SB 1070 Comprehensive Strategy

[There are still some loose ends, such as relationships to SWAMP and to volunteer and regional monitoring efforts, and especially funding. A formal request for an FSR needs to be added, for example.]

Foreword

- Legislative background – SB 1070
- Requirement in the Statute for this report on a comprehensive strategy
- Purpose of report – to lay out a ten-year plan to achieve ambitious goals related to design and implementation of water quality monitoring programs, use of monitoring data in assessments and decision making, and development of tools and supporting infrastructure to enable wide access to data and information products
- Comprehensive Strategy addresses each aspect of the Statute, as illustrated in Appendix 1
- Audience for this report is Legislature, Secretaries of Resources Agency and Cal/EPA, other agency staff, and other interested parties in public and other entities involved in water quality monitoring and assessment

Chapter 1: Introduction

The Monitoring Council has spent the year since the release of its December 2008 recommendations (CWQMC 2008) implementing the first steps called for in that report, empirically testing the assumptions underlying those recommendations, and preparing the technical and institutional infrastructure needed for their full implementation. Four [confirm number as of December] prototype web portals have been developed and been made available for public access on the Monitoring Council's portal website (http://www.waterboards.ca.gov/mywaterquality/), focusing in order on:

- Swimming safety at beaches (Safe to Swim)
- Human health risk associated with sportfish consumption (Safe to Eat Fish and Shellfish)
- Drinking water safety (Safe to Drink)
- Wetlands status (Wetlands)

The Monitoring Council found a high level of enthusiasm for the web portal concept among parties both inside and outside state agencies and had little difficulty establishing productive partnerships with data sources, users of assessment products, and scientists directly involved in the analysis and interpretation of monitoring data.

Developing these web portals showed that the Legislature was correct in its assessment of the status of water quality monitoring programs and data. There is a clear need for a group such as the Monitoring Council to fulfill a coordinating role and to ensure access to standardized data and statewide assessment products. This necessarily involves more than the assembly of data and databases, although this is essential; it also requires developing assessment questions, methods, and products at the statewide level that respond to a variety of users' questions and perspectives. The process of developing these proof-of-concept web portals has also validated key assumptions underlying the Monitoring Council's core philosophy and confirmed the gains in efficiency of analysis, performance assessment, and reporting possible from the portal approach.

Developing the prototype portals also enabled the Monitoring Council to establish a functioning workgroup structure and define the core elements of the infrastructure (both institutional and technical) needed to support complete implementation of the December 2008 recommendations (CWQMC 2008) over the longer term. These accomplishments provide the empirical basis for the Monitoring Council's plan, presented in the following chapters, for moving forward with the ten-year Comprehensive Strategy called for in the Statute.

1.1 The Monitoring Council's approach clarifies the problem

SB 1070 described a number of problems that hamper the ability of managers, scientists, and the public to find, access, and use monitoring data and results. While these problems are widely acknowledged, attempts to solve them have had only limited success because of the diversity of monitoring programs, the sheer volume and variety of data they produce, and the number of databases and data systems in which data are stored. In particular, the absence of clear questions has made it more difficult to develop a useful analysis of data integration and access problems.

In contrast, the web portal that addresses the core question: Is it safe to swim in our waters? (and secondary questions such as: How clean was my beach, lake, or stream during the past month?) provides the context needed to effectively evaluate and then resolve coordination and access problems. The construction of the web portal motivated the Monitoring Council and its "Safe to Swim" workgroup to expand and then organize their knowledge about monitoring programs that focus on this question. As a result, the workgroup has a much clearer picture (Figure 1) of (1) the major sources of data available to answer this question statewide, and (2) which data are currently not being input to databases that can readily be accessed by the web portal. Similarly, attempting to apply assessment methods statewide compelled both the Wetlands and Safe to Swim workgroups to explicitly confront inconsistencies in monitoring designs and data aggregation methods that diminished the statewide applicability of assessment results.

Scientists and managers involved with these monitoring programs had long been aware of these data gaps and inconsistencies and, to be fair, these issues have not prevented individual programs from meeting their objectives. However, without the goal of producing statewide assessments and a mechanism for integrating and displaying information at this scale, there was little motivation (or need) to improve data access, standardization, or coordination.

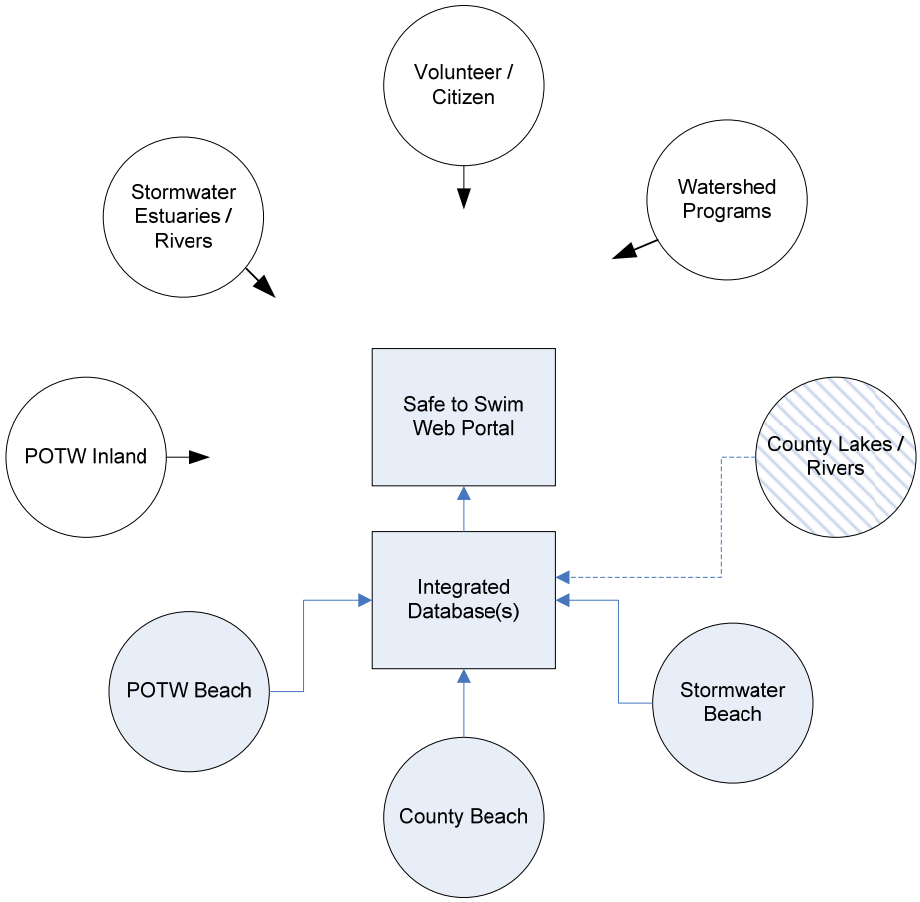

Figure 1. Schematic illustrating the range of categories of monitoring programs that produce data relevant to the Safe to Swim web portal. Data standardization and integration methods have been developed and applied only to a subset of the potential data sources. The workplan for this portal therefore includes efforts to improve the consistency of methods within this subset and to incorporate data flows from the remaining program types into the web portal.

1.2 Web portals foster solutions and improve efficiency

The process of constructing the web portals requires scientists and managers to collaborate on articulating meaningful assessment questions that are both useful to managers and the public and based on credible science. This collaboration, combined with the Monitoring Council's design principles for the web portals, fosters creative problem solving that makes use of a wider range of insights, tools, and resources than are available strictly within state agencies. For example, the Safe to Swim workgroup has proposed a streamlined data management and reporting pathway that makes greater use of technical resources at one of the regional data centers, while both the Safe to Swim and Safe to Eat Fish and Shellfish web portals incorporate mapping features developed by outside partners.

As the web portals continue to develop, they will enable state agencies to dramatically improve the accuracy and efficiency of many of their routine and ad hoc reporting functions. Quicker access to data and assessment products, combined with query and reporting tools built into the web portals, will make it much easier to respond to questions from the Legislature, agency managers, and the public. Such gains in efficiency have been identified in the Statewide Data Strategy Report, released in July 2009 by the Office

of the Chief Information Officer, as one of the major benefits of improved data integration. Even the prototype web portals developed this year by the Monitoring Council have already begun to demonstrate how such dividends can be achieved. For example, the State Water Resources Control Board is planning to use automated outputs from the web portals in annual performance reporting requested by its Office of Research Planning and Performance. And the Safe to Eat Fish and Shellfish web portal makes it possible to quickly create customized assessment products, at scales from individual lakes to the entire state, using monitoring and assessment results that were previously available only in databases (as raw data) and in agency reports (static assessment products). The web portals provide the more powerful ability for users to choose among, or define, multiple perspectives that suit their particular information needs.

1.3 Implementing the Monitoring Council's Comprehensive Strategy

The Monitoring Council's first year of effort has accomplished its primary purpose – to provide the empirical basis for developing a clear plan for moving forward with the Comprehensive Strategy called for in the Statute. The following sections of this report describe the Monitoring Council's core philosophy and approach (Chapter 2), which is fundamental to the success of the ten-year implementation plan (Chapter 3). Implementation will require:

- Further development of the four initial prototype web portals
- Expanding outreach to new partners, both within state agencies and outside of state government
- Identifying the next set of priorities for portal development
- Adapting lessons learned from the 2009 effort to the Monitoring Council's plans and procedures
- Designing and implementing the more permanent technical and institutional infrastructure needed to support this expanded and ongoing effort

Chapter 2: Philosophy and Approach

The Monitoring Council's fundamental vision is of broader and more streamlined access to monitoring data and statewide assessment products through a set of theme-based web portals. A fundamental element of this vision is the philosophy that the theme-based web portals themselves are central to the success of efforts to improve access and create statewide assessment frameworks. As validated by the prototypes developed during 2009, creation of the web portals promotes and organizes critical improvements that are impossible to achieve in a strictly bottom-up effort focused only on technical standardization. This philosophy provides an essential foundation for each element in the Monitoring Council's five-part approach to achieving the goals set by the Statute.

2.1 A philosophy of transparent, continual improvement

The Monitoring Council has established an operating philosophy that defines the complementary roles of the Monitoring Council and the theme-based workgroups, working within an overall context of transparent and continual improvement. As described more fully in Section 2.1.1 (A Flexible Organizational Structure), the Monitoring Council plays a role made up equally of leadership, coordination, and support, while the theme-based workgroups are responsible for the majority of the technical work involved in developing assessment methods and the portals themselves.

For the web portals to work as intended, they must meet all six performance measures described below in Section 2.1.2 (Performance Measures). In order to meet the performance measures, the Monitoring Council has identified the following principles as key elements of its operating philosophy:

- Constantly changing data, technology, and management information requirements mean that the web portals will never be completely "finished" or "perfect"

- The best way to ensure web portals are as responsive as possible to current requirements and constraints is to as open as possible about the strengths and shortcomings of the web portals and the data and assessment methods they are based on (see Sections 1.1 and 1.2 above)
- The Monitoring Council itself should play a central role in critiquing the web portals and facilitating plans for their continual improvement
- Such transparency builds credibility and encourages the involvement of the partners needed to continue developing and improving the web portals
- The web portals should provide the framework to both motivate and guide the effort needed to correct problems and develop enhanced capabilities

Organizations whose success is critically dependent on innovation, high quality, and/or high reliability explicitly cultivate just such a culture of open and transparent self-criticism and continual improvement.

2.2 A five-part approach to assessment and data integration

The Monitoring Council (CWQMC 2008) described a five-part solution essential to achieving its vision of broader data access through theme-based web portals. While these five elements remain central to the Monitoring Council's approach, the practical experience gained during 2009 has added detail and texture to the original concept of how these elements would function together. The five elements are listed here, followed by more detailed descriptions of how the Monitoring Council conceives them to operate after a full year's experience:

- An organizational structure built on decentralized, issue-specific workgroups that operate within common policies and guidelines defined by the Monitoring Council
- A set of performance measures which each theme-based workgroup will use to evaluate, coordinate and enhance monitoring, assessment, and reporting efforts
- A single, global point of entry to water quality data, and a design template for the complete set of theme-based web portals
- Standardization of monitoring and assessment methods that achieves an appropriate balance between statewide consistency and regional flexibility
- Database and data management standards necessary for more efficient data access and integration

It is important to emphasize a crucial difference between the Monitoring Council's approach and past efforts to provide improved data access and standardization. The Monitoring Council will not simply link to monitoring databases and encourage the more widespread use of standards. Rather, the Monitoring Council will use improved data access and standardization as the basis for conducting higher-level syntheses and interpretations of monitoring data at the statewide level. The ready availability of statewide data will enable the Monitoring Council to task its workgroups with developing and applying statewide performance assessments that in the past could not be conducted because of problems like that illustrated in Figure 1.

2.2.1 A flexible organizational structure

The Monitoring Council has established an organizational structure based on theme-specific workgroups operating within common policies and guidelines established by the Monitoring Council. The Monitoring Council will either pose the core assessment questions itself or review and sign off on questions developed by the workgroup. This is a critical initial step because the assessment questions structure the remaining features of the web portal, both the visible ones such as maps, assessment products, and links to other web-based resources, as well as the invisible ones such as methods standardization and data management procedures. The Monitoring Council has established a basic template for the core assessment questions, modeled after those in the four prototype portals, that focuses on map-based depiction of status and trends at a range of spatial scales, and on the success of efforts to correct or improve problems.

Once established, workgroups are responsible for developing the web portal, creating appropriate standards for monitoring and assessment methods and data management procedures, and disseminating these standards to local and regional monitoring programs that generate raw data. The Monitoring Council will encourage and/or assist with outreach to additional potential partners and review and comment on draft assessment products and web portal prototypes. The Monitoring Council will also ensure that data management and integration procedures comply with developing State standards and are compatible with the CEDEN system and its network of regional data centers. Finally, the Monitoring Council will provide technical support as needed. The respective roles of the Monitoring Council and the workgroups are summarized in Table 1.

