future.

- In order to develop and promote shared data management practices the DMWG has been working on a data management best practices guide. This publication can explain useful methods to improve the efficiency, accuracy and sustainability of individual portals while also supporting greater data interoperability between the portals and partner organizations. The initial version of the guide could explain commonly accepted best practices for data management supporting a general improvement in portal operation. Continuing draft revisions could then support dialog between the WQMC partners on standards necessary for data interoperability. Development of these standards are essential but would require resources which have not been available.
- A fourth recommendation of the DMWG was to develop and share a common GIS layer for aquatic resources to support portals requiring a similar GIS layer. A joint meeting was held to move that objective forward. While there was general support for a GIS layer that would meet the needs of multiple workgroups, there was some hesitancy to commit to a common standard until additional analysis could be done and options like CARI are available for review. This serves as one (of a number) of examples of the difficulties revolving around the theme of data sharing and transparency. Consensus

was that there is value in updating and/or modifying this common layer to serve the needs of multiple workgroups as well as to provide this (and other commonly requested data) as web services which are maintained by the appropriate agency or organization (e.g. the authoritative source) but available for integration into portals.

- As mentioned previously, the DMWG has had difficulty maintaining commitments of time and efforts from the relevant agencies and organizations. Because involvement in the DMWG is not directly supported and/or resources are not made available to facilitate its work, it perpetually suffers from uneven and/or a lack of participation by its members. Where appropriate intersections exist with related efforts, members of the workgroup may find the time and resources to attend meetings or call into them. However, since the real work of the group occurs between meetings as the products of the subcommittees, it is only through the good will and efforts of a limited few that any work products are produced.

  Additionally, because there is no explicit mechanism for supporting interaction between the DMWG and each of the theme specific groups, the ability to gather, aggregate and assess their individual and overlapping data and technology needs is limited. While it would be optimal if the DMWG had the personnel and resources to participate with each of the theme-specific workgroups, this has proven impractical. One means to expedite such a process to coordinate the data needs and approaches across all of the workgroups would be to identify specific one-time resources (staff and/or contractor support) to collect, analyze and develop a report summarizing the data requirements across the workgroups. While the DMWG has made multiple attempts to accomplish this task, it has proven too complex to achieve without direct and sustained effort by individuals knowledgeable in both data management and the environmental and water quality data types required.

  Perhaps, a half-dozen individuals have carried the majority of the weight of our efforts over the past 30 months. Provision of concrete support from the Council and resources to develop and implement recommendations of the workgroup could serve to energize those who have remained active, and potentially reinvigorate those who have dropped off or been wary to become involved. These need not be direct funds (though some support for travel to meetings would be beneficial). Support for agency staff and/or contractors to develop the needed web services to make key data sources available would provide a basis for more effective and rapid development and maintenance of all of the current (and future) workgroup portals. Currently recommendations of the DMWG are of little use to the existing workgroups unless there is a means to implement them. Future workgroups and portals would also benefit from having a documents and available base from which to build as opposed to developing their own approaches based on whatever knowledge and experience their particular membership brings to the table.

Conclusions:

Over these past 30 months the DMWG has made reasonable progress in addressing the Monitoring Council Strategy. However without the authority and support to translate these efforts into documented recommendations for data formats, transfer protocols available to the theme specific workgroups and their portals, it has been a frustrating experience. Enhancing support for implementation (from both management and financial perspectives) would provide the DMWG the opportunity to work more effectively with the existing and future workgroups as then update and develop their respective portals. Without such support, technical recommendations of the DMWG will serve little value as each workgroup continues to take whatever path is most expedient in meeting their own needs and without an eye towards the long-term upkeep and maintenance of the portals and the data sources and services upon which they rely.

Attachment 1 - List of Organizations Participating in the DMWG

State Agencies
• California Department of Fish and Game • California Department of Public Health • California Department of Water Resources • California Natural Resources Agency • California Ocean Science Trust (OST), MPA Monitoring Enterprise • California State Water Resources Control Board (SWRCB) • California Technology Agency • Central Valley Regional Water Quality Control Board
Research and Academic Organizations
• California State University (CSU), Council on Ocean Affairs, Science and Technology (COAST) • California State University, Northridge (CSUN), Center for Geographic Studies • Humboldt State University • Klamath Basin Monitoring Program • Lawrence Berkeley National Laboratory • San Francisco Estuary Institute (SFEI) / Aquatic Science Center (ASC) • San Francisco State University • Southern California Coastal Ocean Observing System (SCCOOS) • Southern California Coastal Water Research Project (SCCWRP)
Non-Governmental Organizations (NGOs)
• Council for Watershed Health • Ecolayers • Heal the Bay
Private Industry and Consultants
• 34 North • Esri • IBM • RimuDB