Table 1. Respective roles of the Monitoring Council and the theme-based workgroups on the six main monitoring program elements defined for the Monitoring Council's efforts in CWQMC (2008).

Monitoring program element	Monitoring Council role	Workgroup role
1. Strategy, objectives, design	Define core management questions Collaborate w/workgroup on assessment strategy Ensure compatibility with related themes Comment and review	Develop assessment strategy, detailed monitoring objectives and design(s)
2. Indicators and methods	Set goals for statewide standardization Comment and review	Develop, improve, standardize indicators and measurement methods Improve standardization statewide
3. Data management	Set basic standards, design principles Provide technical support	Implement data management procedures, user interfaces, applications
4. Consistency of assessment endpoints	Ensure assessment targets questions at statewide scale Set goals for statewide standardization Comment and review	Develop new or apply existing assessment methods Improve standardization statewide
5. Reporting	Define reporting standards Set goals for improved efficiency of existing reporting functions Comment and review	Design and produce assessment products Develop reporting functions to support agency reporting functions
6. Program sustainability	Conduct periodic program evaluations Create and update program plans Obtain needed resources	Implement responses to program evaluations Provide needed input to program planning Predict and highlight resource needs

Within this general framework, this year's efforts have highlighted the need for flexibility in both working relationships and technical approaches, given the different points from which each effort started, the level of existing standardization and coordination, and the specific technical challenges posed by each theme. For example, the Wetlands workgroup included a comprehensive range of stakeholders from its

inception, while the Safe to Swim workgroup's membership initially focused only on ocean beaches. Similarly, the Safe to Swim web portal was designed and implemented by State Water Board staff, while the Wetlands and Safe to Eat Fish and Shellfish web portals were developed by external partners. The Safe to Drink web portal is structured around the State Water Board's GeoTracker / GAMA system, which was developed independently to address a separate piece of legislation.

While the Monitoring Council's workgroups are organized around a single theme and have a statewide focus, there are programs that operate at the smaller watershed or regional scale, but that nevertheless are potentially useful partners for the Monitoring Council's efforts. These regional scale programs have a wide range of missions and sponsors, ranging from volunteer water quality monitoring to collaborative watershed assessments and large-scale ecosystem monitoring and restoration programs. The Monitoring Council's organizational structure provides three ways to coordinate with programs focused on the regional scale:

- Supporting standardization of monitoring and data management methods to ensure that key data types are available to and usable by the Monitoring Council's theme-based web portals
- Incorporating specific elements of regional programs into workgroup efforts to develop statewide assessments (e.g., stream bioassessment monitoring, which could be input to the statewide healthy streams sub-theme)
- Creating new sub-themes to represent integrated assessments of aquatic ecosystem health at the regional scale (e.g., integrated assessments of the Delta)

The Monitoring Council is willing to support a range of such relationships, as long as they are compatible with the Monitoring Council's philosophy. Key to any development path, however, is the maintenance of strong relationships with the entities with primary responsibility for conducting statewide assessments for each theme. The Monitoring Council's approach depends on their involvement to assure the accuracy and relevance of all aspects of the web portal and to ensure adequate access to needed data and expertise.

2.1.2 Performance measures

The Monitoring Council adopted a set of performance measures and benchmarks (Table 2) based on USEPA's 2003 report Elements of a State Water Monitoring and Assessment Program (USEPA 2003), but condensed USEPA's list of ten elements to six. A description of these six performance measures can be found in CWQMC (2008) and each workgroup will use these measures to evaluate existing water quality monitoring, assessment, and reporting efforts in order to develop specific actions and estimate funding needs necessary to coordinate and enhance those efforts.

Table 2. Benchmarks associated with each of the six performance measures used by the Monitoring Council and the theme-based workgroups to evaluate existing web-portals and assessment programs and to track the Monitoring Council's progress toward meeting the goals of each web portal development effort.

Evaluation criteria	Rating benchmarks / performance measures
1. Strategy, objectives, design	Low: No core questions; no, or many undifferentiated, target audiences; poorly articulated or conflicting objectives; uncoordinated monitoring efforts not focused on questions or objectives Medium: Core questions and target audiences implicit in program design; objectives implicit but only partly standardized and not directly used to structure design effort High: Core questions standardized, clearly stated, and focused on

Evaluation criteria	Rating benchmarks / performance measures
	specific audience(s); clearly stated and common objectives address standardized core questions and inform all aspects of design
2. Indicators and methods	Low: Indicators and methods uncoordinated, not validated; no QA procedures or plan Medium: Indicators and methods validated but not standardized statewide; QA procedures exist but are poorly matched to objectives and not standardized statewide High: Standardized, scientifically validated, and clearly documented indicators, methods, and QA procedures that match monitoring objectives
3. Data management	Low: No data management procedures or documentation Medium: Data management procedures exist but are not standardized statewide and only poorly support access to data High: Standardized and clearly documented data management procedures are standardized statewide and fully support access to data at multiple levels
4. Consistency of assessment endpoints	Low: No data analysis or assessment procedures used or documented Medium: Data analyzed but methods not standardized; assessment tools exist but not fully validated or standardized High: Data analysis methods and assessment tools fully validated, clearly documented, and standardized statewide
5. Reporting	Low: No reporting process or products Medium: Intermittent reports, available with some effort High: Readily available regular reports focused on core questions and objectives; ability to create user reports from multiple perspectives
6. Program sustainability	Low: No systematic program evaluation, planning, or long-term funding devoted to infrastructure needs related to standardization and data integration Medium: Intermittent internal program review and planning that may or may not include infrastructure needs; limited funding for infrastructure High: Regular external program evaluations and planning for all program needs

2.1.3 A single, global point of entry

A central design feature of the Monitoring Council's approach is that all theme-based web portals, and the water quality data and assessment products they provide, will be accessible through a single, global point of entry. This point of entry has been established at http://www.waterboards.ca.gov/mywaterquality (Figure 2), and the Aquatic Ecosystems theme provides access to a series of sub-themes that address a variety of aquatic ecosystem types (Figure 3). Figures 2 and 3 also illustrate the page design the Monitoring Council has established for the higher-level entry points, and that the theme-specific workgroups must comply with.

The main function of this global point of entry is to solve the long-standing, fundamental data access problem, namely, that it can be confusing and time consuming to find data, assessment products, and background information relevant to a particular question or issue. By providing a direct connection to the individual theme-based web portals, this global entry point will also provide organized access to a broad range of relevant databases and websites maintained by other entities. For example, the Safe to Drink web portal provides a link to the GeoTracker / GAMA program website, the Safe to Swim web portal to Heal the Bay's beach report card website, and the Safe to Eat Fish and Shellfish portal to the website of the Office of Environmental Health Hazard Assessment (OEHHA), in addition to a large number of additional state, federal, and NGO websites and databases.

Figure 2. The Monitoring Council's global point of entry to monitoring and assessment information for all theme-based web portals.[Replace with current version]

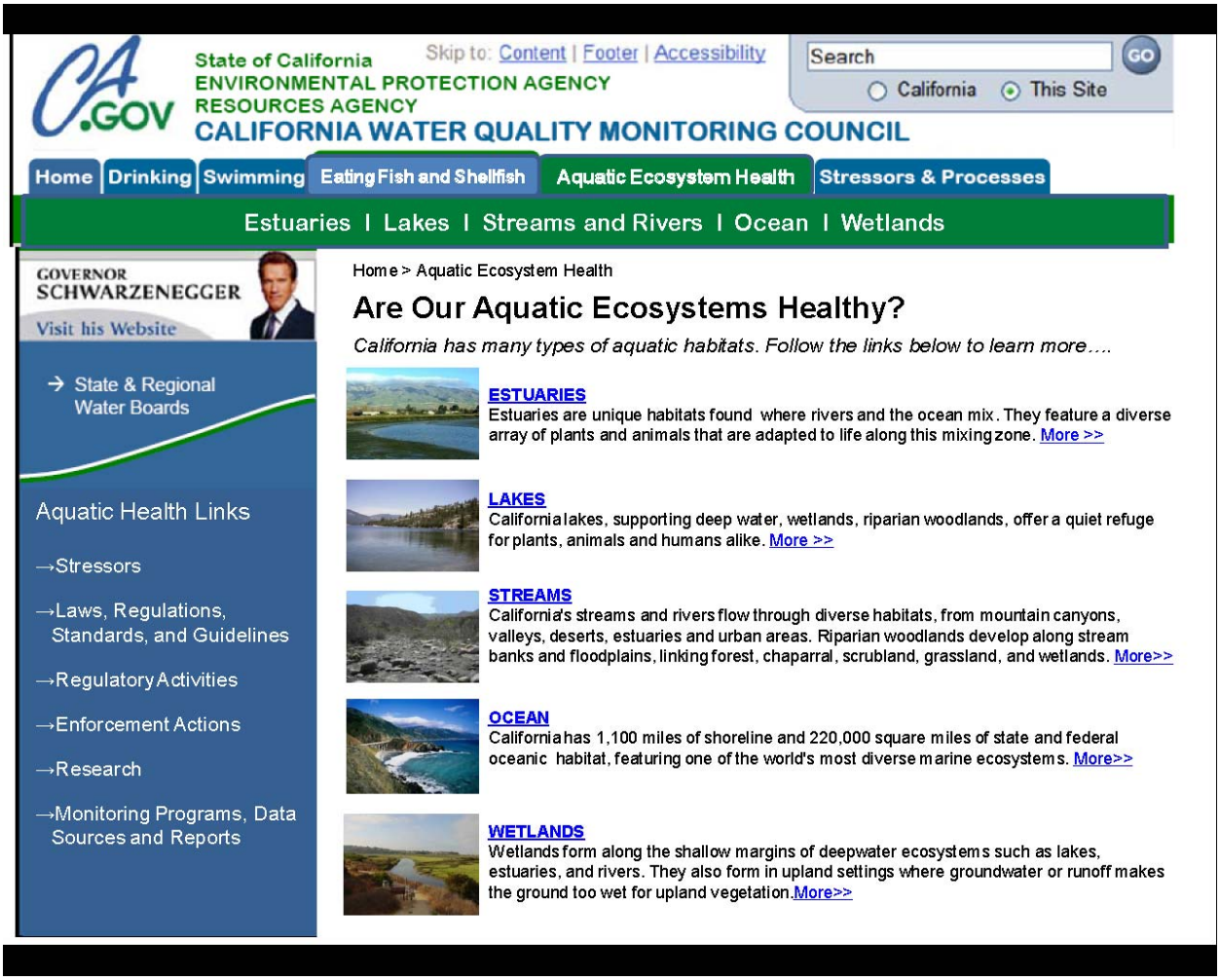

Figure 3. The Aquatic Ecosystems web portal, which provides access to a number of separate sub-themes focused on different categories of aquatic ecosystems. [Replace with more current and higher-quality version]

2.1.4 Standardization of core program elements

Improving the standardization of monitoring program elements is crucial to the successful functioning of the theme-based web portals (see Table 1, especially criteria 1 – 4). Inconsistent monitoring designs and/or methods, indicators, or assessment approaches make it impossible to present credible and reliable assessments at the statewide scale. Thus, making consistent progress toward improved standardization is an important part of the Monitoring Council's workplan (see Chapter 3).

This year's experience with the four prototype portals, as well as experience from past attempts at establishing standards, suggests that the Monitoring Council will encounter a range of situations regarding the standardization of monitoring designs, indicators, measurement methods, and assessment approaches. As a result, standardization will not follow the same pathway, or present the same challenges, for each theme, and different sets of standards will be applicable for different themes. For example, beach water quality monitoring programs apply the same assessment standards, based on AB 411, but have different monitoring design philosophies, with the result that measures of the frequency and magnitude of beach closures have different meanings for different programs. As another example, the wetlands theme is dealing with a situation in which standard monitoring methods have been agreed on, but there is no

agreed-on framework for interpreting monitoring results and arriving at consistent conclusions about wetland status.

As explained in CWQMC (2008), not all aspects of all programs require statewide standardization. The Monitoring Council will therefore work with each workgroup to identify program elements that require statewide standardization to support comprehensive assessments and those that can vary regionally to support local needs. Where national or state standards already exist, the Monitoring Council will encourage adoption of the highest-level standards available. In all cases, however, the Monitoring Council's philosophy (see Sections 1.1 and 2.1) is to present available information in a web portal as soon as some statewide information is available, even if it contains data gaps and/or inconsistencies. As explained above, this approach creates the structure and motivation for a transparent process of continual improvement of data, methods, and assessment products.

2.1.5 Improved data management

The Monitoring Council's approach to improving data access is premised on providing a global point of access to a series of theme-based web portals. These in turn enable access to a wide range of other data sources as needed to fulfill the web portals' analysis, assessment, and reporting functions. This will require comparable data statewide, technical support for infrastructure and tool development, and the ability for users to query and download a variety of data and assessment products.

Work on the prototype web portals during 2009 has demonstrated both the potential and the challenges of this goal. Fully implementing the set of web portals envisioned will require finding, accessing, and integrating many different data types from a large number of sources, and providing monitoring data and products to users with valid, often wide, differences in needs and perspectives. These challenges are not limited to the Monitoring Council's efforts, and are in fact an important issue for the State as a whole. The Office of the Chief Information Officer recently released its Statewide Data Strategy Report, which describes the State's approach to overcoming widespread problems related to data access and integration. While it lays out basic principles for the design, functioning, and integration of the State's data management systems, it also allows for needed flexibility as each agency develops its own solutions and strategies.

The Monitoring Council's approach is compatible with the State's strategy and is based on establishing locally centralized access and data input points at regional data centers. These data centers are linked as the distributed California Environmental Data Exchange Network (CEDEN) network (Figure 4), which also relies on the California Environmental Resources Evaluation System (CERES) metadata catalog. CEDEN is a distributed enterprise system intended to be flexible enough to accommodate multiple requirements and the CEDEN nodes fulfill the role of intermediary between larger state systems and small to medium data providers. CEDEN's architecture (see Appendix 2 for a description of CEDEN's architectural design) has been designed to create a long-term solution for delivering complex, scalable, user-friendly applications and information to a wide variety of users (see Appendix 3 for a description of CEDEN's functional specifications).

As described more fully in Appendix 2, CEDEN is committed to participating in the USEPA Exchange Network and in implementing their standards for service oriented architecture (SOA) and web services. These frameworks structured the initial design and implementation of CEDEN, which became operational in 2009. However, the system still requires a substantial amount of development, both of its basic infrastructure and of applications needed to support the theme-based web portals, and this effort is outlined in the workplan in Chapter 3.

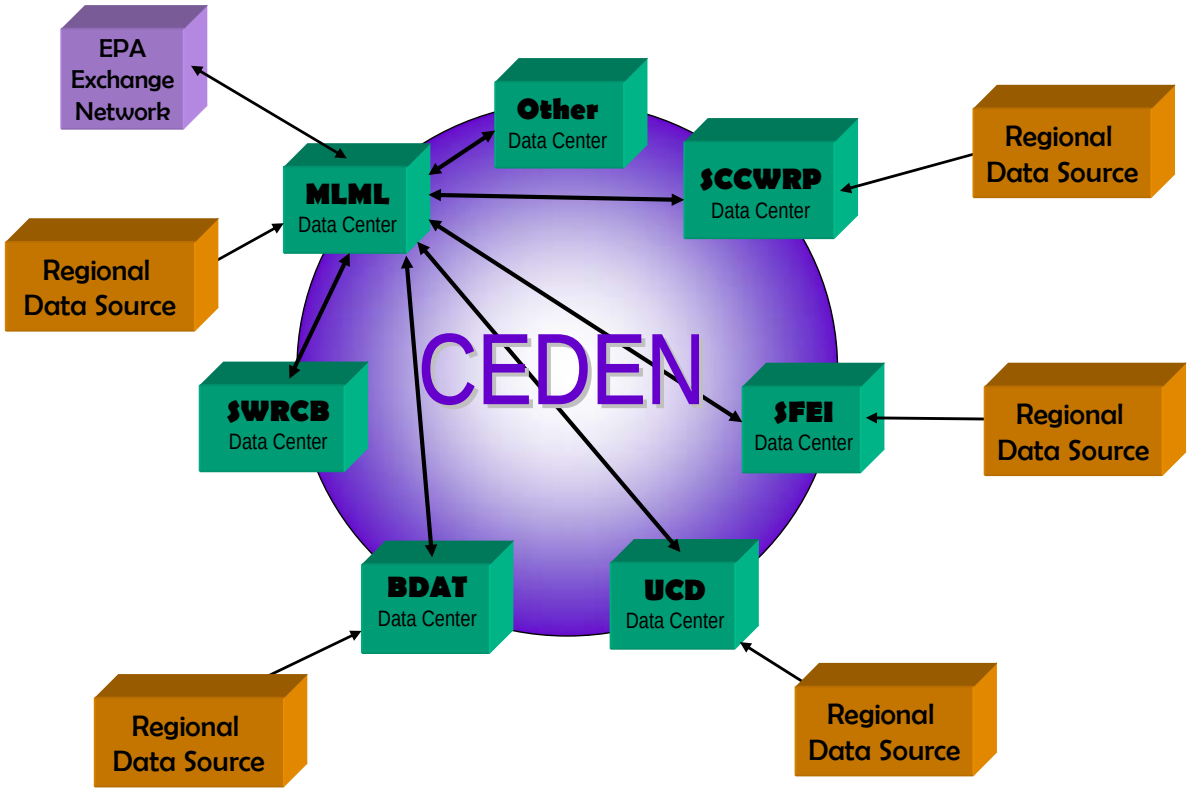

Figure 4. Schematic depiction of the CEDEN network, illustrating the relationships of the regional data centers to each other, to regional data sources, and to the external EPA Exchange Network.

Chapter 3: The Monitoring Council's Ten-Year Workplan

The Monitoring Council has developed a ten-year workplan (Workplan) to implement the approach described in Chapter 2. The Workplan is divided into three phases, with different technical and management challenges and levels of effort allocated to each:

- Start-up: Years 1 – 2
- Development: Years 2 – 8 (overlapping with Start-up)
- Long-term maintenance: Years 9 – 10 (and beyond)

The Workplan includes two complementary and parallel types of effort (Figure 5) essential to accomplishing the five-part solution described in Section 2.2. The left-hand side of Figure 5 represents effort carried out at the level of the individual theme-based workgroups. This effort would in general follow the approach developed in 2009 for the four prototype themes, applying lessons learned during those initial efforts. The right-hand side of Figure 5 represents tasks that are the direct responsibility of the Monitoring Council because they relate to establishing and maintaining the program's technical, management, and financial infrastructure.

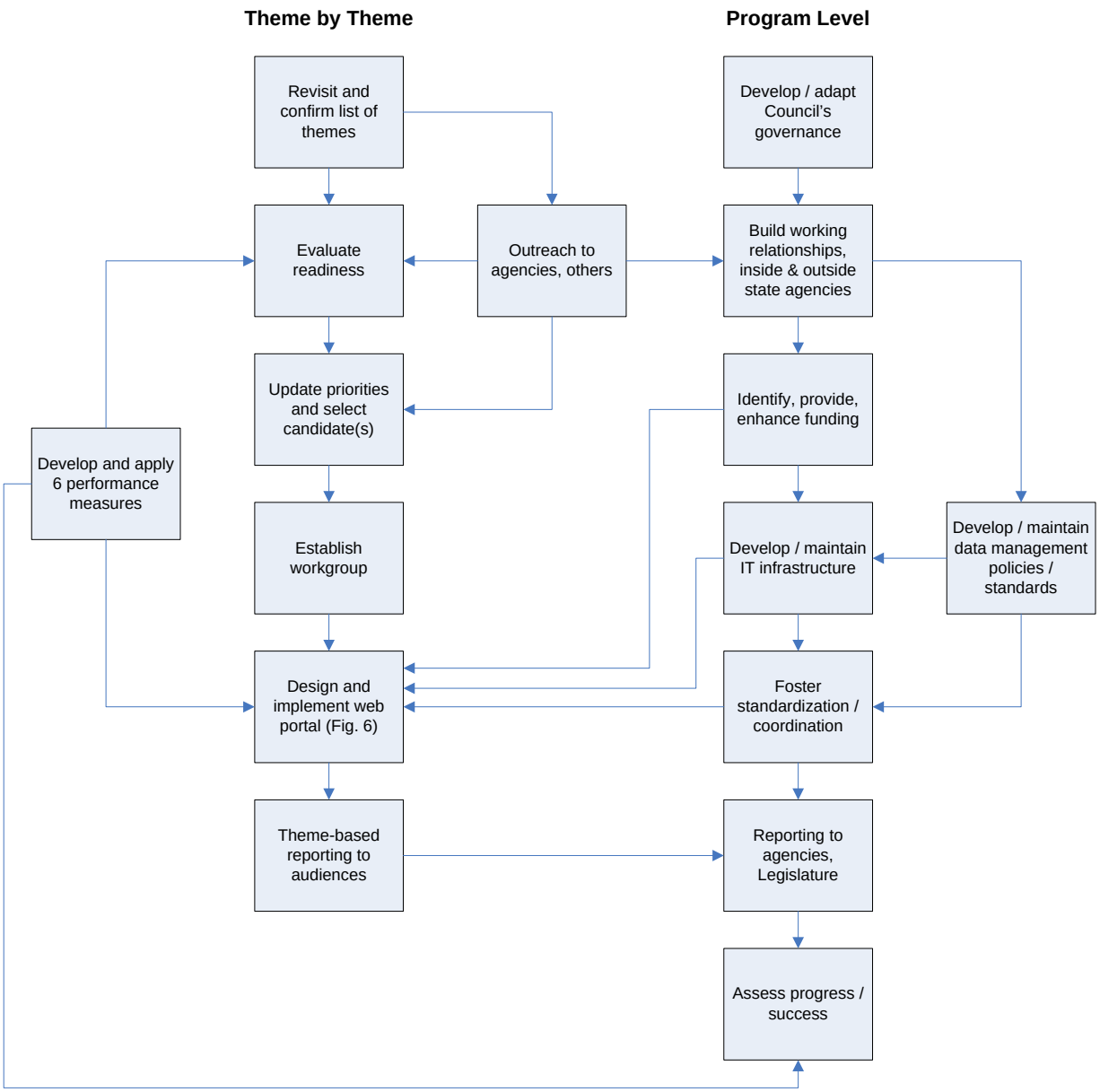

Figure 5. Parallel tracks needed to implement theme-based monitoring and assessment within the context of web portals. The Comprehensive Strategy focuses primarily on the right-hand side of the figure.

3.1 Theme-by-theme tasks

Specific tasks required to prioritize themes for action, establish workgroups, and develop a series of individual web portals are shown on the left-hand side of Figure 5. The following discussion follows the figure from top to bottom.

3.1.1 Prioritize targets for development

The list of potential themes (see Table 3) will be periodically revisited to determine if adjustments are required. For example, the Monitoring Council recently reorganized the Aquatic Ecosystems theme

(Figure 3) to streamline the development of web portals for the associated subthemes. The Monitoring Council will then assess the readiness of each theme by evaluating its performance on each of the six performance measures (see Section 2.2.1, and Appendix 3 of CWQMC 2008).

The Monitoring Council will then prioritize themes for development, using a prioritization scheme based on the following three criteria:

- Level of concern to the public and managers
- Level of effort involved (based on each theme's score on the six performance measures, presented in detail in Appendix 3 of CWQMC (2008))
- Near-term opportunities (i.e., low-hanging fruit) involving interested monitoring / assessment programs, immediate sources of funding, or situations that demonstrate technical methods or institutional arrangements that further the goals of the Statute

This recent prioritization indicates that XXX, XXX, XXX, and XXX are the highest priorities for the next set of web portals.

The Monitoring Council's emphasis on periodic prioritization recognizes the fact that all themes and subthemes cannot be addressed immediately. Implementation must therefore optimize the effectiveness of available resources, address first those issues of most concern to managers and the public, take advantage of existing infrastructure, and build momentum and support for the overall concept of expanding the use of theme-based web portals. Table 3 illustrates how the Monitoring Council has applied the three prioritization criteria. The safety of drinking water received the highest level of concern, with consumption safety and swimming safety the next priority. In general, the status of aquatic life is a lower priority, with exceptions at certain times and places for some audiences. The level of effort needed to meet the goals of the Statute for each portal is rated on four-point scale, based on each theme's scores on the performance measures. High scores correlate with a lower level of effort required. Themes that have expressed an interest in participating in the Monitoring Council's activities, have access to independent sources of funding, and/or have an institutional infrastructure to promote coordination and access are rated as the best opportunities.

Table 3. Summary results of the prioritization exercise. For each criterion, lower numbers represent a higher priority. The overall priority is the simple average of the individual ratings on three separate criteria. [Will be revised to reflect decisions about how to organize aquatic resources and wetlands]

		Prioritization Criteria		
Theme-based portals	**Level of concern**	**Level of effort**	**Opportunity**	**Overall priority**
Is our water safe to drink?				
Surface water	1	1	3	1.7
Groundwater	1	2	1	1.3
Water at the tap	1	3	2	2.0
Is it safe to eat fish and shellfish from our waters?				
Sportfish	2	2	1	1.7
Shellfish	2	1	2	1.7
Is it safe to swim in our waters?				
Freshwater	2	4	3	3.0
Beaches, bays, and estuaries	2	1	1	1.3

	Prioritization Criteria			
Theme-based portals	**Level of concern**	**Level of effort**	**Opportunity**	**Overall priority**
Are our aquatic ecosystems healthy?				
Wadeable streams	3	1	1	1.7
Rivers	3	3	3	3.0
Lakes	3	4	3	3.3
Coastal waters				
Shallow marine reefs	3	1	2	2.0
Intertidal	3	1	2	2.0
Subtidal benthos	3	1	2	2.0
Enclosed bays and estuaries	3	2	2	2.3
Wetlands	3	2	1	2.0
Fisheries				
Anadromous fish	2	2	2	2.0
Freshwater fish	3	4	3	3.3
Marine fish	3	3	3	3.0
Invasive species	3	2	3	2.7
Harmful algal blooms	3	1	1	1.7
What stressors and processes affect our water quality?				
Loadings	3	4	4	3.7
Flows	3	1	4	2.7
Levels of contamination				
Water				
Freshwater	3	4	4	3.7
Marine	3	2	4	3.0
Sediment				
Freshwater	3	4	4	3.7
Marine	3	2	3	2.7
Aquatic life				
Freshwater	3	4	4	3.7
Marine	3	3	2	2.7
Landscape maps	3	3	2	2.7
Measures of climate change	2	1	3	2.0
Ocean acidification	2	4	3	3.0

3.1.2 Establish and task workgroup

The Monitoring Council will then establish workgroups for each of the high priority themes and subthemes. While there is a division of responsibility between the Monitoring Council and the workgroup (Table 1), there is no set formula for how workgroups are established and their members selected. In general, the Monitoring Council anticipates the circumstances shown in Table 4, illustrated with the four prototype web portals addressed in 2009.

Table 4. Possible circumstances the Monitoring Council will face in establishing workgroups to address web portal development for each theme and subtheme. Prototype themes addressed during 2009 are placed in the framework as illustrations.

	Lead responsibility clear	Responsibility split
Workgroup exists and complete	Safe to Eat Fish and Shellfish	Wetlands
Workgroup exists but incomplete	Safe to Swim Safe to Drink	
No workgroup		

Depending on the circumstance, the Monitoring Council could simply adopt an existing workgroup, as it did with the Wetlands and Safe to Eat Fish and Shellfish workgroups, or adopt an existing workgroup and, as work proceeds, reorganize and/or expand the workgroup to include the needed range of expertise and perspectives. For example, the Monitoring Council reorganized the Safe to Swim workgroup to foster a statewide perspective and will shortly expand both the Safe to Swim and Safe to Drink workgroups to capture, respectively, the perspectives of inland monitoring programs and users of the information provided by the web portal. Where no workgroup currently exists, the Monitoring Council will establish one based on discussions with stakeholders both within and outside of State agencies.

The Monitoring Council will meet with representatives of each workgroup to develop a written charge or workplan for the workgroup. Existing web portals will provide examples of the structure, functionality, and look and feel required, and the Monitoring Council at this stage will also clarify the data management and data integration standards. Most importantly, the Monitoring Council will define the core management questions around which the web portal will be constructed. At the moment, the Monitoring Council and its workgroups are operating on the basis of "handshake" agreements. While these have sufficed for the four prototypes, a more formal relationship will be needed as the number and variety of workgroups increases (see Section 3.3.1).

3.1.3 Design and implement web portal

Working from its charge, the workgroup will design and implement the theme-based web portal. The process (Figure 6) will follow that used to develop the four prototypes during 2009, with the addition of more formal procedures for identifying data gaps, applying State and Monitoring Council standards, and feeding adjustments back to monitoring programs to improve their standardization and their ability to support statewide assessments. This process locates detailed design responsibility at the workgroup level, while providing for input and review by the Monitoring Council at appropriate points in the process (see also Table 1). Implementing this process will require additional staff support for the Monitoring Council.

The process illustrated in Figure 6 places the definition of core management questions and assessment products at the front end of the web portal design process. This reflects the Monitoring Council's fundamental philosophy that the web portals will be effective only to the extent that they are question driven and that statewide assessments are targeted directly at answering these questions.

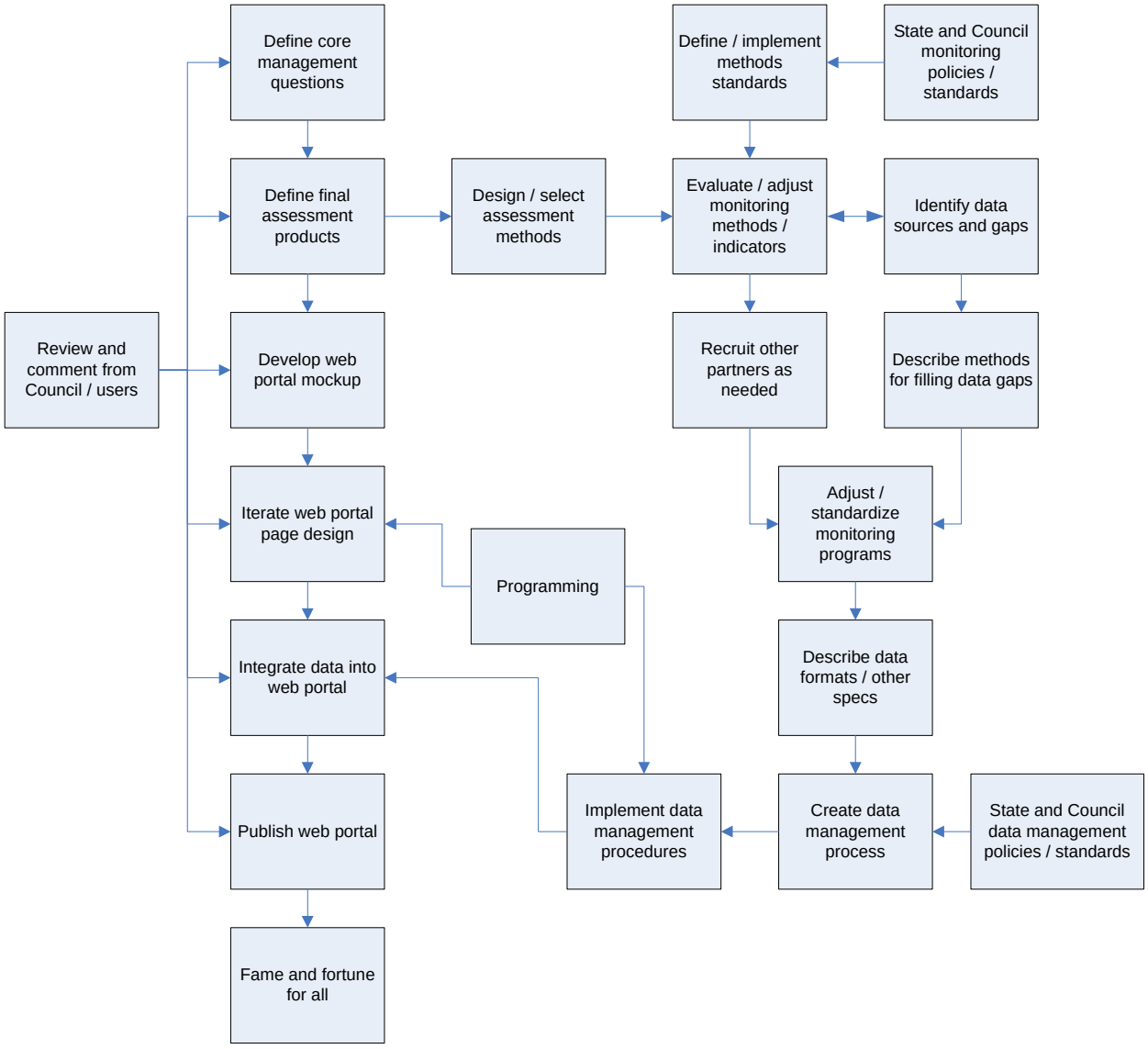

Figure 6. The process for designing and implementing individual theme-based web portals.

3.2 Program-level workplan schedule

Tasks required to develop and implement the Monitoring Council's programmatic infrastructure are shown on the right-hand side of Figure 5 and are the core responsibilities of the Monitoring Council itself. The effort involved in carrying out these tasks, and supporting the theme-by-theme tasks shown on the left-hand side of Figure 5, can be split into three developmental phases:

- Start-up: Years 1 – 2
- Development: Years 2 – 8 (overlapping with Start-up)
- Long-term maintenance: Years 9 – 10 (and beyond)

All tasks shown in Figures 5 and 6, and discussed in Section 3.1, are relevant to each developmental phase. However, the specific technical and management challenges will differ from phase to phase, as will the staffing, cost structure, and level of effort needed to accomplish each task. The following sections

briefly describe the tasks specific to each phase of the Workplan. Tasks are discussed in terms of the five-part solution described above (Section 2.2):

- Organizational structure with common policies and guidelines
- Performance measures applicable to all themes and web portals
- A single, global point of entry
- Standardization of monitoring and assessment methods that achieves an appropriate balance between statewide consistency and regional flexibility
- Database and data management standards necessary for more efficient data access and integration

3.2.1 Start-up: Years 1 – 2

The start-up phase will encompass 2009 and 2010 and will continue and expand the foundation building efforts begun in 2009, targeting a series of specific milestones. Work during this phase will focus primarily on completing the development of policies and procedures, solidifying relationships with key partners, and expanding web-portal development efforts.

Organizational structure: The Monitoring Council will continue to develop its governance structure and formalize it as needed. Written procedures will be established for recruiting replacement members and for deciding whether and how the Council's size and makeup could be adjusted. The respective roles of the Monitoring Council and its workgroups will be described in more detail and a format for a written agreement developed. The Monitoring Council will also further examine the three types of authority described in CWQMC (2008) for ensuring recommendations, especially regarding standardization, are implemented, i.e., voluntary adoption, permit/grant/contract requirements, and legislation. These mechanisms will be described more completely and procedures investigated for implementing them in different situations.

The Monitoring Council will continue its structured outreach to potential partners in State government, local and regional agencies, and non-governmental and volunteer entities. Outreach will be targeted primarily at entities directly involved in monitoring and assessment related to the highest priority themes and sub-themes. However, the Monitoring Council will also respond to spontaneous overtures from other potential partners to investigate whether these may provide unexpected opportunities to progress toward the Monitoring Council's objectives. Further developing relationships with upper-level management in key partner agencies and departments will be a high priority, as will developing a closer working relationship with managers involved in developing the State's data management policies.

The Monitoring Council will assess the workload associated with the developing program described here and determine the staffing requirements needed to support this effort. [Where does this information go to?]

Performance measures: The Monitoring Council will develop more detailed descriptions of the six performance measures (Table 2) and a systematic method for applying them to a wide range of web portals and the monitoring and assessment programs they are based on. It will be important to improve the consistency of the performance measures and to determine whether the existing qualitative scoring system is adequate. The Monitoring Council will develop a plan for applying the performance measures to its web portals on a regular schedule in order to assess progress and highlight specific areas for improvement. The plan will include a means of reporting results to the program's staff, partners, and audiences.

Single, global point of entry: The Monitoring Council will maintain its main web portal, complete the initial phase of development for the first four prototypes, identify and begin needed enhancements to the

prototypes, and begin development of the next set of web portals. This will involve establishing and tasking workgroups, developing core management questions, and embarking on the other tasks described in Section 3.1 and Figure 6.

Standardization: Based on its experience with the four prototypes, the Monitoring Council will develop a more detailed approach to standardization of those aspects of monitoring programs needed to support statewide assessments of the core management questions for each web portal. This will involve developing procedures for the workgroups to use the performance measures to identify data gaps and methods inconsistencies that undermine the breadth and comparability of assessment results. It will also require the Monitoring Council to develop procedures for and tracking workgroups' progress toward resolving these issues. At a another level, the Monitoring Council will identify other sources of inconsistency that cut across individual web portals and that will require more direct involvement by the Monitoring Council to resolve.

Data management: The Monitoring Council will stay abreast of the State's developing data management policies and ensure adequate channels of communication are in place. The Monitoring Council will also use development of the prototype web portals to identify data management issues that must be resolved at a higher level, implement the initial phase of CEDEN, and identify policies and procedures needed to ensure that web portals are both compatible with CEDEN and make effective use of its capabilities. In particular, the Monitoring Council will establish a data management workgroup with representation from CEDEN and the regional data centers, the Monitoring Council, the Office of the Chief Information Officer, policy-making IT staff at Cal/EPA and the Resources Agency, and key user groups. As with the theme-based workgroups, the data management workgroup will operate under a charge established by the Monitoring Council.

3.2.2 Development: Years 2 –8

The development phase will encompass 2010 to and 2016 and will focus on fully implementing the policies and procedures defined in the Start-up phase, revising them as experience dictates, and moving into the routine development and publication of the series of theme-based web portals. [obtain needed funding]

Organizational structure: The Monitoring Council will fully implement all policies and procedures developed during the Start-up phase, including establishing more formal working arrangements with the theme-based workgroups, conducting routine outreach and relationship building/maintenance with existing and potential partners, and formalizing mechanisms for ensuring that standardization policies are fully implemented and complied with.

Performance measures: The Monitoring Council will implement regular assessments of its web portals and report the results to program's staff, partners, and audiences. In addition, the Monitoring Council will routinely apply the performance measures to high priority themes and sub-themes as they are being considered for development, in order to produce more detailed and accurate estimates of effort required for web portal development.

Single, global point of entry: The Monitoring Council will stabilize the design of its main portal and complete the full implementation of all features intended to support data access, analysis, visualization, downloading, and other assessment applications. The second set of web portals will be completed and a series of workgroups established to continue the regular production of additional web portals.

Standardization: The Monitoring Council will make the use of the performance measures to identify inconsistencies at the level of individual themes and web portals a standard workgroup practice, and will support, encourage, and require workgroups to resolve inconsistencies and will track each workgroup's

progress toward needed standardization. The Monitoring Council will also work with its partners to develop more global monitoring standards that cut across individual themes and will publish these standards to all workgroups and incorporate them into the performance measures.

Data management: The Monitoring Council will complete the implementation of CEDEN, including the regional data centers and will publish documentation, policies, and procedures necessary for maintaining the system. The Monitoring Council will also ensure that the data management workgroup stays abreast of new directions in the State's data management policy, as well as of evolving monitoring requirements and users' needs that call for new system capabilities.

3.2.3 Long-term maintenance: Years 9 – 10 (and beyond)

The long-term maintenance phase will extend from 2017 forward and will focus on maintaining and adapting the policies, procedures, funding, and the technical infrastructure needed to ensure the web portals remain both operational and relevant. This will involve periodically reevaluating all aspects of the Monitoring Council's five-part solution to assess their continued relevance and performance.

3.3 Budget

[need more specific guidance on this section]

[Table 5 is one possible way to present budget requirements. See next page.]

Chapter 4: Recommendations

[To be developed.]

Table 5. Sample overall budget structure for Monitoring Council's Comprehensive Strategy.

Strategy Components	Start-up			Development			Maintenance		
	Staffing	Hardware / Software	Contracts	Staffing	Hardware / Software	Contracts	Staffing	Hardware / Software	Contracts
Organizational structure									
Outreach, relationships									
Governance, policies									
Establish workgroups									
Support workgroups									
Performance measures									
Develop measures									
Implement measures									
Technical support									
Reporting									
Portals									
Council portal									
Prototype theme portals									
Multiple theme portals									
Maintenance, updates									
Standardize methods									
Approach to standards									
Implement standards									
Maintenance, updates									
Database / data management									
CEDEN									
Regional data centers									
Interface w/state policy									
Develop policies									
Integrate portals w/CEDEN									
Implement state policy									
Develop tools									
Maintenance, updates									

Appendix 1: SB 1070 Requirements Matched to Comprehensive Strategy Components

The following table illustrates which aspects of the Monitoring Council's Comprehensive Strategy address each specific requirement of SB 1070.

SB 1070 requirement	Detail	Strategy component
Recommend improvements to monitoring	(4) The monitoring council shall review existing water quality monitoring, assessment, and reporting efforts, and shall recommend specific actions and funding needs necessary to coordinate and enhance those efforts. (5) (A) The recommendations shall be prepared for the ultimate development of a cost-effective, coordinated, integrated, and comprehensive statewide network for collecting and disseminating water quality information and ongoing assessments of the health of the state's waters and the effectiveness of programs to protect and improve the quality of those waters. (B) For purposes of developing recommendations pursuant to this section, the monitoring council shall initially focus on the water quality monitoring efforts of state agencies, including, but not limited to, the state board, the regional boards, the department, the Department of Fish and Game, the California Coastal Commission, the State Lands Commission, the Department of Parks and Recreation, the Department of Forestry and Fire Protection, the Department of Pesticide Regulation, and the State Department of Health Services. (C) In developing the recommendations, the monitoring council shall seek to build upon existing programs rather than create new programs. (6) … the monitoring council shall formulate recommendations to accomplish both of the following: (A) Reduce redundancies, inefficiencies, and inadequacies in existing water quality monitoring and data management programs in order to improve the effective delivery of sound, comprehensive water quality information to the public and decision makers. (B) Ensure that water quality improvement projects financed by the state provide specific information necessary to track project effectiveness with regard to achieving clean water and healthy ecosystems.	

Develop a comprehensive monitoring program strategy	(1) Utilize and expand upon the State's existing statewide, regional, and other monitoring capabilities and describe how the State will develop an integrated monitoring program that will serve all of the State's water quality monitoring needs and address all of the State's waters over time. (2) The strategy shall include a timeline not to exceed 10 years to complete implementation. (3) The strategy shall identify specific technical, integration, and resource needs, and shall recommend solutions for those needs.
Develop an agreement on Indicators	Agreement, including agreement on a schedule, with regard to the comprehensive monitoring of statewide water quality protection indicators that provide a basic minimum understanding of the health of the state's waters. Indicators already developed pursuant to environmental protection indicators for statewide initiatives shall be given high priority as core indicators for purpose of the statewide network.
Develop a Quality Assurance Management Plan	Quality management plans and quality assurance plans that ensure the validity and utility of the data collected.
Develop a method for compiling, analyzing, and integrating readily available information	This is to include data from waste discharge reports; volunteer monitoring groups; local, state, and federal agencies; and state and federal grant recipients of water quality improvement projects.
Develop an accessible and user-friendly electronic Data Management System	To the maximum extent possible, include the geospatial information on the data sites.
Develop a method for producing timely and complete water quality reports and lists	The reports and lists required are those required under Sections 303(d), 305(b), 314, and 319 of the Clean Water Act, and Section 406 of the BEACH Act.
Develop an update of the SWAMP needs assessment	The SWAMP program needs will change in light of the benefits of the increased coordination and integration of information from other agencies and information sources.

Appendix 2: CEDEN Architectural Design_1

[A more recent and up to date version may be available.]

Introduction

The purpose of this document is to describe the CEDEN enterprise architecture and the technologies that will be used. It will also explain why the design and technologies were selected, contrasting these choices with other options. The main components of CEDEN are listed below.

I. **Data Input** – Data will come into a Regional Data Center (RDC) from various sources. The system is designed to be flexible and will accept data submissions in a variety of ways. Examples of some of these ways are:

- **a. ODBC Connections** - The data may come from a regional data provider via an ODBC connection to a transfer database.
- **b. Web Data Entry** – The data may come from the web data entry application to the transfer database.
- **c. Web Services** – Data input via web services will be available using standards developed by the Environmental Protection Agency (EPA).
- **d. Miscellaneous Sources –** Data from other sources will also be accepted. This may include spreadsheets, flat files, etc.

II. **Transfer Databases -** The incoming data are loaded into transfer databases. The transfer database serves a very important purpose. Here the data remains in its original form from the provider, and this data store provides a complete transactional history from the moment of inception. The transfer database also provides a staging area so that the data can be cross-walked for load into a normalized Relational Database Management System (RDBMS). These data stores will be located in the DMZ.

III. **Relational Database Management System (RDBMS) -** Data is mapped from the transfer databases into the comprehensive, normalized RDBMS. The comprehensive database will store data in a canonical form and will be secured behind a firewall.

IV. **Application Data Marts** – The comprehensive database may grow to be very large over time. The sheer volume of data will make application data delivery a challenge, which is why the star schemas and other types of data marts will be created. Here the data can be aggregated and transformed for application specific purposes. The design of these marts will ensure good performance for customized applications. These data stores will be located in the DMZ for public access.

V. **Data Sharing -** Data and the information derived from it will be made available from CEDEN through a variety of methods. Some of these methods are:

- **a. Web Applications** - There will be a variety of web applications for converting data into information and delivering the information in a user-friendly way. Customized applications can be created upon request for any given group of data

users. They will serve a diverse population from casual users to data modelers and engineers. Data query tools will include maps and other forms of data selection.

b. **Data Synchronization -** CEDEN partners can share data by database replication. This is a powerful feature that provides synchronization between peers. Data partners may elect to have a copy of all or a subset of the comprehensive database in their own RDC.
c. **Web Services** – Data output via web services will be available using standards developed by the Environmental Protection Agency (EPA).

Discussion

CEDEN is a distributed, enterprise system that serves many purposes. One of CEDEN's goals is to be flexible enough to accommodate multiple requirements in order to serve the evolving business needs of the user community. This is not a "one-size-fits-all" sort of solution. The correct solution will be applied to each situation based on the requirements analysis and review of the specific business needs of any given case. This is why CEDEN provides a number of methods for data sharing, for instance.

Persistent Comprehensive Data Store

Why does the architecture include a persistent comprehensive data repository? Could the system have been designed without this? It is important to understand that there is an inherent difficulty in receiving data from multiple sources. One reason is that not all data definitions will be the same and the data may not be comparable. Data mapping will be necessary to provide a view of the data using a common vocabulary. In order to do comparisons from one data provider to another, sometimes across regions and involving multiple data categories, the data must be in a canonical structure. Certainly it would be possible to have each data user do the data mapping themselves for every data source they use, but this is no small feat. It is expensive, time-consuming, and requires extensive knowledge of the data semantics involved. It would be a massive effort to perform such a task for each new data request from every new data source. Having it done once at the RDC will ensure that the data is accessible to data consumers in a useable, consistent format.

Another reason to maintain a persistent data store is to maintain historical data. Data providers do not always store data indefinitely. After a certain period data is no longer available from the original source, but it would be kept in CEDEN's data stores permanently.

Also, since the data in this comprehensive data store is comparable and in a consistent format, highly sophisticated and complex aggregations and transforms can be applied. Online Analytical Processing (OLAP) data marts would be created as needed. This would offer reliable, efficient data access for extremely large queries and would meet the high-performance requirements for custom applications such as simulation modeling, geographical interfaces (maps), and Web Services.

Service Oriented Architecture

CEDEN is committed to participating in the US EPA Exchange Network and in implementing their standards for SOA and Web Services. The CEDEN architecture will provide Web Services from the comprehensive databases for any organization using the EPA standards. CEDEN data that is provided to data requestors will come from the comprehensive databases. This design does not include merely leaving all data at the local data stores and creating a web site that delivers data by doing real-time web service queries to these local data providers. Why did

CEDEN not design an SOA that limits data retrieval to real-time service requests from local providers? In other words, instead of serving data from a persistent, comprehensive data store, why not just get data using Web Services from the actual data provider at the time of the request? This issue relates to that of the persistent local data stores mentioned above. The same problems would arise, as well as others related to Web Services in general. Let us begin with the issues concerning Web Services first.

There are many problems and misconceptions concerning SOA and Web Services. First, SOA is not a new concept. It has been around for a long time. Also, SOA is not synonymous with Web Services. SOA could be implemented in other ways. But one of the most serious misconceptions is that Web Services provide an easy method of enterprise integration, making access to services and data from disparate vendors simple. This could not be further from the truth. In order to see why this is the case, we need to review what an SOA actually is.

A Service Oriented Architecture provides methods for system integration where resources are available as loosely coupled, interoperable "services". Whether the goal is to access services from different legacy systems within a particular organization or to create an application with a collection of services from different organizations, the problem is one of interoperability. In essence, these resources are difficult to integrate when they are on completely different platforms, written in different programming languages, etc. SOA promises to solve these problems by creating standard ways of creating services that can be used by anyone who uses the given paradigm, such as Web Services (SOAP/XML), REST, CORBA, etc.

In recent years, some companies have chosen to implement SOAs using Web Services. But why has the reality not lived up to the hype surrounding SOA and Web Services? [1,2] Is it because applying the SOA paradigm to a real-time system has problems of response time, support of event-driven, asynchronous parallel applications, complicated human interface support, reliability, etc.? Or perhaps it is because of complex, bloated, competing and conflicting specifications? [3,4,5]

The fact is, the main reason application integration has not been easy is because it is very difficult to get groups to use common semantic definitions.[6] SOA and Web Services have been marketed and available for many years, and yet we have seen that many integration efforts have not been successful. While it is nice that Web Services provides standards for message formats and transport, these are the least of the problems involved in integration. Harvard Business School professor Andrew P. McAfee argues that the benefits of Web services, in terms of data-sharing and communication exchanges, are not going to happen unless managers better integrate common standards. "In many ways simply normalizing or standardizing the connection protocols and data formats are the *easiest* of all problems in integration and business process management. The hard problem -- but the one with the most potential ROI -- is the *semantic* normalization of applications." [6] Indeed, the real problem goes back to semantic interoperability – that is, the meaning of the data. "The organizational challenge comes as all stakeholders get together and hammer out common definitions." [7]

So why don't companies just do that? McAfee points illustrates the problem with this example: "A couple of years ago Cisco called a halt to all new IT efforts and spent over eighteen months and $300 million to resolve exactly these kinds of dissimilarities within the company. If any Web services tools could have made this work cheap, fast, and 'easy', don't you think Cisco would have used them?" He also notes that some companies do have a way to create standards for their business partners: "Big companies have the power to convince or compel their partners to participate, and to shortcut negotiations by simply dictating terms. [With] Amazon and eBay […]

it's a ‘take it or leave it’ proposition. Amazon and eBay don't renegotiate Web services standards with each seller; they simply publish their standards and wait for other companies to adopt them.” [7]
It is clear that having each organization concoct its own brand of semantic definitions and service descriptions will not bring us any closer to interoperability, no matter how well Web Services can define communications protocols. Luckily for environmental data groups, however, the United States Environmental Protection Agency has succeeded in creating a standard that can be used by all participants. [8] These standards are in production and are being used all over the country. CEDEN proposes to utilize these standards for data sharing with Web Services. There are a number of important uses for Web Services and CEDEN is planning to make data available using these recognized, proven standards.

So while CEDEN will certainly support Web Services, the question remains: Why not just use Web Services for data delivery? Why not leave the data on the data providers systems and access it there real-time instead of integrating it into a comprehensive data store and providing services from there? The simple answer is that this would create a very limited functionality that would not begin to meet the business needs of CEDEN’s data users.

First, the data as stored by local providers is not in a canonical form. The mapping *must* occur somewhere. Would it be wise to place such complex logic into client-side transforms to convert to a common schema for each client request? These transforms would occur with every access to the system. So a query that involves fifteen different sources would have to make data mappings on the fly for each one, as well as any complex aggregations or transforms required. If one were to evade this sort of data mapping, the data could be presented in its raw form from the data provider, but this would render the data unusable in cases where data comparability was important. We are not talking about simplistic sorts of data mapping here, such as merely transforming location information into latitudes and longitudes. This issue concerns the data semantics of the entire data set we retrieve from the foreign data store.

Second, clients that are either temporarily or permanently out of service would eliminate the data source completely. Queries to these data sources would fail, or at least fail to return data during a real-time service request. And there would be no historical data if the local providers did not elect to maintain it. The only data that is available through the service request is that which is available at any given time on the data provider’s system.

Another issue with using XML as the only method of data sharing is that it is very inefficient and slow for large data sets. The requests for environmental data can result in extremely large data sets. Data compression may help move the effort from the network to the local machines, but this requires some degree of client configuration, either with extensions or with headers, and still increases the load on the servers performing the compression/decompression.

It is important to realize that not all of the data providers have local data systems. Some data is collected and stored in flat files or spreadsheets on various user’s laptops. Many of the data are collected by projects that are transitory. Once the project is over, there is no system to be connected. How would one set up web services to gather their data in a real-time application?

These things combined would make it very difficult to create complex applications using Web Services to gather data real-time from disparate, sometimes transitory, unmapped environmental data stores. Most Web Services applications in use today are limited to short, online transactions such as simple data queries, transmitting an order or acknowledging a shipment. [7] Even with increased bandwidth and computing power, attempting to create the kind of complex,

sophisticated real-time applications that CEDEN will be providing with that sort of environment would certainly be ill-advised.

Conclusion

CEDEN is a multi-faceted system that is intended to serve multiple purposes for many types of users. This architecture has been designed to meet business objectives and to create a long-term solution for delivering complex, scalable, user-friendly applications and information to data consumers. It cannot succeed at this with a simplistic paradigm or a limited set of tools. The goal is to create a powerful, flexible system that meets user requirements. This means performing business needs analyses, doing systems integration, collaborating with CEDEN's business partners, using established standards and protocols, implementing data security, and converting data into information for data users. CEDEN's design and technologies are being chosen with these concepts in mind.

[1] http://www.gartner.com/DisplayDocument?doc_cd=141033
[2] http://soa.sys-con.com/node/46564
[3] http://blog.webservices.or.kr/hollobit/roadmap/ws-specs/map.htm
[4] http://www.ibm.com/developerworks/library/ws-rmpaper/
[5] http://xml.coverpages.org/Chappell-WSRelSOAP.doc
[6] http://www.edithere.com/barry/2005/05/20#a1933
[7] http://hbswk.hbs.edu/item/4800.html
[8] http://www.epa.gov/webservices/index.htm

Appendix 3: CEDEN Functional Specifications

[A more recent version of this document may be available.]

California Environmental Data Exchange Network
CEDEN

Business Needs and Functional Specifications
DRAFT

Business Needs

State Water Resources Control Board (SWRCB), Regional Water Quality Control Board (RWQCB), university groups, private entities, bond-funded monitoring programs, and stakeholder entities collect large amounts of environmental data; in many cases, there is a great demand for these data to be available as a comprehensive, interoperable, and standardized data set by SWRCB, RWQCB technical/enforcement personnel, and decision makers. Unfortunately, within California, the many groups who collect monitoring data store them in different databases with inconsistent formats, Quality Assurance Quality Control (QA/QC), and data collection procedures. To help provide better access and improve compatibility, the Legislature has passed SB-1070 mandating the collaboration of the SWRCB and Resources Agency departments to coordinate, collect, and disseminate data, and passed SB-1049 requiring bond recipients collecting environmental data to follow data quality and dissemination standards developed by the

SWRCB's Surface Water Ambient Monitoring Program (SWAMP). To date, much of the ambient monitoring data collected with grant funds from Propositions 13, 40 and 50 are not being collected using required standards and these data are not being integrated and made accessible in a way that benefits the state's data users. Ambient monitoring data collected using Proposition 84 funds will also need a system for standardization, management, and dissemination and requires that they meet SWAMP standards. The California Department of Water Resources' Bay/Delta and Tributaries Program (BDAT) initiated the California Environmental Data Exchange Network (CEDEN) in collaboration with the SWRCB's SWAMP, the Resources Agency's (RA) California Environmental Resources Evaluation System (CERES) Program, the Environmental Protection Agency's (EPA) Environmental Exchange Network and Office of Water in region nine to help meet the state's need for integrated standardized ambient monitoring data. Coordination with the EPA was conducted by the California Environmental Protection Agency (Cal-EPA).

Currently the state's lead for the program is the Office of Information Management and Analysis (OIMA) at the SWRCB with continued participation from DWR, Cal-EPA, RA and EPA. CEDEN will help the SWRCB meet Goal Seven of its business plan, "Ensure that the Water Boards have access to information and expertise, including employees with appropriate knowledge and skills, needed to effectively and efficiently carry out the Water Boards' mission." Providing data and information in a standardized system promotes a comprehensive understanding of the status, trends, and environmental processes and mechanisms in California, and leads to more robust adaptive management strategies, improved information for 305(b) reporting, 303(d) listings, and Total Maximum Daily Load (TMDL) calculations.

Through CEDEN, the SWRCB will help meet Goal Five of its business plan and "Improve transparency and accountability by ensuring that Water Board goals and actions are clear and accessible, by demonstrating and explaining results achieved with respect to the goals and resources available, by enhancing and improving accessibility of data and information, and by encouraging the creation of organizations or cooperative agreements that advance this goal, such as establishment of a statewide water data institute." Access to data in the CEDEN network will lead to creating applications that convert monitoring data into refined information, which will help the SWRCB convey important aspects of surface water information to the public and legislature using media such as the Web. The distributed nature of CEDEN provides the opportunity for membership from Non-Governmental Organizations (NGOs), non-profit groups, universities and any groups who can benefit the network through their participation, in addition to state and federal agencies.

The standards specified in SWAMP and CEDEN will help "enhance consistency across the Water Boards, on an ongoing basis, to ensure our processes are effective, efficient, and predictable, and to promote fair and equitable application of laws, regulations, policies, and procedures." SWAMP and CEDEN will meet Goal Six of the business plan by providing a consistent way for the Water Boards to collect, QA, and disseminate ambient monitoring data.

This report identifies the business need for CEDEN and its functional specifications based on user needs.

Functional Specifications

The CEDEN Program works collaboratively with the monitoring community to promote comprehensive access to data through the promulgation of data and monitoring standards, data sharing using distributed architectures, and improved local database management.

CEDEN is comprised of members dividing into three broad categories:

- Data providers,
- Regional Data Centers (RDC), and
- Data users.

Data Providers

Data providers are monitoring programs that collect environmental monitoring data needed by various groups both within and outside the SWRCB. Data providers use local systems for management of collected data. These systems could be as simple as an Excel Spreadsheet or an Access database or as complex as an enterprise system built using Oracle or Informix databases. In the CEDEN system, data providers within a geographic region work with their local Regional Data Centers to form a regional environmental recording network.

Regional Data Centers

The RDCs are groups that work with many data providers within a geographic region of the state. The RDC cooperates with their data providers to:

- Integrate data in their region,
- Promulgate data standards,
- Provide information technology (IT) and appropriate tools to obtain data from data providers,
- Provide program information with metadata, and
- Provide and receive data from CEDEN.

RDCs are crucial to improving the data sharing between existing local monitoring programs and CEDEN in order to effectively and efficiently assess water bodies statewide. RDCs have advanced data management practices, which includes organizational and infrastructure resources to support the technological, scientific, and community outreach needed for an integrated environmental data management system within their region.

RDCs often provide data management, exchange, retrieval, and interaction services to data providers. RDCs require reliable mechanisms to upload the provider's data. In addition, they require devices to combine the data from different providers into a seamless interface, which provides data access to CEDEN. The RDC provide data using the CEDEN standards for naming attributes such as analytes and QA/QC operations to

determine the quality of data, and hence their potential applicability for various uses and decisions, as well as its interoperability relative to the SWAMP. This requires an intensive and ongoing effort to review and translate data.

Moss Landing Regional Data Center

The Moss Landing Regional Data Center (MLML RDC) is managed by the Marine Pollutions Studies Laboratory at Moss Landing Marine Laboratories (MPSL-MLML). The MPSL-MLML works with many different organizations to assist with program management, data collection, data management, verification and reporting of environmental data. MPSL-MLML past and current collaborations include EPA's Environmental Monitoring and Assessment Program (EMAP), Department of Fish and Game and Fish and Wildlife Service's Introduced Species Study (ISS), San Francisco Bay Regional Monitoring Program's Fish Survey, and California's Surface Water Ambient Monitoring Program (SWAMP). The MLML RDC is an extension of the data management, data coordination and quality assurance services provided to the SWAMP program. The MLML RDC is one of the regional data centers that comprise the California Environmental Data Exchange Network (CEDEN). The purpose of this document is to outline the anticipated roles and activities of this data center with respect to CEDEN.

I. MLML RDC will act as the primary source of data aggregation for all the data participants in CEDEN. That is, data will be uploaded, converted to a canonical form, and aggregated into a database that houses both current and historical water quality and environmental data for California. The data will be normalized into a schema of an Online Transaction Processing (OLTP) database on the MLML server. Other CEDEN RDCs will synchronize their data periodically with this MLML server.
II. This data center will also be a conduit for data going out to other RDCs. Other CEDEN participants may wish to have the aggregated database replicated in full or in part to their own local sites. This synchronization can be customized to serve the requested portion of the main database either on demand or in a scheduled manner.
III. The MLML RDC will also provide server setup and maintenance where needed for other RDCs. The personnel at the MLML RDC can install existing hardware and software and create the software to synchronize to the aggregate database. They can also assist with requirements for gathering, creating server specifications, and vendor contacts for purchasing.
IV. The CEDEN web site will be maintained by this data center. It will provide data access to the public as well as other customized, web-based applications.
V. The MLML RDC will create Online Analytical Processing (OLAP) data marts for various applications. These data marts will be created for the CEDEN web site applications as well as for other regional data centers as requested.
VI. The MLML RDC staff will be available to create applications for various regional data centers when necessary. The software development lifecycle will be followed in producing applications to fit the requirements of the requesting data center.

VII. The MLML RDC will continue to work with data users to transfer existing applications at regional and statewide level.
VIII. The MLML RDC will continue to work with the other RDCs to maintain consistent and comprehensive standards used within CEDEN, and provide documentation to assist in transferring data to CEDEN.
IX. The MLML RDC will also function as the RDC for the central coast region. The MLML RDC is currently working with regional data managers at SWAMP, National Oceanic and Atmospheric Administration (NOAA), the Central Coast Ambient Monitoring Program (CCAMP) and other groups to help format and transfer their data into the CEDEN system.
X. Moss Landing's RDC will assist new data providers in transferring their data into CEDEN using existing web and database data loading applications.
XI. In conjunction with the SWAMP data management team, the MLML RDC will maintain a helpdesk for general information and as a gateway to the CEDEN regional data centers.
XII. The MLML RDC will check all uploaded data for minimal data standards prior to transfer into CEDEN.
XIII. The MLML RDC will be available to evaluate any data sets against existing QA/QC standards at the data provider's request.

There are a variety of human resources required to accomplish these tasks. It will be important that personnel have the requisite skill sets to meet the goals of the Moss Landing Marine Labs Regional Data Center.

I. Systems Administration – the installation and maintenance of server hardware and software, creation of user accounts, data backups and restores, system troubleshooting, performance tuning, application of operating system updates, patches, and configuration changes, system audits and system documentation, establishment of network connectivity, and creation of a secure data processing environment.
II. Security Administration – the establishment of computer and network security, including the administration of security devices such as firewalls, as well as consulting on general security measures.
III. Network Administration - the deployment, configuration, maintenance and monitoring of active network equipment.
IV. Web Administration – the maintenance of web server software and configuration.
V. Application Development – the programming of application software.
VI. Software Engineering – implementation of the techniques and procedures of the software development process or lifecycle to systems development.
VII. Project Management – the analysis and design of objectives, creation of scope statement, project planning, risk management, resource management, cost estimation, resource allocation, management of quality, issues and change control.

VIII. Database Administration – the installation and maintenance of database software, database tuning, database backups, data analysis, data dictionary maintenance, and ensuring database system availability.

IX. Data Modeling - the process of creating a data model by applying formal data model descriptions using data modeling techniques. This would include the analysis, design and creation of conceptual, logical and physical data models of the data.

X. Environmental Biologist – assists in the creation of crosswalks, QA/QC and data translators to move data from existing systems into the MLML RCD.

Southern California Regional Data Center

The Southern California Coastal Water Research Project (SCCWRP) is a research institute in Southern California whose primary focus is to provide decision makers and other stakeholders with scientific information that allows them to manage our coastal and watershed resources. SCCWRP will act as the Southern California Regional Data Center and will collect and collate ambient water quality data collected within our region. SCCWRP is actively participating in projects that encourage the sharing of data for regional assessments and has developed web-based technologies for data submission processes that increase data quality.

SCCWRP will work with regional partners and use web-based data technologies to collect and provide ambient water quality data to CEDEN. This will include data not only collected by regional partners, but also data from historical surveys and regional surveys conducted every five years. SCCWRP is currently working with members of the Stormwater Monitoring Coalition, which includes both stormwater regulators and dischargers, and has obtained most of their current and historical water quality data for streams and rivers. Union of water quality data and CEDEN allows large-scale comparisons with other regions throughout the state of California. SCCWRP is also collaborating with various citizen-monitoring groups in southern California to add their water quality monitoring data to the Southern California Regional Data Center.

SCCWRP will continue to work with local monitoring groups in Southern California with the goal of expanding participation and the extent of data collected.

San Francisco Estuary Institute Regional Data Center

The San Francisco Estuary Institute (SFEI) RDC provides coordination, data management, and information access for its five major programs: Contaminant Monitoring and Research, Wetlands, Watersheds, Biological Invasions, and Historical Ecology.

SFEI hosts several high profile programs and projects that collect and distribute ambient monitoring data including:

- Regional Monitoring Program for Water Quality (RMP),
- Wetlands Regional Monitoring Program (WRMP)

    - Includes: California Rapid Assessment Method (CRAM) that evaluates the ecological health of a wetland,
- Wetland Tracker that provides project information in a geospatially referenced tool for finding out about projects undertaken in wetlands in the region (and state),
- CalFed - Fish Mercury Project (FMP) to investigate contaminants in fish of the Bay-Delta watershed, and
- Other studies those are relevant to managing and understanding the ecological conditions of the region.

SFEI's programs focus mainly on the San Francisco Estuary and Delta regions, but state, national, and global programs provide scientific and technical expertise and services.

SFEI's RDC participates in the ongoing implementation and infrastructure of SWAMP and CEDEN, and provides technological, scientific, and community outreach needed to integrate the environmental data management system. The broader role of the SFEI RDC is to serve as a portal for stakeholder coordination and scientific information exchange with the scientific and regulatory community, as well as the interested public and stewardship groups.

The SFEI RDC employs a customized version of the SWAMP v2.5 database structure and has three primary functions:

1) Data compilation – where data management services for discrete ambient monitoring data submissions (including formatting templates, documentation and training, data format verification, QA/QC validation, and compiling data into the v2.5 database system)
2) Data exchange – where datasets are exchanged with CEDEN (on a periodic basis) making regionally relevant data accessible to the public through CEDEN website and the SFEI information portal
3) Information access – providing web access to interpretive scientific reports, data, interactive data summary tools, project inventories (metadata catalogs), and stakeholder coordination and management tools.

These functions form the foundation of the CEDEN RDC's business plan, and RDCs share and customize the tools for these primary functions. SFEI is also providing services to assist interested parties to participate in the web services and data management and access models employed by the SFEI RDC.

SFEI's RDC currently manages data and provides access to data and interpretive reports for environmental managers and the public for several regional monitoring projects. The data compilation services provided to these projects include templates for laboratories (or projects) to format and report their data to the RDC, data review for compliance with CEDEN's formatting standards, QA/QC review for compliance with each projects expected quality assurance measures, formatting documentation, and training services for laboratories or project/programs wishing to submit data to SFEI's RDC.

The SFEI RDC also has two active project directories (metadata catalogs) that summarize project information for wetlands projects around the estuary and monitoring projects in the San Joaquin River basin. SFEI is currently developing a third directory to use for all project/program data that is loaded to SFEI's RDC. These directories can periodically submit data to the Resource Agency's California Environmental Resource Evaluation System (CERES) catalog, which is the centralized metadata repository for the state.

Plans include expanding these data compilation services to allow non-SFEI data providers (regional agencies, state grant projects from Propositions 13, 40, 50, and 84, citizen monitoring groups, etc.) to upload data to CEDEN. These services will include data format templates, a web-based data uploading tool that will review datasets for compliance with CEDEN's formatting standards and training services for laboratories or agencies wishing to load data to CEDEN via SFEI's RDC. All datasets will exchange with the CEDEN via SFEI's RDC.

SFEI's portal allows data users (the public, environmental managers, and the scientific community) to access reports and summary information about a host of contaminants and ecological measures relevant to understanding the condition and trends of key environmental indicators in the Estuary and Delta. Through this portal, interactive data access and summaries are accessible and include:

- Charts that compare contaminant concentrations to ecological-effects and/or regulatory thresholds, or comparisons of average concentrations of pollutants of concern around the Estuary, and
- Maps that present the relative contaminant concentrations (or ecological measure) around the Estuary and/or Delta
- Subsets of "analysis ready" data available for further analyses.

The portal also serves as a resource to:

- Access information about stakeholder meetings, agendas, minutes, and scientific reports generated by SFEI (and possibly other agencies),
- Link to other websites involved in environmental activities in the region related to SFEI's programs including (such as the Wetland and Beaches portals, SWAMP, the Marine Mammal Center, USGS and DWR sites, and CEDEN), and
- Participate in informal discussions about environmental issues related to the health of the San Francisco Estuary and its adjacent watersheds.

The SFEI portal is in the first phase of development with a redesigning of SFEI's website to utilize a Content Management System that will improve the website management. The following new interactive tools address the data reporting needs of specific, funded projects:

- The RMP - Status and Trends Program's contaminant monitoring data for water, sediment, and bivalve bioaccumulation (1993-2006) are available through a "data access tool" that complies with the SWAMP/CEDEN database standards. The program's Annual Report (summary charts and maps formerly reported in .pdf

report formats) is updating interactive query and display tools that will be available on the web in December 2008.

- The Fish Monitoring Project (FMP) will have similar data access and display tools for fish contaminant, bioaccumulation data from the Bay-Delta watershed (2007-2008). Developments of these tools are with project funding outside the SWAMP/CEDEN funding support.

Additional funding is needed to adapt these services and tools to address the information needs of a broader public, scientific, and management community and to implement non-SFEI data management and access services. The large amount of standard-format data that will be available through CEDEN will greatly expand the interpretive capabilities of monitoring information for the region, state, modeling, interpreting, and communicating environmental condition and trends. The SFEI RDC is committed to coordinating with the OIMA at the SWRCB, other RDC, and the CEDEN coordinator to make the statewide-federated data management system and portal services functional, flexible, informative, and relatively easy to use.

UC Davis Regional Data Center

The UC Davis Regional Data Center currently maintains water quality monitoring data collected by Central Valley monitoring programs including the Organophosphate Total Maximum Daily Load Program (OP TMDL), Irrigated Lands Regulatory Program (ILRP), Phase I and II Central Valley Regional Water Quality Control Board Monitoring, ILRP Coalition Monitoring, Proposition 50 projects, and Pesticide Reduction and Investigation of Source Mitigation (PRISM) investigations. Aquatic Ecosystems Analysis Laboratory (UCD-AEAL) manages UC Davis RDC at UC Davis, which also conducts monitoring for a variety of water quality projects. UCD-AEAL collects, verifies and validates prior to uploading data onto the CEDEN website. UCD-AEAL Database Management staff ensures comparability between data from outside sources and CEDEN.

The UC Davis RDC is working closely with the ILRP Database Management team to review, transform and verify ILRP data that will move onto the CEDEN website. Currently, the UC Davis RDC assists the ILRP Database Management team by reviewing older data for CEDEN comparability and uploads the data into the design master database housed at the UC Davis RDC. Within the next year, the ILRP will upload their own data and synchronize their database with the UC Davis RDC where the data will receive a final review before transferring into CEDEN. The goal of this relationship is to ensure that ILRP data is CEDEN comparable and is publically accessible in a timely manner.

Goals for the future of the UC Davis RDC include working with monitoring programs across California to verify, validate and move water-monitoring data to CEDEN. Specifically, the UC Davis RDC will expand its role in data management by:

- Assisting the State Water Resources Control Board and the Regional Water Quality Control Boards in bringing data from various assessment and monitoring programs into CEDEN

- Assisting Irrigated Lands Regulatory Programs from across the state to incorporate their data into CEDEN
- Assisting the incipient regional monitoring programs in the Central Valley to insure their data is publically accessible through CEDEN
- Serving as a resource for grantees from various proposition grant programs (Proposition 40, 50, 84) in developing SWAMP comparable databases and migrating their data to CEDEN through the UC Davis RDC
- Identifying the data analysis needs of stakeholders and facilitate the development of web accessible tools
- Serving as a resource for stakeholders to resolve SWAMP and CEDEN comparability issues including quality assurance analyses, data comparability, and data business rules

DWR RDC INFO goes here!!

Data Users

Data users need integrated access to monitoring data across different monitoring programs. This group is comprised of the casual, which is the more dominant user group, and advanced users. Casual users include those with simple queries on the Internet. Advanced users are characterized by planners and researchers that include agency and industry scientists, academicians, consultants, and other non-governmental organizations and advocacy groups. Researchers include SWRCB and RWCB staff working on 305(b) reporting, 303(d) listings, and TMDL calculations and listings. In addition to raw data users, the public, legislators, and decision-makers can develop Web-based systems that transform raw data into information and provide refined analysis of data on CEDEN for general consumption.

Advanced users utilize the data in conjunction with user-supplied analytical tools, such as simulation models, or apply the data to TMDL calculations and 303(d) reporting, which can all be used to help make decisions that can greatly affect water resources management and regulation in the state. This group in most cases needs access to large volumes of high-quality multi-agency data. Alternatively, they may need a representation of these data in other formats such as charts, plots, or maps. These users are typically willing to spend a couple of hours to learn how to use sophisticated query tools that may also have utilities to convert data into information; in many cases, they can develop their own custom applications to retrieve and display data. Examples include a web display of contaminated water bodies for public consumption, but could be third party portals accessing the same corpus of data (anything from Google, through educational web sites creating language-appropriate interpretations for K-6 graders, to supercomputer center-based simulated flyovers).

Casual users require simple web based tools. They expect to obtain information shortly after reaching the web site. These tools may be provided by the RDC themselves or part

of the general suite of CEDEN query tools. In addition to raw data users, the public, legislators, and decision makers can develop Web-based systems that provide refined analysis and display of data on CEDEN for general consumption.

Although advanced users make the substantial contributions to planning studies, water exports, adaptive management, and use larger amounts of data, they represent a small percentage of the user community at this time, which leaves the majority of users in casual user group.

Significant research and application development with input from stakeholders' academics, users, management, and the public will successfully convert data into meaningful information. One of the advantages of the distributed system is provides the opportunity for this type of research and development at many different locations across the state at the RDCs.

CEDEN Principles

Any successful approach to integrating data developed by a diverse group of data collectors and programs will have:

- Architecture that is distributed, flexible about platforms and formats, readily scalable, and accessible to meet the needs of the different user groups.
- Common standards for naming database objects, parameters, and QA/QC, and
- Metadata that thoroughly describes the data.

Highly Distributed Architecture

A highly distributed solution is needed to meet the needs of the data providers and the RDC in CEDEN. The distributed solution should allow for a reliable replication of data and updates.

Platform Flexibility

Platform flexibility is required since the CEDEN members can have different computing infrastructures without a single standard although some of the RDC are adopting the architecture and data systems developed by SWAMP and OIMA. This provides the opportunity to share application and developmental costs. Differing platforms must be able to communicate with each other to achieve the distributed solution.

Readily Scalable

The solutions used for CEDEN should be scalable to accommodate more data providers as well as the RDC. The spectrum of solutions should cover small systems using Microsoft Access and Excel spreadsheets as well the larger providers with specialized databases, web services, on-line models, and visualization tools. External participants will need tools to both provide and validate data to RDC.

Accessibility

Accessibility of the data should be considered as an important factor in implementing this system. RDCs should make their data available online both to their user community and to CEDEN. Data from CEDEN will also need to be provided to the RDC. Web

accessibility standards should be considered when designing the interfaces for data retrieval.

Version Control

The authoritative versions of data should be held as closely as possible to those who generate and maintain them, though it is often desirable to distribute copies of data to a more centralized facility. If users can access multiple copies of data sets, it is essential that changes in the primary source of data cascade to all copies and track the data changes.

Metadata

Data needs to be documented by extensive metadata using accepted metadata standards. At present, federal partners are mandated to provide documentation compliant with the Federal Geographic Data Committee (FGDC) metadata standards, at least for geospatial and biological datasets and most important GIS datasets held by CEDEN members have FGDC metadata. There are several existing metadata systems available including one developed by the Resources Agency. Utilization of an existing metadata system should be endorsed to help reduce costs and take advantage of the expertise available from others who already developed metadata documentation. For CEDEN purposes, it is likely that a more expanded metadata profile will need to be adopted for non-FGDC metadata.

Where feasible, peer-review of both data and metadata should be encouraged.

CERES INFO GOES HERE

Use of SWAMP/CEDEN Standards

In many cases, data providers will be required to use vocabularies, QA/QC procedures, and monitoring protocols developed by the SWAMP program and adapted by CEDEN. Historical data and other data types that are needed by the data user community that do not meet standards will need to by clearly indentified and cross-walking and other tools will need to be developed to include these data in CEDEN. Without standardization of data, complete interoperability among different data providers is all but impossible to achieve.

QA/QC Levels

Metadata and data dictionaries need to record the nature of QA/QC for data (*trust* characteristics). The user can then use this information to help identify the suitability of the data for different analysis.

Technological Solutions

There are multiple possible technological paths and combination of technologies consistent with the user needs and the principles outlined. What is presented in this report is a hybrid of technologies that will provide a proven approach to answering the enterprise need for data and information in California, based on CEDEN members'

experience managing large data sets and their knowledge of the various institutional practices of data collectors in the state.

The hybrid solution will be comprised of data storage components, data exchange components, and the data retrieval or query components. Each component will be further discussed in this report. Candidate technologies will be identified for each component and will be evaluated in light of the user needs and the CEDEN principals.

Data Storage Component

RDC and data providers must store and manage data reliably in order to contribute it to a distributed data and information system. Relational Database Management Systems (RDBMS) will greatly assist with CEDEN data management because of their ability to store and relate the diverse types of physical, chemical (e.g. water quality, hydrodynamics, meteorological), biological, terrestrial, wetland, fisheries, GIS, and modeling information collected in the state from many different data providers. Data submitted from many different data providers to a RDBMS located at RDC will be stored in tables related to each other according to key fields (location, date, time, data type, et cetera) and made accessible online. The resulting database would represent comprehensive data sources for the region. This database will be an invaluable tool for the RDC to conduct analysis and reporting results and information; RDC staff could perform simple and refined queries, obtaining data they need from numerous sources quickly and efficiently. Data can be retrieved from the RDBMS at a specific RDC or the ones located at other CEDEN RDC. RDBMS also have other tools important to the success of CEDEN including linking to RDBMS to other RDC so that data can be shared and the ability to propagate data changes throughout the CEDEN.

RDBMS may be implemented as user customized databases using specified table structures, controlled vocabularies for monitored attributes, and database objects; integration of project business rules as well as extensive tools for: reporting data, conducting QA/QC analysis, data management, and loading data can also be included. There are several examples of customized databases in the state including the implementation of the SWAMP data management system at Moss Landing Marine Labs and several programs being implemented by the San Francisco Estuarine Institute (SFEI) and Southern California Coastal Water Research Program (SCCWRP), UC Davis, and else where in the state. Program specific databases that use the SWAMP standards make excellent data sources for the comprehensive RDBMS.

The comprehensive RDBMS will utilize the CEDEN standards and staff at the RDC will establish and implement QA/QC and documentation regarding the source data sets. The comprehensive RDBMS will be used to share data with other RDC in CEDEN.

The comprehensive RDBMS located at the RDC have a multi-tiered RDBMS architecture that includes transfer databases, a normalized database, and various data marts and star schemas that facilitate application development (Figure I). As shown in Figure III, data to be included in the main data store, or normalized databases, is first loaded into a permanent transfer database. Data in the transfer database is mapped to the

normalized database schema. Mapping from the transfer database includes determining where data needs to be placed in the normalized database shown in Figure I as the center blue circle. Primary look up values in the transfer database are then cross-walked to the corresponding values in CEDEN before they are placed in the normalized database. Great care and much staff time is needed to make sure cross-walks do not misrepresent data and when new values are received they need to be reviewed to determine if they are valid and should be added to the CEDEN standards. Processing data into CEDEN can integrate data, but without the use of standardized QA/QC, monitoring and data handling procedures, and the interoperability of data from one monitoring program to another is greatly reduced. Promulgation of the SWAMP/CEDEN data standards will improve both the use of data, greatly reduce the cost, and improve the accuracy of data processing procedures since data will no longer need to be cross-walked.

The process of placing data into the normalized RDBMS consist of a program that identify new records in the transfer databases and moves them into the normalized database using the standardized cross-walks. The program writes back the primary key information from the normalized database to the transfer databases, along with the time and date the record was added. Data from the normalized database are then used to populate data marts and star schemas, which are stored in the application databases, displayed in Figure II, and are used by data access applications, which are shown on the right hand side of Figure II. Applications include programs that provide web access to monitoring data, highly refined information for the public, web services, and other types of data dissemination programs.

Moving data through the RDBMS components of CEDEN involves the transfer and aggregation of hundreds of millions of records and maintaining version and quality control over the various physical representations of these data. This requires an extensive optimized code set that not only moves data accurately, but also moves it quickly so that it can be implemented at regular intervals throughout the week.

Data loaded into the CEDEN RDBMS can come from multiple local and client databases. These local databases are typically developed in MS Access or SQL server by the CEDEN RDC staff working together with the data provider, a process that also serves to help providers better organize their data and manage it once the system is in place. Data in various client databases are combined and synchronized with existing data on the server using database replication. A record of all transactions between the main server and local databases can be kept such that this, combined with the method of synchronization, ensures version control between incoming and stored data. Data in the transfer databases need to be evaluated for adherence to QA/QC and QAPPs to determine their comparability at each RDC. QA/QC procedures for all data handing procedures will also need to be run to help ensure data quality.

The RDC' staff works with their monitoring community on other non-IT aspects of data management such as how to comply with the CEDEN/SWAMP monitoring and data standards, QA/QC, and coordinating monitoring activities in their region. The RDC' role in CEDEN is further defined earlier in this report.

Data will also be provided to CEDEN either directly from other group's enterprise data systems, static data loads, or via the web through an application that is in development at the SWRCB.

Data Exchange Component

Three data exchange mechanisms are used to meet the data exchange component requirements: 1) direct database connection, 2) web services, and 3) web input.

Direct database connection is the simplest mechanism, which provides linkage between RDMBS. This technology is readily available and typically does not require extensive software or system administration support. Data can be synchronized using database replication and synchronization procedures and programs.

A web service is an infrastructure strategy that promotes development of individually accessible distributed components using loosely coupled and reusable software architecture**.** For example, web services can be used in an Internet/Intranet configuration to develop web and standalone applications. The web services utilized by CEDEN are specified by the EPA's Exchange Network and will be used to supply data to the EPA. Other CEDEN groups can exchange data through Web services.

Web Data Entry

Several forms of web data entry will be provided, including a batch upload system that utilizes standard data and transfer formats to enable file upload data deliveries to the MLML data center, and web forms for direct transcription of field sheet data into the standard data transfer formats. The web data entry system builds on work previously completed as a component of the California NPS Citizen Monitoring Program data collection system. (http://www.ccamp.net/simon/index.php/Main_Page)

The Consolidated Grants Program requires that grantees meet certain electronic data reporting requirements for assessment of grant effectiveness. Grant agreements require that data be delivered in an electronic format that is consistent with the Surface Water Ambient Monitoring Program data management system. No technical support is currently provided to aid grantees in meeting this requirement. The web data entry system will address these grant requirements, the intent of SB 1070, and other legislation by providing a data delivery framework to move data from grant projects to the CEDEN data management system. The types of data that will be supported by the tools include water quality, toxicity, bacteria, and bio-assessment data records. The web-based system is designed to check for errors in data formatting and content. The upload system will simplify the process of submitting and sharing data and will support data transfers from CALFED Watershed grant projects and other programs to the CEDEN data management system.

Data Retrieval Component

The data retrieval components of CEDEN should address the needs of the data users. The candidate system for the casual users includes a Web browser that is a map enabled for

extraction of data spatially and the ability to refine queries based on analytes, methods, etc. This technology can be adopted at all RDCs. More customized retrieval systems will need to be developed to meet specific group's needs.

The great advantage of working on CEDEN is most of the participating RDC have extensive IT staff that will contribute to data retrieval systems both for the scientific community and public outreach. The ability to develop retrieval systems that convert the monitoring data in CEDEN into information greatly enhances many operational, adaptive management, reporting, public outreach, and research efforts already underway in the state. Using distribution technologies provides the opportunity for many groups to develop customized data retrieval systems that meet their specific needs or to develop processes that convert data into information that they can share with other interested parties via the Internet or other types of media.

Conclusion and Recommendations

The use of distributed data technology and the SWAMP/CEDEN standards is proposed to facilitate data access, sharing, and standardization of the state's environmental monitoring data so data can be used interchangeably between multiple monitoring programs and reporting systems. The proposed mechanism is flexible in that it can be set-up to deliver specifically formatted data to decision support groups used by many participants to provide comprehensive data for integrated research projects. In addition, data considered preliminary and not ready for broad distribution, can be controlled and made accessible only to specific data user groups.

An *ad hoc* CEDEN data committee was formed and is beginning the coordination, implementation, and management of CEDEN, but ultimately this effort will require dedicated leadership. The OIMA at the SWRCB with its integration of IT and biological and engineering staff will provide the necessary state support; however, the ultimate success of the program will require strong institutional commitments between the groups hosting the RDC, EPA and SWRCB. CEDEN participants should consider the creation of a more formal committee that includes representatives from the RDC, EPA, SWRCB, and as needed basis representatives from various agency, academic, stakeholder, and private sector participants who have experience and expertise in distributing data/information systems or can provide user specifications for the system. This group should address tactical planning and implementation of the ideas discussed in this report, help determine how changing technologies should be applied to the effort, and determine the staffing needs of a distributed data/information system,

To help move the above ideas forward, a series of user needs assessment meetings are being planned to help develop the CEDEN web site and initial basic web interface design. The SWRCB's OIMA has hired a data coordinator. The California Department of Water Resources has disclosed they are currently unable to support CEDEN so work has begun to finish setting up a replacement system at the Moss Landing Marine Labs.

Figure I

CEDEN Regional Data Centers
Enterprise Components

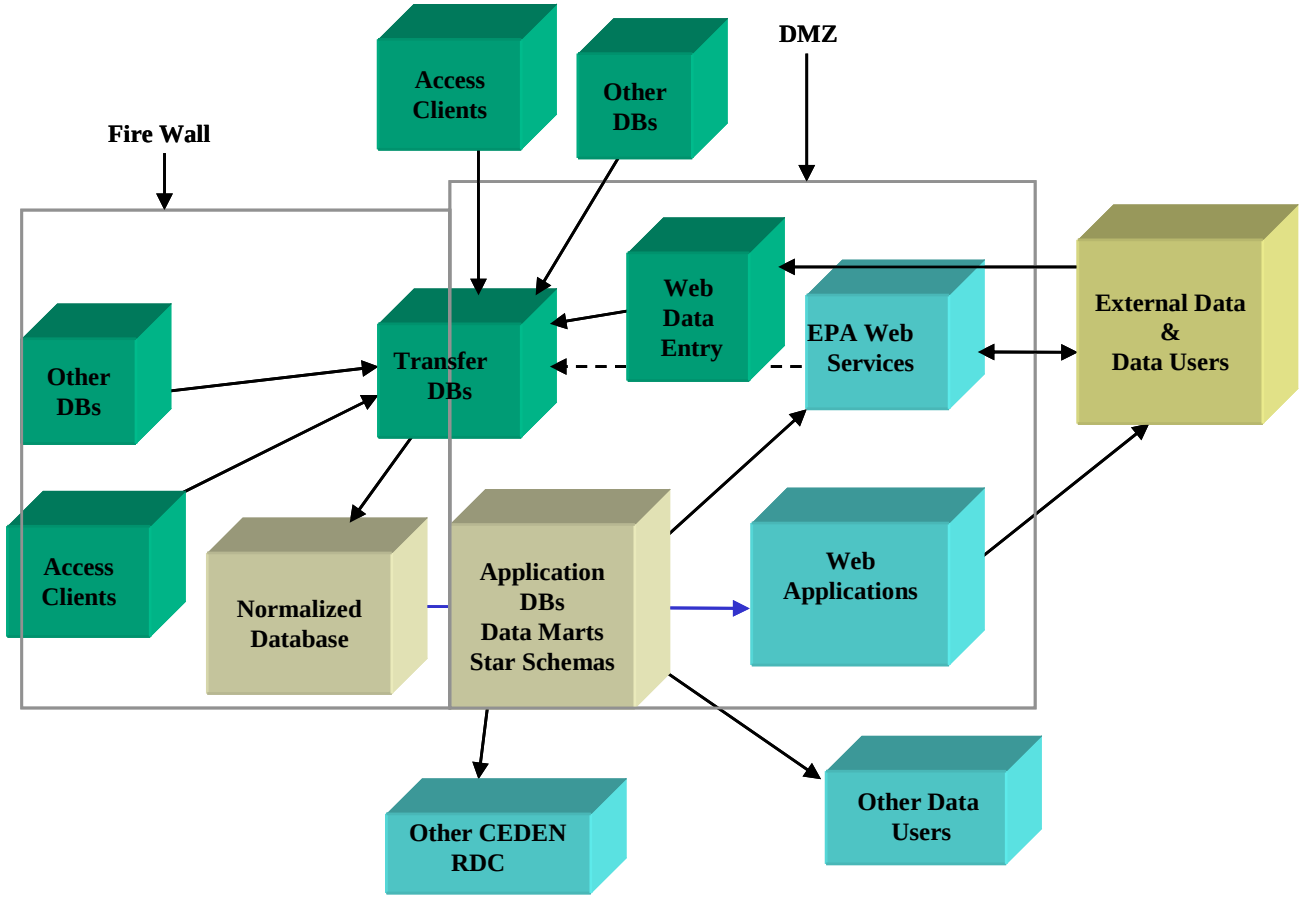

Figure II

CEDEN Network

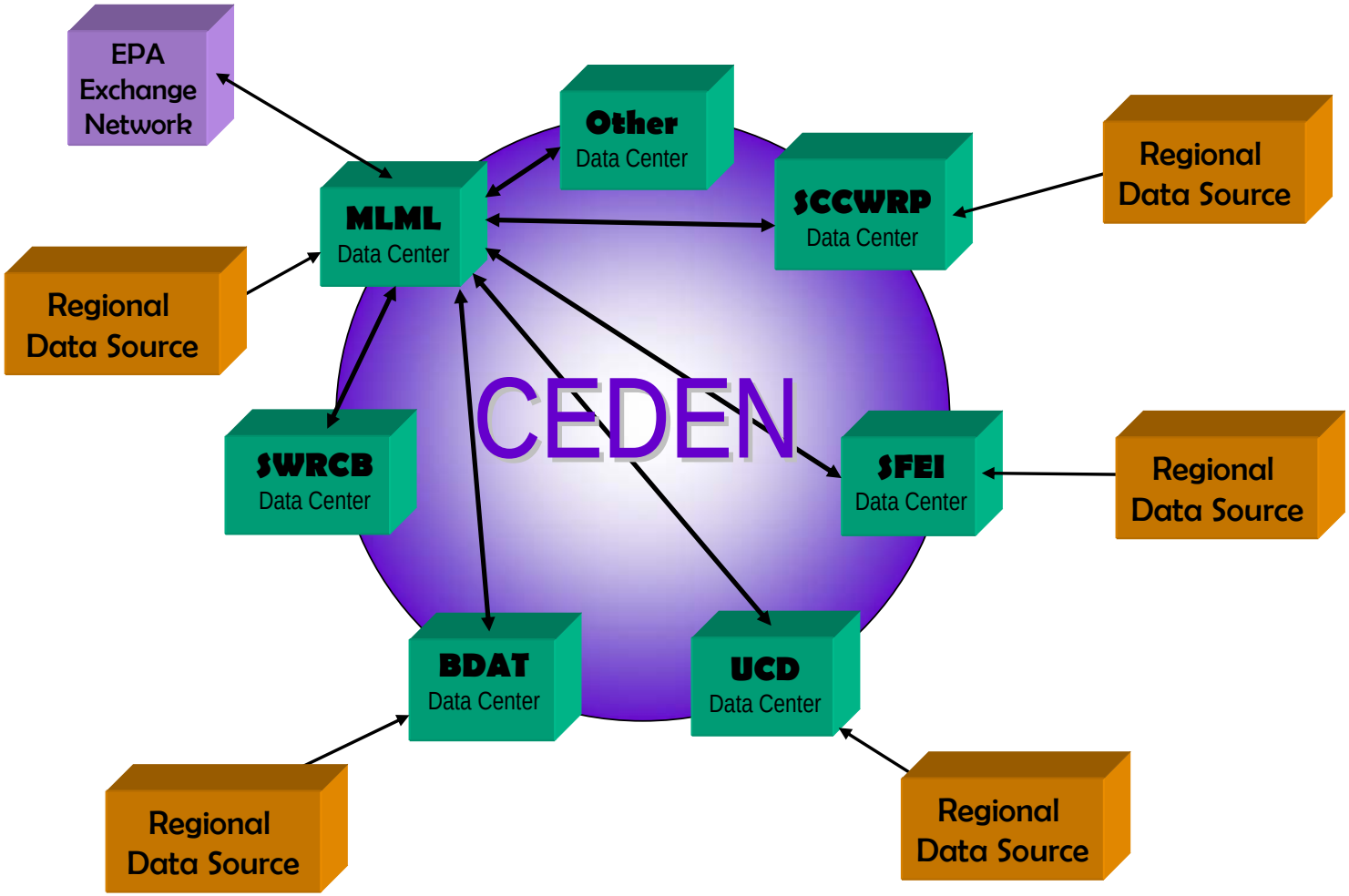

Figure III

Background Data Flow Characteristics